SUPPLEMENTARY METHODS, TABLES AND FIGURES

Supplementary Methods 1

Tissue staining and scoring

TMEM staining

Tissue sections were stained in batches of twenty, with positive controls for TMEM, blood vessels, and macrophages. Positive controls were identified from prior breast cancer cases and these served as external positive controls; also, both the benign and malignant areas of sections from breast tissue of the included subjects contained an inherent population of blood vessels and macrophages that were examined as internal positive controls. As a negative control, we used non-immune mouse IgG diluted to the same concentration as the test antibodies.

Blank sections were cut at 5 µm, baked at 60ºC for 30 minutes, and deparaffinized.

All immunostains for the triple stain were performed on the Bond Max Autostainer. Three antibodies were applied sequentially, and developed separately with different chromogens. The sequence was anti-CD31, then anti-CD68, followed by anti-panMena.

CD 31: Antigen retrieval was performed with Bond Epitope Retrieval Solution 2 (Leica Biosystems) at 99-100°C degrees for 20 minutes. Endogenous alkaline phosphatase was quenched in Dual Endogenous Enzyme Block (DAKO) for 5 minutes. Incubation with anti-CD 31 (clone JC70A, DAKO) was performed at 1:800 dilution for 30 minutes. Slides were then incubated in Post Primary AP (Bond Polymer Refine Red Detection, Leica Biosystems) for 20 minutes, followed by incubation for 30 minutes in Polymer AP (Bond Polymer Refine Red Detection, Leica Biosystems). Colorimetric detection was done with Vector Blue (Vector Laboratories) for 10 minutes.

CD 68: Antigen retrieval was performed with Bond Epitope Retrieval Solution 1 at 99-100°C for 30 minutes. Incubation with anti-CD 68 (clone PGM1, DAKO) was at 1:300 dilution for 15 minutes. Incubation with Bond Post Primary (Bond Polymer Refine Detection, Leica Biosystems) was for 8 minutes followed by incubation with Polymer solution from the Bond Polymer Refine Kit for 8 minutes. Endogenous peroxidase was quenched with Peroxide Block (Bond Polymer Refine Detection, Leica Biosystems) for 5 minutes. Colorimetric detection was with DAB for 10 minutes (Bond Polymer Refine Kit, Leica Biosystems).

Pan Mena: Incubation with anti-pan Mena (clone A351F7D9)(1) was at 1:200 (5ug/ml) for 15 minutes. Slides were then incubated in Post Primary AP (Bond Polymer Refine Red Detection, Leica Microsystems) for 20 minutes, followed by incubation for 30 minutes in Polymer AP (Bond Polymer Refine Red Detection, Leica Biosystems). Colorimetric detection was with Fast Red (Bond Polymer Refine Red Detection, Leica Biosystems) for 5 minutes. Counterstain was with Light Green SF Yellowfish (Fisher Scientific).

Positive control slides were evaluated for crisp, clear staining of each antibody. The expected results of immunohistochemical staining were as follows: CD68 – granular, dense and diffuse brown cytoplasmic staining of all macrophages and of macrophages only; CD31 – smooth, glassy and diffuse blue cytoplasmic staining of all endothelial cells and of endothelial cells only; and pan Mena – granular and diffuse red cytoplasmic staining of all carcinoma cells and most intense in carcinoma cells associated with TMEM (Supplementary Figure 1). The light green counterstain allowed sufficient visualization of the stromal and cellular (including nuclear) detail in order to facilitate identification of invasive carcinoma, in situ carcinoma, and benign breast epithelium and allow them to be distinguished by morphology from stromal elements and inflammatory cells. If the results of the staining of either the controls or of the tissue from the

study subjects was suboptimal for any of the antibodies used, the procedure was reviewed and repeated for the entire batch.

Reference

1. Lebrand C, Dent EW, Strasser GA, et al. Critical role of Ena/VASP proteins for filopodia formation in neurons and in function downstream of netrin-1. *Neuron*. 2004;42(1):37-49.

