INTERNATIONAL JOURNAL of CANCER

Clinical and molecular characteristics of estrogen receptor-positive ultralow risk breast cancer tumors identified by the 70-gene signature

Annelie Johansson[1] | Nancy Y. Yu[2] | Adina Iftimi[2] | Nicholas P. Tobin[1] | Laura van 't Veer[3,4] | Bo Nordenskjöld[5] | Christopher C. Benz[6,7] | Tommy Fornander[1] | Gizeh Perez-Tenorio[5] | Olle Stål[5] | Laura J. Esserman[8] | Christina Yau[7,8] | Linda S. Lindström[1]

[1]Department of Oncology and Pathology, Karolinska Institutet and University Hospital, Stockholm, Sweden

[2]Department of Biosciences and Nutrition, Karolinska Institutet, Stockholm, Sweden

[3]Department of Laboratory Medicine, University of California San Francisco, San Francisco, California, USA

[4]Department of Pathology, University of California San Francisco, San Francisco, California, USA

[5]Department of Biomedical and Clinical Sciences and Department of Oncology, Linköping University, Linköping, Sweden

[6]Department of Medicine, University of California San Francisco, San Francisco, California, USA

[7]Buck Institute for Research on Aging, Novato, California, USA

[8]Department of Surgery, University of California San Francisco, San Francisco, California, USA

Correspondence
Annelie Johansson, Department of Oncology and Pathology, Karolinska Institutet and University Hospital, BioClinicum, Visionsgatan 4, 17164 Stockholm, Sweden.
Email: annelie.johansson@ki.se

Funding information
This work was supported by the Swedish Research Council (Vetenskapsrådet, grant no: 2020-02466 to Linda S. Lindström); the Swedish Research Council for Health, Working life and Welfare (FORTE, grant no: 2019-00477 to Linda S. Lindström); the Gösta Milton Donation Fund (Stiftelsen Gösta Miltons donationsfond, to Linda S. Lindström); the Swedish Cancer Society (Cancerfonden, grant no: 180385 to Olle Stål and 190140 to Linda S. Lindström); Stockholm Cancer Society (Cancerföreningen i Stockholm, to Tommy Fornander); the California Breast Cancer Research Program (CBCRP, grant no: 180B-0065 to Laura J. Esserman); and the National Institutes of Health (NIH, grant no: U01-CA196406 to Laura J. Esserman). The funding sources had no role in the design and conduct of the study;

Abstract

The metastatic potential of estrogen receptor (ER)-positive breast cancers is heterogeneous and distant recurrences occur months to decades after primary diagnosis. We have previously shown that patients with tumors classified as ultralow risk by the 70-gene signature have a minimal long-term risk of fatal breast cancer. Here, we evaluate the previously unexplored underlying clinical and molecular characteristics of ultralow risk tumors in 538 ER-positive patients from the Stockholm tamoxifen randomized trial (STO-3). Out of the 98 ultralow risk tumors, 89% were luminal A molecular subtype, whereas 26% of luminal A tumors were of ultralow risk. Compared to other ER-positive tumors, ultralow risk tumors were significantly (Fisher's test, $P < .05$) more likely to be of smaller tumor size, lower grade, progesterone receptor (PR)-positive, human epidermal growth factor 2 (HER2)-negative and have low Ki-67 levels (proliferation-marker). Moreover, ultralow risk tumors showed significantly lower expression scores of multi-gene modules associated with the AKT/mTOR-pathway, proliferation (AURKA), HER2/*ERBB2*-signaling, IGF1-pathway, PTEN-loss and immune response (IMMUNE1 and IMMUNE2) and higher expression scores of the PIK3CA-mutation-associated module. Furthermore, 706 genes were significantly

collection, management, analysis and interpretation of the data; preparation, review or approval of the manuscript; and decision to submit the manuscript for publication.

(FDR < 0.001) differentially expressed in ultralow risk tumors, including lower expression of genes involved in immune response, PI3K/Akt/mTOR-pathway, histones, cell cycle, DNA repair, apoptosis and higher expression of genes coding for epithelial-to-mesenchymal transition and homeobox proteins, among others. In conclusion, ultralow risk tumors, associated with minimal long-term risk of fatal disease, differ from other ER-positive tumors, including luminal A molecular subtype tumors. Identification of these characteristics is important to improve our prediction of nonfatal vs fatal breast cancer.

What's new?

The metastatic potential of estrogen receptor (ER)-positive breast cancers is heterogeneous, and distant recurrences may occur months to decades after primary diagnosis. However, the long-term risk of metastatic disease in breast cancer remains largely unexplored. Using a previously-established 70-gene signature to identify breast cancer patients with minimal long-term risk of fatal disease, here the authors show that ultralow-risk tumors differ from other ER-positive breast cancer tumors, including luminal molecular subtype tumors. Differences were found in both clinical breast cancer markers and molecular features. These results are important for the characterization and prediction of non-fatal vs fatal breast cancer.

1 | INTRODUCTION

Estrogen receptor (ER)-positive breast cancer patients have a steady long-term risk of fatal disease, and distant metastatic recurrence can occur anywhere between a few months to several decades after primary diagnosis.[1-6] Mammographic screening enables detection of early breast cancer and has reduced the disease mortality, but can introduce overdiagnosis of tumors that might never have come to clinical attention.[7,8] Adding molecular risk prediction tools to standard clinical breast cancer markers may improve risk assessment and reduce overtreatment in patients with low risk of metastatic disease.[9] However, to confidently offer less aggressive treatment to patients, a better understanding of the long-term risk of metastatic disease in breast cancer is needed.