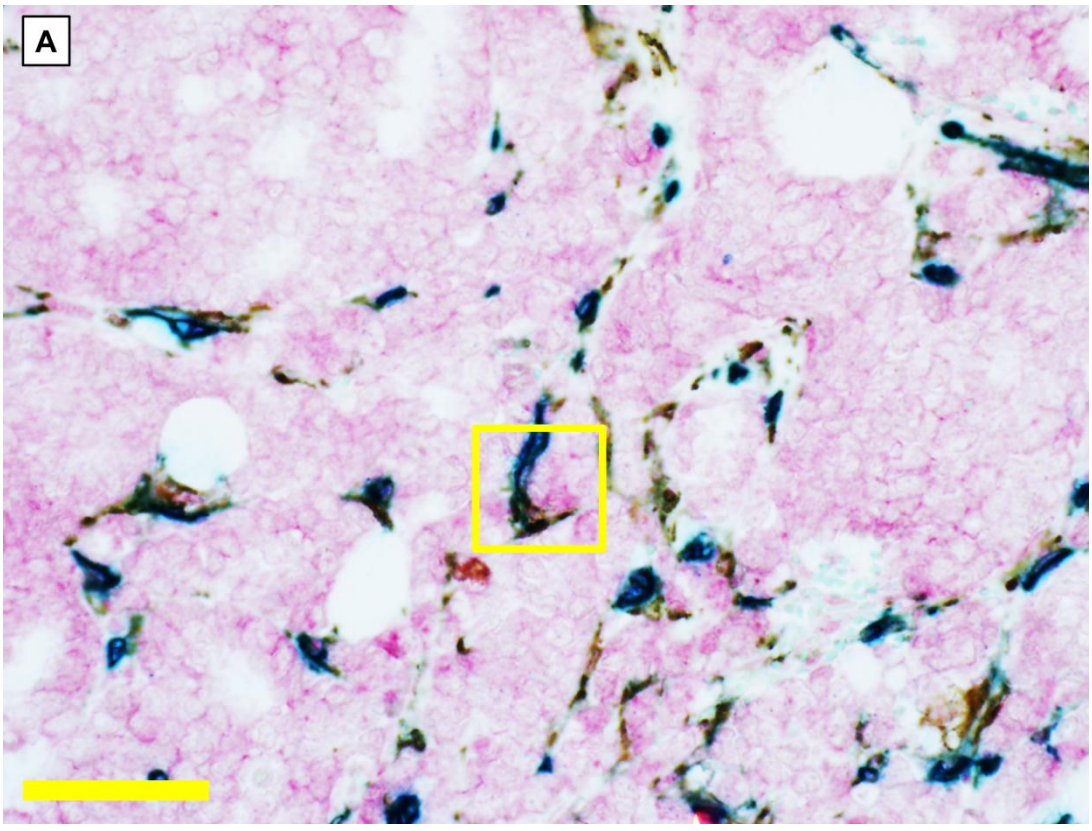
TMEM assessment

TMEM assessment was performed using Adobe Photoshop on 10 contiguous high power digital images. The appropriate area was identified by low power scanning assessment, focusing on representative high density of tumor, adequacy of tumor, lack of necrosis or inflammation, and lack of artifacts such as retraction or folds. Once a representative area was identified, ten digital images were acquired at 400x total magnification. The pattern is a "9 + 1" sequence and forms a square with one extra field adjacent to the square to make a total of ten fields. Opening each image one at a time, the entire image was scored for TMEM. One TMEM is defined as a structure composed of the direct contact between an invasive pan Mena-overexpressing carcinoma cell (detected with pan-Mena, an antibody that recognizes all Mena isoforms), an endothelial cell, and a perivascular macrophage, with no discernible stroma between the tumor cell and the perivascular macrophage. Using the "circle" tool available in Photoshop, all TMEMs were "marked" using circles, and the "marked" images were saved as separate files. The total number of TMEMs for each image was tabulated, and the scores from all ten images were then summed to give a final TMEM score for each patient sample, expressed as the number of

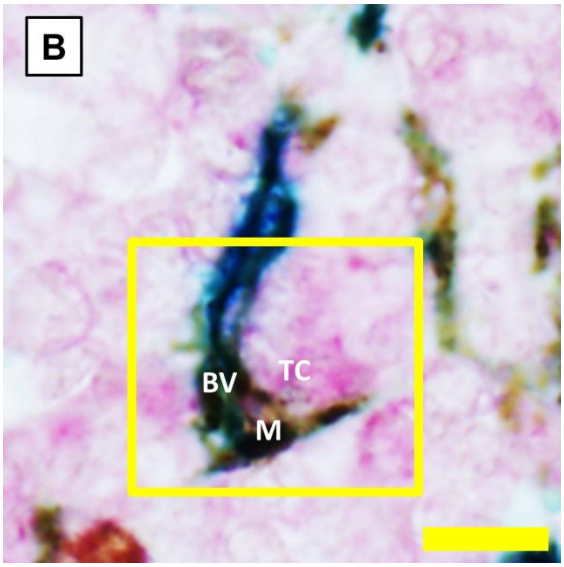
TMEMs per 10 high power fields (400x total magnification each). Examples of TMEMs are shown below in Supplementary Figure 1.

Supplementary Figure 1: Fields showing TMEM.

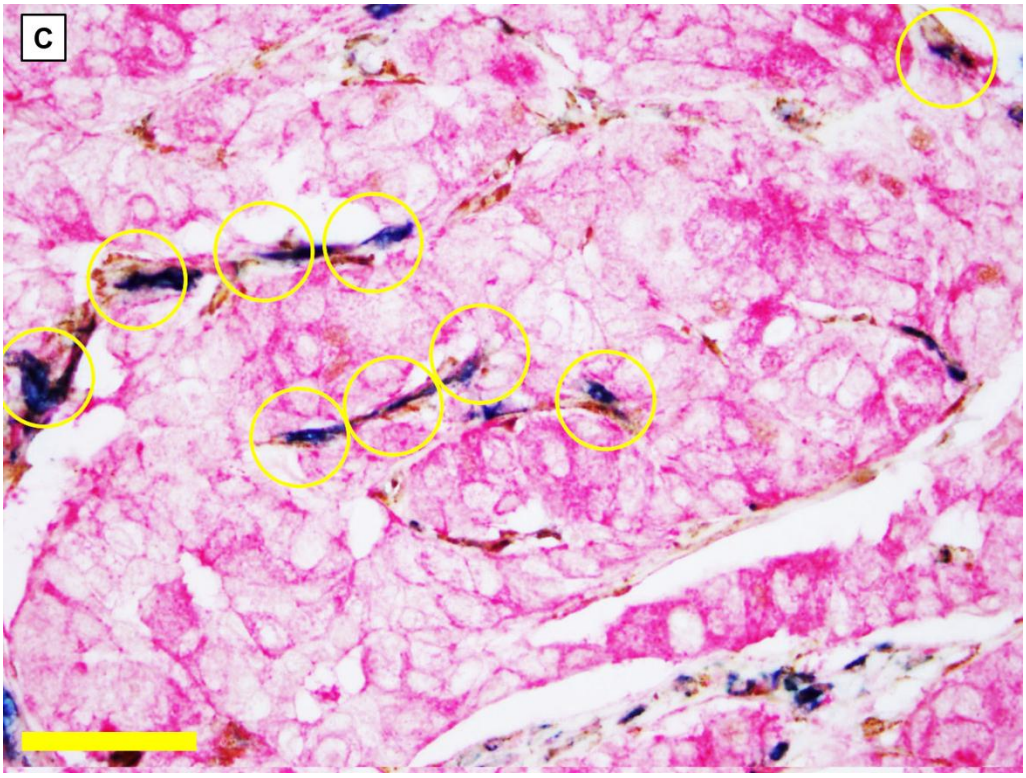
A. Field at 800x showing typical TMEM (box). Scale bar = 75um. B. TMEM (box) in A at 2x higher magnification with cell types in TMEM labeled (TC, pan-Mena red stained tumor cell; M, CD68 brown stained macrophage; BV, CD31 blue stained endothelium). Scale bar = 20um. C. Field as scored for TMEM at 400x with TMEMs circled. Scale bar = 75 um.



A



B



C

Cases were randomly assigned to five pathologists for TMEM scoring. To assess the intra- and inter-reader repeatability of the TMEM scoring, the first 60 triple-immunostained slides generated as part of the study were read twice by all pathologists. The repeatability study was conducted in two phases. During phase one, the stained slides were divided into five equal-sized groups (12 per group) and then distributed using random assignment to the five pathologists. Each pathologist created 10 images for each of his/her assigned slides and uploaded the images into a custom designed web-based application. The 600 uploaded images (60 samples*10 images per sample) were replicated in the system for each pathologist and access to these 600 original (not yet annotated) images was provided to the five pathologists. Each pathologist was then able, via the custom application, to access their copy of all 600 images and annotate each one using Adobe Photoshop. The pathologists only had access to the annotated images that they created. After annotation of each set of 10 images was completed, the pathologists were prompted by the system to provide a TMEM score for each of the 10 high-powered fields selected. The web-based application ensured that data were not shared among pathologists so that each pathologist was blinded to the scores assigned by the other pathologists to a particular section. When this phase was completed, the 600 original images were assigned new ID numbers so that the pathologists were also blinded to their own original scores. The process for phase one was then repeated, wherein each pathologist was instructed to access the 600 original re-labeled (not yet annotated) images and then annotate and score them. The data were saved via the web-application and the repeatability analysis was conducted based on the data from each phase. The intraclass correlation coefficients were 0.63, 0.74, 0.82, 0.85, and 0.91 for within-rater agreement, and 0.75 for between-rater agreement.