The 70-gene signature (MammaPrint) was originally designed to identify breast cancer patients with high or low risk of early relapse within 5 years after primary diagnosis to identify which patients require adjuvant therapy.[10] The signature was developed in lymph node-negative patients under the age of 55, but has shown to also be prognostic in lymph node-positive and older patients, and up to 25 years after primary diagnosis.[11-13] Genes included in the signature and upregulated in high-risk patient tumor samples have been shown to be associated with the cell cycle, invasion and metastasis and angiogenesis.[10] Clinical trials have validated that ER-positive patients of otherwise high clinical risk, but classified as low risk by the 70-gene signature, may not benefit from adjuvant chemotherapy,[14] and this molecular risk prediction tool is implemented in several breast cancer treatment guidelines.[15-17] Moreover, the additional information from the 70-gene signature has been shown to improve clinicians' confidence in their treatment recommendations.[18-20] Furthermore, we

have demonstrated that the "ultralow risk" threshold derived from the 70-gene signature identifies patients with a very low long-term risk of fatal breast cancer. The 20-year disease-specific survival was 97% and 94% for ER-positive lymph-node negative patients randomized to tamoxifen vs no adjuvant therapy, respectively.[9] Consequently, it is important to understand the underlying characteristics of ultralow risk breast cancer tumors, given that the long-term risk in breast cancer is largely unexplored.

Therefore, here we aimed to assess the clinical and molecular characteristics of ultralow risk tumors in ER-positive lymph-node negative postmenopausal breast cancer patients from the Stockholm tamoxifen randomized trial (STO-3). Ultralow risk tumors were compared to other ER-positive tumors, including PAM50 molecular subtype luminal A and B tumors, regarding the clinically used breast cancer markers, as well as the expression scores of 19 multigene modules representative of specific biological processes and pathways. Furthermore, differentially expressed genes on the single-gene level in ultralow risk tumors vs other ER-positive tumors were identified and categorized by their associated Hallmark gene set to better understand the molecular characteristics of ER-positive tumors with very low long-term risk of fatal disease.

2 | MATERIALS AND METHODS

2.1 | Patients

The Stockholm breast cancer study group conducted randomized trials 1976 to 1990 in lymph-node negative postmenopausal

patients with tumors ≤30 mm in diameter.[21,22] The Stockholm tamoxifen trial (STO-3) enrolled 1780 patients that were randomized to 2 years of adjuvant tamoxifen (40 mg daily) vs no adjuvant treatment.[21,22] In 1983, patients who re-consented and were recurrence-free after 2 years of tamoxifen treatment were randomized to 3 additional years of tamoxifen. As a result, patients in the tamoxifen randomization arm were treated with tamoxifen for 2 or 5 years. No significant differences in survival in the comparison of 2 vs 5 years of tamoxifen have been observed in the STO-3 trial.[21]

Molecular analysis was possible for 808 patients with available formalin-fixed paraffin-embedded (FFPE) tissue blocks from the primary breast cancer tumor.[23] The 808 patient subset was well balanced to the original STO-3 trial cohort with regards to tumor characteristics.[23] Eighty-one patients were excluded from analysis due to insufficient invasive tumor cells, leaving 727 samples available for further analysis (Figure 1).

2.2 | Immunohistochemistry

Immunohistochemistry (IHC) analysis of ER, progesterone receptor (PR), human epidermal growth factor 2 (HER2) and proliferation marker Ki-67 were performed 2014.[2] The annotation of whole-tumor
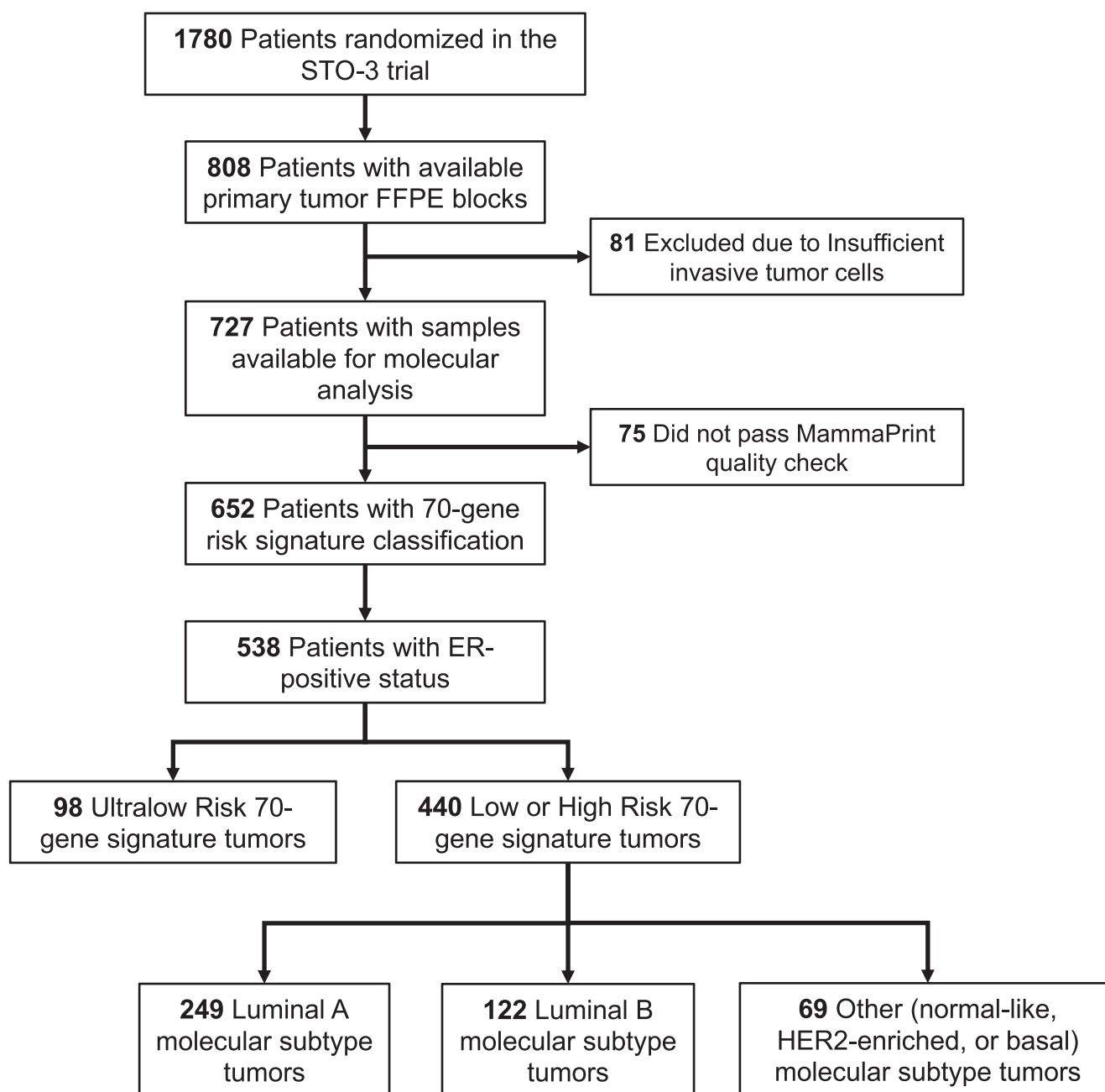

FIGURE 1 Participant flowchart of the Stockholm tamoxifen randomized trial (STO-3)

sections (5 μm) from FFPE tumor blocks was conducted in a random order at a single medical laboratory (University of California Davis Medical Center, UCDMC) following standard recommended procedures. The slides were stained using DAKO Link48 Autostainer with antibodies for ER (SP1; Spring Bioscience M301), PR (PgR 636; DAKO IR068), HER2 (HercepTest; DAKO SK001) and Ki-67 (MIB-1; DAKO M7240). The percentage of cancer cells positive for ER, PR, HER2 and Ki-67 was scored by breast cancer pathologists.[2] ER- and PR-positivity was defined by a threshold of 10% or greater, according to the Swedish National Guidelines,[24] HER2-positivity as intensity 3+ by IHC and Ki-67 was categorized as low (<15%) and medium/high (≥15%).

2.3 | Tumor grade

Tumor grade was retrospectively assessed by one pathologist according to the Nottingham Histologic Score system (Elston grade).[23]

2.4 | Agilent microarray gene expression profiling

Agilent microarray profiling was performed in 2014.[2] Gene expression data were independently generated using custom-designed arrays, Agilent Technologies (Santa Clara, CA), containing approximately 32.1K probes, representing approximately 21.5K unique genes from FFPE breast cancer tumor tissue. Out of 727, 652 breast cancer tumors passed the RNA quality check according to the diagnostic quality model and were used in the analysis, of which 538 were ER-positive (Figure 1).

2.5 | 70-Gene signature risk classification

The 70-gene signature (MammaPrint) was used to classify primary tumors into "high risk", "low risk" or "ultralow risk." The molecular risk prediction tool was originally designed to identify patients with low or high risk of early relapse,[10-13] and the ultralow risk threshold was added to identify patients with indolent tumors associated with minimal long-term risk of fatal disease.[9] From the primary tumor microarray gene expression data, the 70-gene signature risk classification was performed according to standard protocols as previously described, including the use of 465 normalization genes and over 250 probes for hybridization and printing quality control.[11,25] Patient tumor samples were classified into "high risk" (<0), "low (but not ultralow) risk" (>0, <0.355) and "ultralow risk" (≥0.355) using thresholds previously developed.[9,11]

2.6 | Molecular subtyping

Tumors were assigned to one of five molecular subtypes: luminal A, luminal B, HER2-enriched, basal or normal-like, by the PAM50

classification,[26] see details in Supporting Information Materials and Methods.

2.7 | Multigene modules expression scores

The expression scores of 19 multigene modules, proxy-signatures for activation of biological processes or pathways and associated with clinical outcome, were analyzed in the study.[27-29] Each module comprises genes that are positively or negatively associated with its biological process/pathway. Continuous expression scores for each multigene module were calculated for all 652 STO-3 patients with gene expression data using R package Genefu version 2.8.0.[30] The continuous multigene expression scores were categorized to tertiles, which were then converted to two values: the most aggressive tertile (low or high expression scores suggested to be associated with worse prognosis, see details in Supporting Information Materials and Methods) vs the two less aggressive tertiles combined.[27] For example, the highest expression tertile for the proliferation-marker AURKA, associated with higher proliferation and clinical aggressiveness, was compared to the two lower expression tertiles. The multigene module expression scores in tumors classified as "ultralow risk" according to the 70-gene signature were compared to three other groups of tumors classified as either "low risk" or "high risk": (a) all other ER-positive tumors, (b) ER-positive luminal A molecular subtype tumors and (c) ER-positive luminal B molecular subtype tumors.

2.8 | Differentially expressed genes

On a single-gene level, an exploratory analysis of genes differentially expressed in ultralow risk tumors compared to all other ER-positive tumors of low or high risk was conducted. Prior to this, probes from the gene expression microarray data were filtered out if the median expression value was among the bottom 5%, if the variance was among the 25% lowest, if containing any missing values, or if the probe was not annotated to any gene. The differentially expressed genes were categorized by Hallmark gene sets (version 6.2)[31] and Gene Ontology biological processes (C5 collection version 6.2),[32] which were combined into 27 Hallmark gene sets, see Table S1 and Supporting Information Materials and Methods. A heatmap for the differentially expressed genes was produced using Z-scores of normalized gene expressions of all 538 ER-positive tumors, ordered by the 70-gene signature classification, molecular subtype, tumor grade, tumor size, PR status, HER2 status and Ki-67 status.

2.9 | Gene set enrichment analysis

Using the 27 Hallmark gene sets, a gene set enrichment analysis (GSEA) was performed to identify which gene sets are significantly enriched of genes differentially expressed in ultralow risk tumors

compared to other ER-positive tumors of low or high risk. The GSEA calculates enrichment scores (ES) for each gene set to explore if genes belonging to the gene set tend to occur at the top (or bottom) of a specific preranked gene list.[33] Genes were ranked using the t-statistics output from the analysis of differentially expressed genes. A false-discovery rate (FDR) of 5% was used to adjust for multiple testing.

2.10 | Statistical analysis

Given that all ultralow risk tumors were ER-positive, only ER-positive patients were included in our study (n = 538; Figure 1). Fisher's exact test was used to compare ultralow risk tumors to other ER-positive tumors by the clinically used breast cancer markers (ie, tumor size, tumor grade, PR, HER2 and Ki-67) and the 19 multigene modules expression scores. A P-value less than .05 was considered statistically significant. The analysis to identify differentially expressed genes was conducted using R package OCplus[34] with t-statistics and FDR cutoff of 0.001.