Methods for ER, PR, Ki67 and HER2 Immunostaining

The methods for ER, PR, Ki67 and Her 2 immunostaining are summarized in the table below.

Assay	Scoring	Staining					Counter-stain
	Scoring method	**Antibody & Supplier**	**Antigen retrie-val**	**dilut-tion**	**incub-ation**	**Detection system**	**Hematoxyl-in type & supplier***
ER	H-score	Dako M7047, Clone 1D5	Dako target retrieval solution, S1699	1:50	30 min.	Dako-diaminoben-zidine (DAB)	Surgipath Hema-toxylin, Leica Biosystems
PR	H-score	Dako M3569, Clone PGR636	Dako target retrieval solution, S1699	1:200	30 min.	Dako-diaminoben-zidine (DAB)	Surgipath Hema-toxylin Leica Biosystems
HER2	ASCO/ CAP guideline (1)	HER2, 28-0004, Clone Z4881, Invitrogen	Dako target retrieval solution, S1699	1:50	30 min.	Dako-diaminoben-zidine (DAB)	Surgipath Hema-toxylin Leica Biosystems
Ki67	% positive cells (2)	Ki67, M7240, Clone MIB-1 Dako	Dako target retrieval solution, S1699		30 min.	Dako-diaminoben-zidine (DAB)	Surgipath Hema-toxylin Leica Biosystems

* All staining times were 3 minutes

1. Wolff AC, Hammond ME, Schwartz JN, et al. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. *J Clin Oncol.* 2007; 25(1):118-45.
2. Dowsett M, Nielson TO, A'Hern R, et al. Assessment of Ki67 in breast cancer: recommendations from the international Ki67 in breast cancer working group. *J Natl Cancer Inst*. 2011;103(22):1656-64.

Ki67 and ER/PR receptor assessment

Ki67 and PR scores were recorded as the percentage of positively staining malignant cells. These percentages were determined as a visual estimate of the Ki67 positive and PR positive invasive tumor cells over the entire tissue section.

ER was quantified for IHC-4 by using the H-score, which is defined as the percentage of cells staining weakly plus two times the percentage of cells staining moderately plus three times the percentage of cells staining strongly.

Supplementary Methods 2

Calculation of IHC4

As per Cuzick et al. (2011), the equation used to calculate the IHC4 score was

$$IHC4 = 94.7 \times \{-0.100\ ER_{10} - 0.079\ PgR_{10} + 0.586\ HER2 + 0.240\ \ln(1 + 10 \times Ki67)\}.$$

The variable ER_{10} was obtained by dividing the H-score by 30 to obtain a variable with a range of 0 to 10. PgR was scored as a percentage of cells staining positive for the progesterone receptor with a positive cutoff of 10%. PgR_{10} was obtained by dividing this percentage by 10 to obtain a variable with a range of 0 to 10. HER2 was scored according to the manufacturer's recommendation: 3+ was positive. Ki-67 scores were recorded as the percentage of positively staining malignant cells. (The distribution of IHC4 scores, overall and by breast cancer subgroup, is summarized below in Supplementary Table 3.)

Reference

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):4273-8.