All data preparation and analysis were done using R version 3.5.2 and SAS software version 9.4. All statistical tests were two-sided.

3 | RESULTS

A total of 652 patients in the STO-3 trial with 70-gene signature risk classification were available for analysis. The 70-gene signature classified 15% (n = 98) of the tumors as of ultralow risk (Figure 1). Given that all ultralow risk tumors were ER-positive as determined by IHC, the analyses were focused on the 538 patients with ER-positive breast cancer tumors only (Figure 1). In Table 1, patient and tumor characteristics are presented for ultralow risk tumors, all other ER-positive tumors (of low or high risk) and ER-positive luminal A and B molecular subtype tumors (of low or high risk). Noteworthy, a majority (n = 87 out of 98, 89%) of the ultralow risk tumors were of luminal A molecular subtype (Table S2). On the other hand, only 26% (87 out of 336) of the luminal A molecular subtype tumors were classified as ultralow risk. Moreover, four ultralow risk tumors were of luminal B subtype, seven normal-like and none of HER2-enriched or basal molecular subtype.

TABLE 1 Primary patient and tumor characteristics of ER-positive patients in the STO-3 trial

| Primary patient and tumor characteristics | Ultralow risk tumors (n = 98) | ER-positive tumors of low or high risk | | | | | |
| | | All other ER-positive tumors (n = 440) | | Luminal A tumors (n = 249) | | Luminal B tumors (n = 122) | |
	No (%)	No (%)	Pb	No (%)	Pb	No (%)	Pb
Tumor size							
pT ≤ 20 mm	89 (90.8)	348 (80.2)	.013	205 (83.7)	.123	83 (68.0)	<.001
pT > 20 mm	9 (9.2)	86 (19.8)		40 (16.3)		39 (32.0)	
Unknown	0 (—)	6 (—)		4 (—)		0 (—)	
Tumor grade							
1	39 (39.8)	77 (17.8)	<.001	58 (23.5)	<.001	5 (4.2)	<.001
2	59 (60.2)	279 (64.6)		172 (69.6)		74 (62.2)	
3	0 (0.0)	76 (17.6)		17 (6.9)		40 (33.6)	
Unknown	0 (—)	8 (—)		2 (—)		3 (—)	
PR status[a]							
Positive	82 (84.5)	285 (66.0)	<.001	174 (71.6)	.012	74 (61.2)	<.001
Negative	15 (15.5)	147 (34.0)		69 (28.4)		47 (38.8)	
Unknown	1 (—)	8 (—)		6 (—)		1 (—)	
HER2 status[a]							
Negative	98 (100.0)	415 (94.5)	.012	246 (98.8)	.562	117 (95.9)	.067
Positive	0 (0.0)	24 (5.5)		3 (1.2)		5 (4.1)	
Unknown	0 (—)	1 (—)		0 (—)		0 (—)	
Ki-67 status[a]							
Low	89 (95.7)	306 (73.0)	<.001	196 (82.7)	.001	69 (58.5)	<.001
Medium/High	4 (4.3)	113 (27.0)		41 (17.3)		49 (41.5)	
Unknown	5 (—)	21 (—)		12 (—)		4 (—)	

[a]PR-positivity was defined as ≥10%, HER2-positivity as intensity 3+ and Ki-67 was categorized as low (<15%) and medium/high (≥15%).
[b]Fisher's exact test, using ultralow risk tumors as reference. Significant P values (P < .05) are marked in bold.

3.1 | Clinical characteristics of ultralow risk breast cancer tumors

The ultralow risk tumors were significantly ($P < .05$) more likely to be of a smaller (pT ≤ 20 mm) tumor size, except in the comparison of luminal A tumors and of lower tumor grade, as compared to the other groups (Table 1). Furthermore, ultralow risk tumors differed from other ER-positive tumors in the IHC markers by being significantly more likely to be PR-positive, HER2-negative and Ki-67-low (<15%; Table 1). Note, however, that all ultralow risk tumors were of grade 1-2 and HER2-negative, and that HER2-status did not significantly differentiate ultralow risk tumors to luminal A or B molecular subtype tumors.
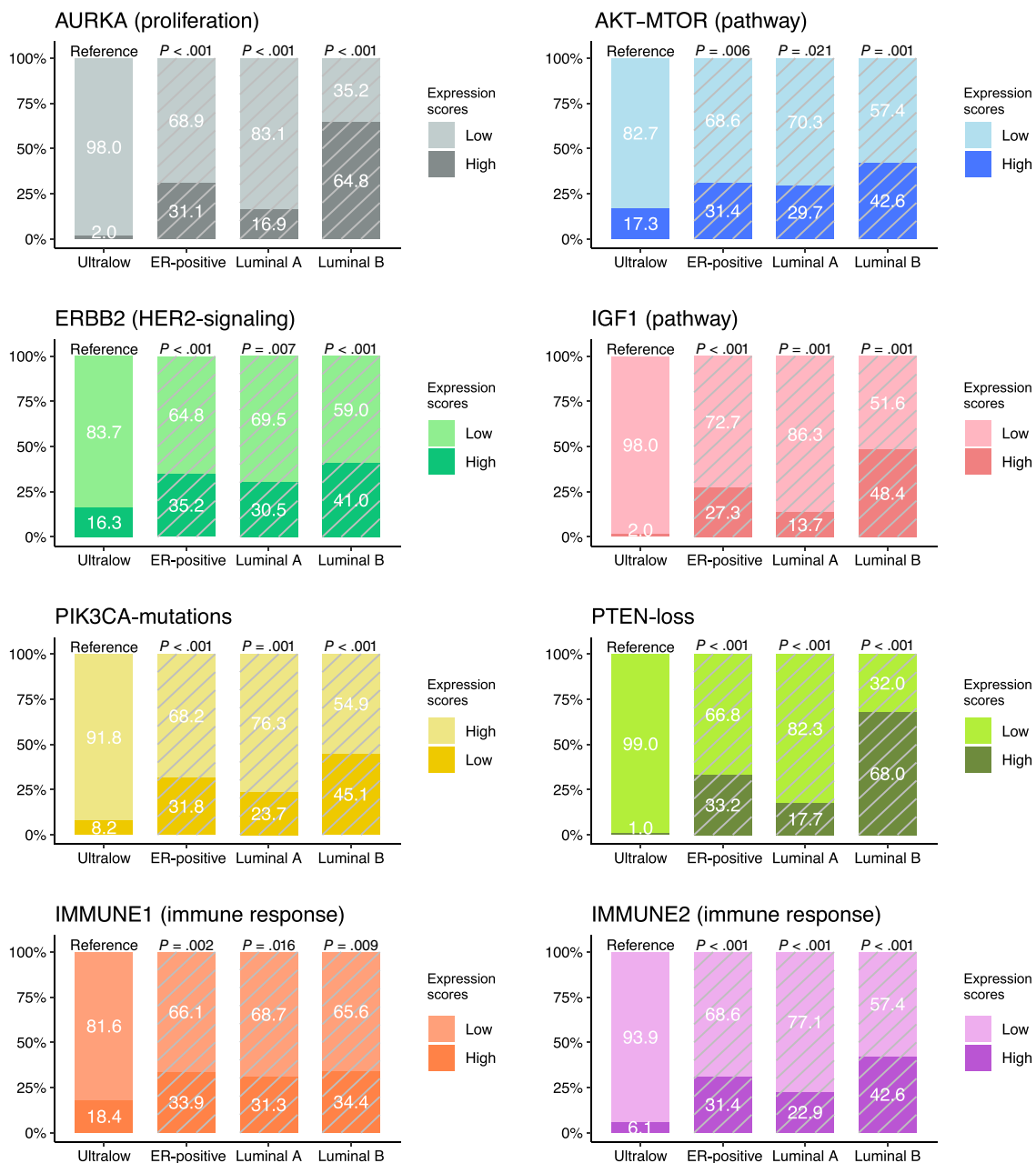