Supplementary Table 1

Adherence to REMARK guidelines

REMARK Checklist

Checklist Item	This manuscript
Introduction	
1. State the marker examined, the study objectives, and any pre-specified hypotheses.	1. The marker examined was TMEM, assessed using a triple immunostain which identifies Mena-expressing invasive tumor cells, macrophages and endothelial cells in direct contact with each other; this microanatomic structure is the site at which intravasation occurs, as a prelude to hematogenous dissemination. The study objective was to determine whether TMEM score provides prognostic information regarding distant metastasis in operable breast cancer independent of other clinicopatholgic features (defined in Table 1), including the IHC4 score (a validated prognostic marker in ER-positive breast cancer). Our pre-specified hypothesis was that TMEM score would provide prognostic information that was independent of classical clinicopathologic features, and that it may provide prognostic information within pre-specified breast cancer subtypes (ER-positive, HER2-negative; HER2-positive, and triple negative).
Materials and Methods	
Patients	
2. Describe the characteristics (for example, disease stage or co-morbidities) of the study patients, including their source and inclusion and exclusion criteria.	2. Cohort of 3,983 women in the Kaiser Permanente Northwest (KPNW) health care system who received a first diagnosis of invasive ductal carcinoma of the breast, were aged ≥ 21 years at initial diagnosis, were treated surgically, and did not have evidence of metastasis at initial diagnosis
3. Describe treatments received and how chosen (for example, randomized or rule-based).	3. Nested case-control study nested within the cohort. Patients received standard treatment of the primary tumor (surgery +/- radiotherapy) and systemic adjuvant therapy in accordance with contemporary practice guidelines.
Specimen characteristics	
4. Describe type of biological material used	4. Archival formalin-fixed, paraffin embedded primary tumor tissue.

(including control samples) and methods of preservation and storage. ***Assay methods*** 5. Specify the assay method used and provide (or reference) a detailed protocol, including specific reagents or kits used, quality control procedures, reproducibility assessments, quantitation methods, and scoring and reporting protocols. Specify whether and how assays were performed blinded to the study endpoint.	5. TMEM stain is a triple immunostain in which 3 antibodies are applied sequentially, including an anti-CD31 (clone JC70A;1:800 dilution; DAKO), anti CD-68 (clone PG-M1; 1:300 dilution; DAKO), and anti pan-Mena (1:200 dilution; Gertler laboratory); additional details regarding methods for staining, identification, and counting are described in the manuscript and supplemental material. ER, PR, and HER2 were also done in a central lab in accordance with ASCO-CAP guidelines, and Ki67 was also done in a central lab as described in the manuscript and supplemental material. All readings were done blinded to the study endpoint.
<u>Study design</u>	
6. State the method of case selection, including whether prospective or retrospective and whether stratification or matching (for example, by stage of disease or age) was used. Specify the time period from which cases were taken, the end of the follow-up period, and the median follow-up time.	6. Nested case-control ("prospective-retrospective") study design including women with first diagnosis of invasive ductal carcinoma of the breast between 1/1/1980 and 12/31/2000 diagnosed at KPNW. Cases were women who developed a distant metastasis, whereas controls were women who were alive and had not developed a distant metastasis by the date of metastasis of the corresponding case. Controls were individually matched to cases (1:1) on age at and calendar year of the diagnosis of invasive breast cancer (both generally ± 1 year) and were selected randomly from risk sets, with replacement.
7. Precisely define all clinical endpoints examined.	7. The primary endpoint was the occurrence of distant metastasis.
8. List all candidate variables initially examined or considered for inclusion in models.	8. Clinicopathologic features evaluated included age, tumor size, histologic grade, number of positive axillary lymph nodes, treatment, and ER, PR, HER2, and Ki67 expression, and IHC4 score (Table 1).
9. Give rationale for sample size; if the study was	9. Our pre-determined target sample size was 250 case-control pairs, sufficient

designed to detect a specified effect size, give the target power and effect size.	for detection of an odds ratio (OR) of 1.5 per 50 unit increase in TMEM score with 80% power and a two-sided type I error rate of 5% in the entire cohort. Evaluation of the TMEM score in 3 breast cancer subtypes was prespecified (ER-positive, HER2-negative; HER2-positive, triple negative).
Statistical analysis methods	
10. Specify all statistical methods, including details of any variable selection procedures and other model-building issues, how model assumptions were verified, and how missing data were handled.	10. Both univariate and multivariate conditional logistic regression models were used to estimate odds ratios (OR) and 95% confidence intervals for the association between TMEM score and risk of distant metastasis (treated as a binary outcome) for all subjects combined. Clinical/tumor characteristics were adjusted for in the multivariate model. Potential controls considered to be unsuitable (e.g., inadequate tumor tissue) were replaced by another control randomly selected from the same risk set.
11. Clarify how marker values were handled in the analyses; if relevant, describe methods used for cutpoint determination.	11. TMEM score was treated both as a continuous variable as well as a categorical variable, with categorization by tertiles (based on the distribution of the TMEM score among the controls in the entire study population). For breast cancer subgroup analyses, the matching was broken, and ORs were obtained using unconditional logistic regression, with additional adjustment for the matching variables. As a sensitivity analysis, we also used conditional logistic regression on the entire dataset, with interaction terms between TMEM score and subgroup indicator to allow the association with TMEM score to differ by subgroups. All analyses were conducted using SAS (SAS 9.3, SAS Institute, NC) and all p-values are two sided. To compare TMEM score with IHC4 score, we conducted analyses according to the following steps: (1) a multivariate regression model was fitted for IHC4 and the clinical risk factors; (2) a multivariate regression model which included both IHC4 score and TMEM score as well as the clinical risk factors was used to examine if the prognostic value of TMEM score/IHC4 score was affected by the other; (3) a multivariate regression model was used with TMEM score and the four markers in IHC4, as well as the clinical risk factors. Where TMEM score was significantly associated with risk of distant metastasis while controlling for clinical risk factors, we further performed a receiver operating