FIGURE 2 Multigene module expression scores in ultralow risk breast cancer tumors. Bar plot of the eight multigene modules with significantly different expression scores in ultralow risk breast cancer tumors compared to all other ER-positive tumors (of low or high risk), and ER-positive luminal A and luminal B molecular subtype tumors (of low or high risk). The plot shows the percentages of tumors with low or high expression scores. The modules included are proliferation-marker AURKA, AKT/mTOR-pathway, HER2/*ERBB2*-signaling, IGF1-pathway, PIK3CA-mutation-associated module, PTEN-loss and immune response modules IMMUNE1 and IMMUNE2. Darker colors represent the category of module-expression scores suggested to be associated with worse prognosis (ie, high AURKA expression scores, but low PIK3CA-mutation expression scores) and diagonal lines indicates significant differences from ultralow risk tumors (reference) by Fisher's exact test ($P < .05$)

3.2 | Multigene module expression scores in ultralow risk breast cancer tumors

In order to further understand the molecular characteristics of ultralow risk tumors, the expression scores of 19 multigene modules (proxy-signatures associated with different biological pathways/processes and clinical outcome) in the ultralow risk tumors were compared to all other ER-positive tumors (of low or high risk), and ER-positive luminal A and B molecular subtype tumors (of low or high risk). Eight multigene modules showed significantly ($P < .05$) different expression scores in ultralow risk tumors (Figure 2). Lower expression scores of the proliferation-marker AURKA, AKT/mTOR-pathway, HER2/*ERBB2*-signaling, IGF1-pathway, PTEN-loss and the immune response-modules IMMUNE1 and IMMUNE2 were observed for the ultralow risk tumors, and higher expression scores were observed for the PIK3CA-mutation-associated module (Figure 2). Moreover, compared to all other ER-positive tumors, but not the luminal molecular subtypes, ultralow risk tumors showed significantly higher expression scores of the *ESR1*-module (Table S3). Furthermore, lower expression scores of the pathway-associated gene-modules MYC and E2F3 were observed in ultralow risk tumors compared to all other ER-positive tumors and luminal B molecular subtype tumors (Table S3). Other multigene modules differentiating ultralow risk tumors from luminal B molecular subtype tumors included lower expression scores of CASP3, associated with apoptosis and higher expression scores of MAPK, STROMA1 and STROMA2, associated with its pathway, tumor invasion/metastasis and the stromal environment, respectively.
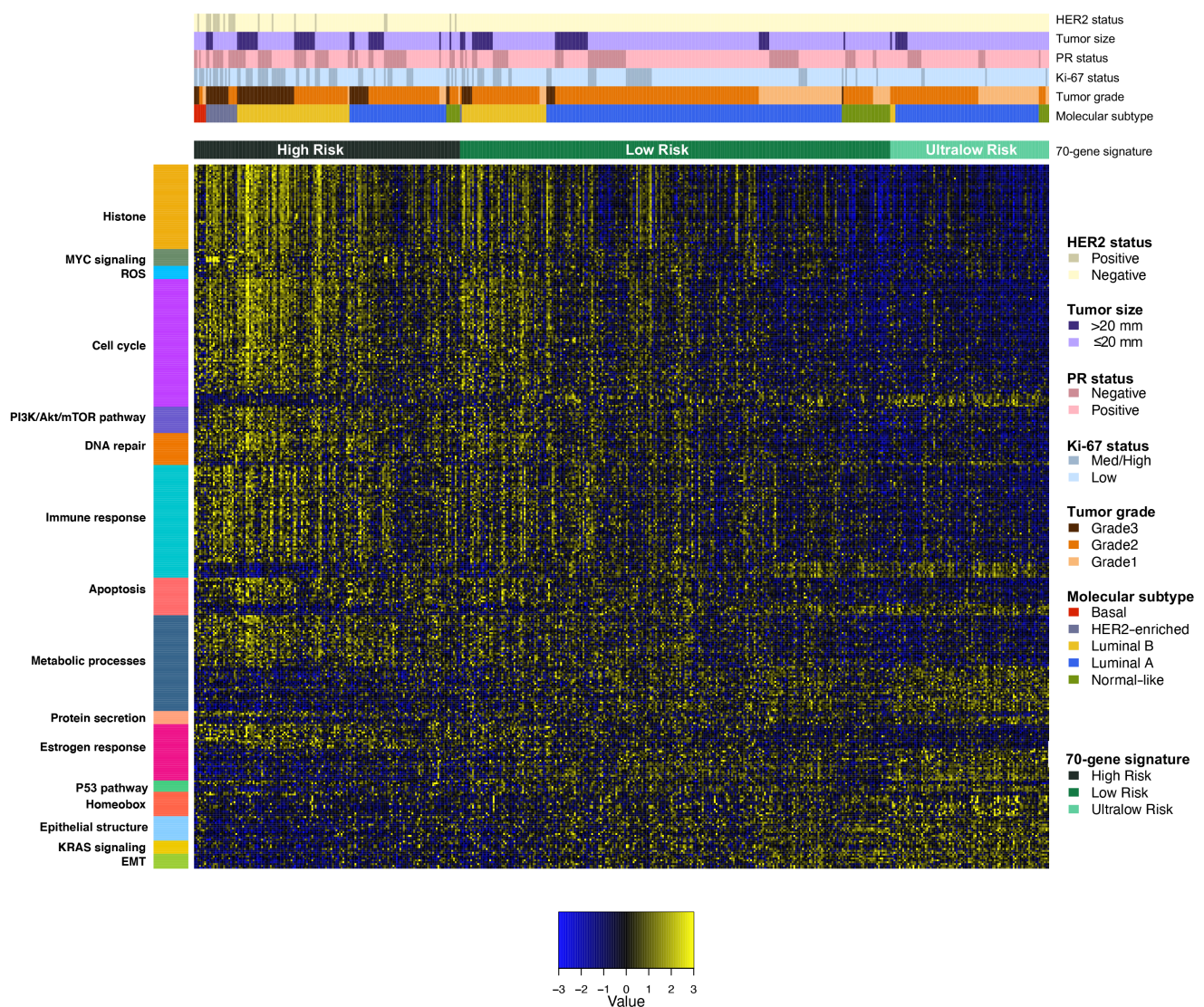