	characteristic (ROC) analysis and the area under the curve (AUC) was computed. Separate ROC analyses were conducted based on TMEM score alone and TMEM plus clinical variables (tumor size, number of lymph nodes, grade, and hormone therapy) through the use of a TMEM composite score, defined as a linear equation of all the variables in the model with the estimated coefficients from the logistic regression model as their coefficients. An ROC analysis was also performed for IHC4 score.
Results	
Data	
12. Describe the flow of patients through the study, including the number of patients included in each stage of the analysis (a diagram may be helpful) and reasons for dropout. Specifically, both overall and for each subgroup extensively examined report the number of patients and the number of events.	12. Summarized in Figure 1 (CONSORT diagram).
13. Report distributions of basic demographic characteristics (at least age and sex), standard (disease-specific) prognostic variables, and tumor marker, including numbers of missing values.	13. Summarized in Table 1 (Patient characteristics of case and control populations).
Analysis and Presentation 14. Show the relation of the marker to standard prognostic variables.	14. TMEM score distribution by tumor size, histologic grade, and number of positive lymph nodes shown in Figure 2. TMEM score was only weakly correlated with IHC4 score (Spearman correlations: 0.09 in the entire study population, 0.12 in ER+/HER2- disease).
15. Present univariable analyses showing the relation between the marker and outcome, with the estimated	15. Shown in Table 2 (association between TMEM score and risk of distant metastasis overall, and by tumor subtype), and Table 3 (association between

effect (eg, hazard ratio and survival probability). Preferably provide similar analyses for all other variables being analyzed. For the effect of a tumor marker on a time-to-event outcome, a Kaplan-Meier plot is recommended.	IHC4 score and risk of distant metastasis overall, and by tumor type). Kaplan-Meier plot is not appropriate given the case-control design.
16. For key multivariable analyses, report estimated effects (eg, hazard ratio) with confidence intervals for the marker and, at least for the final model, all other variables in the model.	16. Shown in Table 2 (association between TMEM score and risk of distant metastasis overall, and by tumor subtype), and Table 3 (association between IHC4 score and risk of distant metastasis overall, and by tumor type).
17. Among reported results, provide estimated effects with confidence intervals from an analysis in which the marker and standard prognostic variables are included, regardless of their statistical significance.	17. Shown in Table 2 (association between TMEM score and risk of distant metastasis overall, and by tumor subtype), and Table 3 (association between IHC4 score and risk of distant metastasis overall, and by tumor type).
18. If done, report results of further investigations, such as checking assumptions, sensitivity analyses, and internal validation.	18. Results of receiver operating characteristic curves shown in Figure 3.
Discussion	
19. Interpret the results in the context of the pre-specified hypotheses and other relevant studies; include a discussion of limitations of the study.	19. TMEM score provides independent prognostic information for distant metastasis, does not correlate with tumor size, number of positive lymph nodes, or IHC4 score, and correlates weakly with grade.
20. Discuss implications for future research and clinical value.	20. TMEM score is positively associated with risk of distant metastasis in women with ER+/HER2- breast cancer and provides prognostic information that is independent of IHC4 score and other clinicopathologic risk factors. This study provides level 1, category B evidence supporting the prognostic information provided by the TMEM score. Additional validation studies are required in order to provide confirmatory information supporting the clinical

	validity and utility of the TMEM assay.

Supplementary Table 2

Summary statistics for TMEM score by case-control status, overall and by breast cancer subgroup

Tumor type	Status	N Obs	Minimum	25th Pctl	50th Pctl	75th Pctl	Maximum	Mean	Std Dev
All subjects	Case	259	0.0	6.0	19.0	45.0	174.0	30.4	32.4
	Control	259	0.0	3.0	14.0	32.0	184.0	24.3	29.9
ER+/HER2 -	Case	147	0.0	8.0	24.0	51.0	174.0	34.7	35.3
	Control	148	0.0	2.5	12.0	24.0	184.0	19.8	26.4
HER2 +	Case	47	0.0	5.0	13.0	40.0	92.0	25.5	25.6
	Control	28	1.0	8.0	21.0	52.0	121.0	34.9	34.2
Triple Negative	Case	56	0.0	5.0	15.5	32.0	116.0	22.7	23.5
	Control	42	0.0	3.0	13.5	42.0	137.0	28.8	33.2

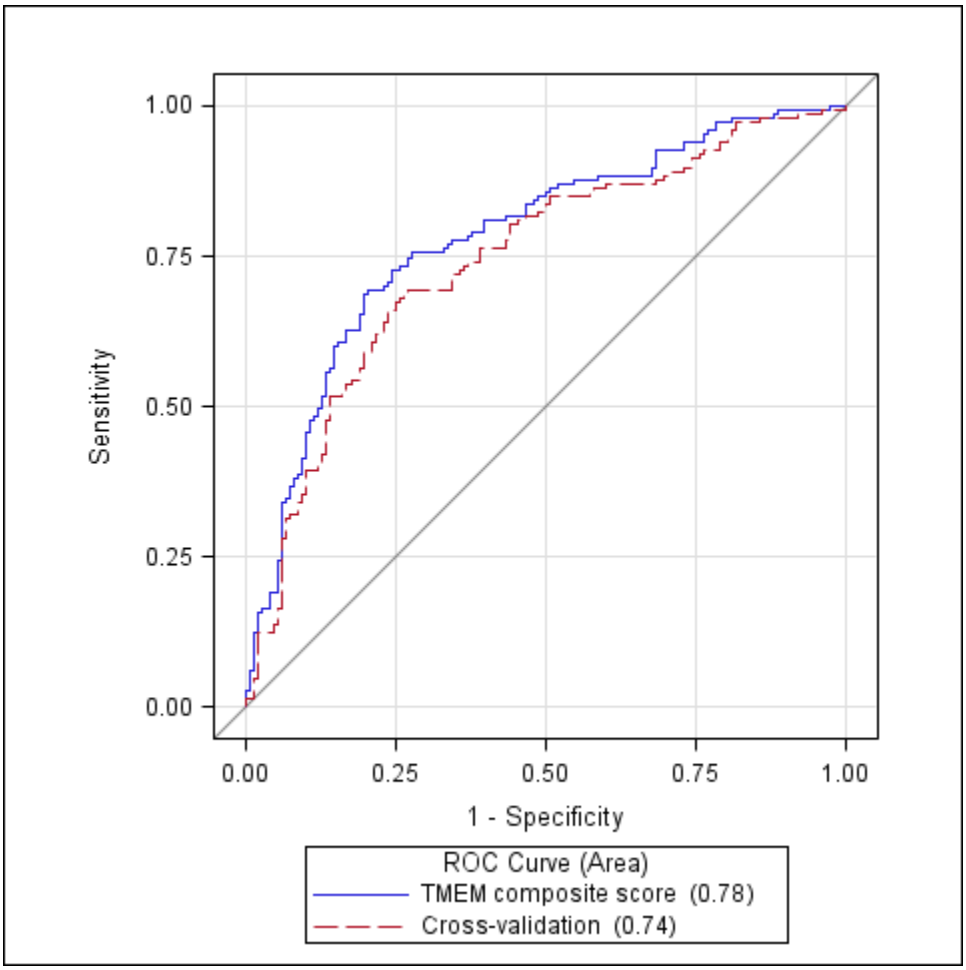
Supplementary Table 3

Summary statistics for IHC4 score by case-control status, overall and by breast cancer subgroup

Tumor type	Status	N Obs	Minimum	25th Pctl	50th Pctl	75th Pctl	Maximum	Mean	Std Dev
All subjects	Case	259	-97.1	9.4	72.8	125.5	204.4	67.7	71.8
	Control	259	-97.7	-18.7	22.9	114.0	204.4	41.6	74.6
ER +/ HER2 -	Case	147	-97.1	-19.3	17.6	58.9	125.8	18.8	49.7
	Control	148	-97.7	-29.9	-5.8	26.3	150.7	-0.3	45.2
HER2 +	Case	47	55.5	114.0	160.4	181.1	204.4	150.1	38.1
	Control	28	79.5	112.9	144.9	169.5	204.4	142.0	37.0
Triple negative	Case	56	54.5	120.5	129.7	141.3	154.6	126.6	21.4
	Control	42	54.5	120.5	136.2	148.9	152.0	131.2	23.3