FIGURE 3 Differentially expressed genes in ultralow risk breast cancer tumors. Heatmap of genes differentially expressed between ultralow risk tumors and all other ER-positive tumors (of low or high risk). Patients (columns) are ordered by the 70-gene signature classification, molecular subtype, tumor grade, Ki-67 status, PR status, tumor size and HER2 status. Genes are categorized by their involvement in different biological processes or pathways (Hallmark gene sets), and ordered by *t*-statistics between and within each gene category. EMT, epithelial-to-mesenchymal transition; ROS, reactive oxygen species

3.3 | Differentially expressed genes in ultralow risk breast cancer tumors

To further understand the molecular characteristics of ultralow risk tumors, an analysis on the single-gene level was performed to identify genes differentially expressed in ultralow risk tumors compared to all other ER-positive tumors of low or high risk. Overall, 706 genes were significantly (FDR < 0.001) differentially expressed in ultralow risk tumors as compared to all other ER-positive tumors of low or high risk (Table S4). Of these, 454 genes were expressed at lower levels in ultralow risk tumors and 252 genes at higher expression levels.

To further understand the biological function associated with ultralow risk definition, the 706 differentially expressed genes were categorized into different cancer-related Hallmark gene sets (Table S4). Figure 3 shows a heatmap of a subset of the differentially expressed genes ordered by their Hallmark gene category and with patients ordered by clinical and molecular characteristics. The gene expression patterns in the heatmap were clearly different between the ultralow risk and high risk tumors, whereas the tumors classified as low risk showed a more heterogenous gene expression pattern (Figure 3). Consistent with the previous results of the multigene modules, ultralow risk tumors showed lower expression levels of genes associated with the PI3K/Akt/mTOR pathway and immune response (Figure 3 and Table S4). Moreover, ultralow risk tumors generally expressed lower levels of genes coding for histones, MYC-signaling, reactive oxygen species, cell cycle, DNA repair and apoptosis (Figure 3 and Table S4). Genes coding for homeobox proteins, epithelial structure, KRAS-signaling and epithelial-to-mesenchymal transition (EMT) were expressed in higher levels in ultralow risk tumors. Genes involved in different metabolic processes, protein secretion, estrogen response or P53 pathway showed both higher and lower expression levels in ultralow risk tumors.

Further gene set enrichment analysis (GSEA) showed significant (FDR < 0.05) enrichment of gene sets consistent with previous analyses (Table S5). The gene sets MYC-signaling, cell cycle, DNA-repair, unfolded protein response, PI3K/Akt/mTOR-pathway, immune response and apoptosis were downregulated in ultralow risk tumors, and EMT upregulated. Furthermore, the GSEA also showed the gene sets metabolic processes and P53-pathway to be downregulated in ultralow risk tumors, whereas myogenesis, hedgehog signaling and estrogen response were upregulated.

4 | DISCUSSION

We have previously shown that the ultralow risk threshold of the 70-gene signature identifies patients at minimal long-term risk of death from breast cancer.[9] To further understand the long-term risk of breast cancer, we here aimed to explore the clinical and molecular characteristics of primary ultralow risk tumors from the STO-3 trial. Our study shows that ultralow risk tumors are significantly more likely to be of a smaller tumor size, of lower tumor grade, PR-positive, HER2-negative and Ki-67-low, compared to other ER-positive tumors.

Moreover, ultralow risk tumors exhibited substantially different expressions of "hallmark" gene sets as well as other important cancer-related biological processes or pathways.