Supplementary Figure 2

ROC curves for TMEM composite score and cross-validated predicted probabilities in ER+/HER2- subgroup



Supplementary Methods 3

Estimation of the absolute risk of metastasis in low, medium and high risk groups defined by the TMEM composite score

The TMEM composite score was derived from a multivariate logistic regression model with TMEM and the clinical variables in the ER+/HER2- breast cancer subgroup. This yielded a linear equation with the estimated coefficients from the multivariate logistic regression as weights. Specifically:

TMEM composite score = -.18818 + 0.0147TMEM score + 1.1356 x I(2cm<Tumor Size≤3cm) + 1.2697 x I(3cm< Tumor Size) + 0.6630 x I(Tumor size missing) + 0.4570 x (Poor tumor grade) + 1.7358 x I(Lymph node number≥4) + 0.5215 x I(Hormone therapy) + 0.0010 x Age at diagnosis + 1.004 x Years of follow-up – 0.0396 x I(Diagnosis year before 1988).

Using the estimates of sensitivity and the specificity obtained from our ROC analyses, we define the low risk group to be those with composite scores ≤ -0.94 (sensitivity=0.90) and the high risk group to be those with composite scores ≥ 0.70 (specificity=0.90). Let P_D be the proportion of breast cancer patients who developed distant metastasis by the end of the follow-up. We estimate that P_D for the entire cohort, which has a median duration of follow-up of 11.8 years, is 14.1% (i.e., 530/3760 (see Fig. 1)). Furthermore, we assume that the proportion of the ER+/HER2- subgroup (representing about 60% of all cases) that developed a distant metastasis is the same as that for the entire cohort. The proportion of individuals falling into the putative low risk group (i.e., composite score ≤-0.94) can be calculated as follows:

(1-sensitivity at the low risk group cut point) *P_D + specificity at the low risk cut point*(1-P_D)=28.9%.

Similarly, the proportion of individuals falling into the high risk group (i.e., composite score≥0.70) can be calculated as the follows:

Sensitivity at the high risk cut point *P_D + (1-specificity at the high risk cut point)*(1- P_D) =14.7%.

The proportion of individuals who fall into the medium risk group (i.e., -0.94<composite score<0.70) is then 100%-28.9%-14.7%=56.4%.

As indicated above, we further assume that proportion of the medium risk group that develops distant metastasis is close to that for the entire cohort. Let P_low risk, P_med risk and P_high risk denote the metastasis proportion for the low risk, medium risk, and high risk groups, respectively. Then P_med risk≈14.1%. In addition, since women in the high risk group and women in the low risk group differ in their composite risk scores by at least 0.70-(-0.94)=1.64, we conservatively assume this difference in risk score when calculating the relative risk of distant metastasis between the high and low risk groups. This relative risk is then approximated by the OR between the two risk groups, i.e, exp(1.64). Therefore, solving the following equation: P_low risk *28.9%+P_med risk *56.4% +P_high risk*14.7% = 14.1%, where P_high risk ≈ exp(1.64)*P_low risk, we find that P_low risk=5.9% and P_high risk=30.3%.

Our study was conducted as a case-control study, which therefore precluded direct estimation of the absolute risk of metastasis. For this, a prospective study would have been required. Nevertheless, we attempted to estimate the absolute risk of metastasis for low, medium, and high risk groups by making an assumption about the absolute risk for the ER+/HER2- subpopulation and a conservative assumption about the average difference in TMEM composite risk score between the high risk and low risk groups. For the same reason, the confidence intervals (CI) of the absolute risk estimates cannot be calculated directly. In place of this, a simulation study was conducted to compute empirical CI for the estimates of absolute risk by taking into consideration the variation in the estimation of the absolute risk for the subpopulation

as well as the variation in estimating the sensitivity and specificity of the TMEM composite score associated with specific cut-points of the score. Using this approach, the 95% CI for the risk of distant metastasis for the low risk group was estimated to be 5.1% - 6.9%, and that for the high risk group was estimated to be 26.1% - 35.4%. The risk of distant metastasis for the medium risk group was based on the assumption that it is same as that for the entire cohort and therefore we adopted the 95% CI for the risk of an event for the entire cohort, namely 13.0% to 15.2%.