The risk of distant recurrences in ER-positive breast cancer remains steady decades after primary diagnosis.[1-6] Consequently, it is important to identify distinct biological characteristics that predict patients' long-term recurrence risk to improve our understanding and distinguish our management of nonfatal vs fatal breast cancers—something that has proven a great challenge. Here, our findings suggest that ultralow risk tumors have differential tumor characteristics as compared to other ER-positive tumors, including luminal A molecular subtype tumors which are generally considered to be of low risk. Noteworthy, far from all luminal A molecular subtype tumors (1 out of 4) are classified as ultralow risk, which indicates that the ultralow risk classification differs from the molecular subtyping. We have previously shown that patients with luminal A tumors have a long-term risk of distant recurrences, which is reduced by tamoxifen therapy.[3] Thus, the luminal A molecular subtype may yet consist of a heterogeneous group of tumors, were the ultralow risk threshold can be used to identify patients with a very low long-term risk. Therefore, the 70-gene signature ultralow risk threshold may be helpful to minimize overtreatment and reassure patients in terms of their low long-term risk of distant recurrences, together with more traditionally used clinical markers,[1] and other prognostic predictors such as urokinase plasminogen activator (uPA) levels and its inhibitor,[35] molecular subtype classification and risk of recurrence (ROR) score[36] and ER-intratumor heterogeneity levels.[2]

Ultralow risk tumors differ significantly from other ER-positive tumors in relation to the clinically used breast cancer markers. PR and Ki-67 have prognostic value in breast cancer,[5,37,38] and we show that ultralow risk tumors are significantly more likely to be PR-positive and Ki-67-low (<15%), as determined by IHC. Together with lower expression of genes involved in the cell cycle and the proliferation-associated multigene module AURKA, this suggests that lower proliferation is an important clinical characteristic of ultralow risk tumors. Interestingly, the expression scores of the multigene module reflecting HER2/ERBB2-signaling significantly differ between the ultralow risk tumors as compared to the other ER-positive tumor groups. However, since this difference was not observed by the HER2 IHC assay (with no ultralow risk tumors categorized as HER2-positive, whereas 16% were categorized as high HER2-expression by the HER2/ERBB2 multigene module) this finding should be confirmed in further studies. Furthermore, ultralow risk tumors exhibited higher expression scores of the ESR1-module, indicating higher ER expression which is associated with better prognosis.[27]

Ultralow risk tumors differ significantly from other ER-positive tumors on the molecular level by the expression of cancer-associated biological processes and pathways. In brief, genes involved in the PI3K/Akt/mTOR pathway, which has a role in endocrine sensitivity and breast cancer survival,[39,40] show lower expression in ultralow risk tumors, as observed both by the multigene module AKT/mTOR and on the single-gene level. Furthermore, loss of PTEN, a PI3K-regulator and tumor suppressor, is more frequent in the ultralow risk tumors,

and PIK3CA-mutations, which have been associated with more favorable outcome in ER-positive breast cancer,[41] are more common. Moreover, the difference in expression in transcriptionally important histone genes and homeobox genes implicate a potentially altered transcriptional and epigenetic state in ultralow risk tumors.[42]

Lower expression of immune response genes is recognized in ultralow risk tumors, as observed by the multigene modules IMMUNE1 and IMMUNE2 and on the single-gene level. While the immune response mostly has been investigated in the context of ER-negative or HER2-positive breast cancer were higher expression often has been associated with better survival and reduced metastatic risk,[28,43] it has also been associated with better pathologic complete response (pCR) in ER-positive/HER2-negative patients.[29] Contrary to this, our observed lower expression of immune response genes in ultralow risk tumors is a new and clinically provocative finding that deserves further validation.

Another finding in our study indicate that ultralow risk tumors exhibit lower expression scores of the IGF1-pathway-associated module, which is involved in breast cancer development and progression.[44,45] Interestingly yet relatively small gene groups identified as differentially expressed in ultralow risk tumors include MYC, reactive oxygen species, P53, KRAS and EMT. These, as well as genes involved in metabolic processes and estrogen response, which showed both lower and higher expression levels in ultralow risk tumors, require further analyses to understand their role in relation to long-term metastatic risk. Noteworthy, even though higher expression of EMT genes might be a sign of metastatic development,[46] a majority of the differentially expressed genes in ultralow risk tumors associated with EMT have been recognized as tumor-suppressor genes, including *TPM1*, *WNT5A*, *TGFBR3*, *SLIT3* and *PDLIM4*.[47-51]

To summarize, our study suggests an ultralow risk phenotype characterized by clinically used breast cancer markers (tumor size, tumor grade, PR status, HER2 status and Ki-67 status) and molecular characteristics. Importantly, the ultralow risk classification differs from the luminal A molecular subtyping. Furthermore, ultralow risk tumors show lower activation of the PI3K/Akt/mTOR-pathway, immune response and IGF1-pathway and lower expression levels of genes involved in proliferation, cell cycle, DNA repair and apoptosis and higher expression levels of EMT tumor suppressor genes. Also, an altered transcriptional state in ultralow risk tumors is suggested, including lower expression levels of genes coding for histones and higher expression levels of homeobox genes, which less is known about. These findings are supported by analysis of multigene expression modules and the identification of differentially expressed genes and their associated Hallmark gene sets.

Our study was conducted using tumors from postmenopausal patients with generally low metastatic risks given their lymph-node negative status and smaller (≤30 mm) tumors. There are limitations to our study. The STO-3 trial was conducted mainly prior to the introduction of mammographic screening in Sweden; therefore, the majority of the breast tumors were clinically detected. Modern mammographic screening has increased the number of newly diagnosed low-risk breast tumors, thus, a better clinical understanding of

low-risk tumors characteristics is needed. As with most long-term follow-up studies, clinical recommendations for disease management and treatment have changed since the trial. Molecular analysis was possible for approximately half of the tumors in STO-3; however, we have confirmed that patient and tumor characteristics were well balanced to the original STO-3 cohort.[23] Further, when dealing with IHC assays, there is often some degree of subjective inaccuracy. However, the clinical IHC markers were recently re-assessed at a single medical laboratory by dedicated breast cancer pathologists.[52] Despite our relatively small study size with 98 ultralow risk tumors, we were able to identify significant and informative differences between the analyzed groups. Another clear strength of our study includes the recent performance and annotation of genome-wide gene expression analyses.

5 | CONCLUSIONS

Ultralow risk tumors, associated with minimal long-term risk of fatal breast cancer, have distinct clinical, biological and molecular characteristics as compared to other ER-positive tumors, including luminal A molecular subtype tumors that generally are considered as a low-risk disease. A better understanding of the characteristics of breast cancer tumors of very low long-term risk of metastatic disease is important to improve our prediction of fatal vs nonfatal disease.

CONFLICT OF INTEREST
Dr van 't Veer is a cofounder, stockholder and part-time employee of Agendia. All remaining authors declare no potential conflicts of interest.

AUTHOR CONTRIBUTIONS
Conceptualization: Annelie Johansson, Nancy Yiu-Lin Yu, Adina Iftimi, Laura van 't Veer, Christopher C. Benz, Laura J. Esserman, Christina Yau and Linda S. Lindström. Data curation: Bo Nordenskjöld, Tommy Fornander, Olle Stål, Laura J. Esserman and Linda S. Lindström. Formal analysis: Annelie Johansson, Nancy Yiu-Lin Yu, Adina Iftimi, Christina Yau and Linda S. Lindström. Funding acquisition: Tommy Fornander, Olle Stål, Laura J. Esserman and Linda S. Lindström. Investigation: Annelie Johansson, Nancy Yiu-Lin Yu, Adina Iftimi, Gizeh Perez-Tenorio, Christina Yau and Linda S. Lindström. Methodology: Annelie Johansson, Nancy Yiu-Lin Yu, Adina Iftimi, Christina Yau and Linda S. Lindström. Project administration: Annelie Johansson and Linda S. Lindström. Resources: Bo Nordenskjöld, Tommy Fornander, Olle Stål and Linda S. Lindström. Software: Annelie Johansson, Nancy Yiu-Lin Yu, Adina Iftimi, Nicholas P. Tobin, Christina Yau and Linda S. Lindström. Supervision: Annelie Johansson and Linda S. Lindström. Visualization: Annelie Johansson, Nancy Yiu-Lin Yu, Adina Iftimi, Gizeh Perez-Tenorio, Christina Yau and Linda S. Lindström.

Writing - original draft: Annelie Johansson, Nancy Yiu-Lin Yu and Linda S. Lindström. Writing - review & editing: All authors.

ETHICS STATEMENT

The STO-3 trial was approved by the ethics committee at Karolinska Institutet and participants provided oral consent. The trial was conducted at the Regional Cancer Center Stockholm-Gotland, Sweden, and began in 1976, which was well before trial registration started in Sweden; therefore, information on trial registration number is not available.

DATA AVAILABILITY STATEMENT

The raw RNA microarray data generated in our study is deposited on a secure Swedish server and has been assigned a DOI (https://doi.org/10.5878/3vxa-3c28). Data access requests may be submitted to the Swedish National Data Service through the DOI link, and will be reviewed by the STO Trialists Group. The R-code to reproduce the results and figures are publicly available at https://github.com/annelieewa/Ultralow_risk_STO3. Other data that support the findings of our study are available from the corresponding author upon request and with the permission of the STO Trialists Group.

ORCID

Annelie Johansson https://orcid.org/0000-0001-9029-1996

REFERENCES

1. 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.
2. Lindström LS, Yau C, Czene K, et al. Intratumor heterogeneity of the estrogen receptor and the long-term risk of fatal breast cancer. *J Natl Cancer Inst*. 2018;110(7):726-733.
3. Yu NY, Iftimi A, Yau C, et al. Assessment of long-term distant recurrence-free survival associated with tamoxifen therapy in postmenopausal patients with luminal A or luminal B breast cancer. *JAMA Oncol*. 2019;5:1304-1309.
4. Colleoni M, Sun Z, Price KN, et al. Annual Hazard rates of recurrence for breast cancer during 24 years of follow-up: results from the international breast cancer study group trials I to V. *J Clin Oncol*. 2016;34(9):927-935.
5. Davies C, Godwin J, Gray R, et al. Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. *Lancet*. 2011;378(9793):771-784.
6. Richman J, Dowsett M. Beyond 5 years: enduring risk of recurrence in oestrogen receptor-positive breast cancer. *Nat Rev Clin Oncol*. 2019;16(5):296-311.
7. Nyström L, Andersson I, Bjurstam N, Frisell J, Nordenskjöld B, Rutqvist LE. Long-term effects of mammography screening: updated overview of the Swedish randomised trials. *Lancet*. 2002;359(9310):909-919.
8. Pace LE, Keating NL. A systematic assessment of benefits and risks to guide breast cancer screening decisions. *JAMA*. 2014;311(13):1327-1335.
9. Esserman LJ, Yau C, Thompson CK, et al. Use of molecular tools to identify patients with indolent breast cancers with ultralow risk over 2 decades. *JAMA Oncol*. 2017;3(11):1503-1510.
10. van 't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*. 2002;415(6871):530-536.

11. Drukker CA, van Tinteren H, Schmidt MK, et al. Long-term impact of the 70-gene signature on breast cancer outcome. *Breast Cancer Res Treat*. 2014;143(3):587-592.
12. van 't Veer LJ, Yau C, Yu NY, et al. Tamoxifen therapy benefit for patients with 70-gene signature high and low risk. *Breast Cancer Res Treat*. 2017;166(2):593-601.
13. van de Vijver MJ, He YD, van 't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med*. 2002;347(25):1999-2009.
14. Cardoso F, van 't Veer LJ, Bogaerts J, et al. 70-gene signature as an aid to treatment decisions in early-stage breast cancer. *N Engl J Med*. 2016;375(8):717-729.
15. Cardoso F, Kyriakides S, Ohno S, et al. Early breast cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up†. *Ann Oncol*. 2019;30(8):1194-1220.
16. Duffy MJ, Harbeck N, Nap M, et al. Clinical use of biomarkers in breast cancer: updated guidelines from the European group on tumor markers (EGTM). *Eur J Cancer*. 2017;75:284-298.
17. Krop I, Ismaila N, Andre F, et al. Use of biomarkers to guide decisions on adjuvant systemic therapy for women with early-stage invasive breast cancer: American Society of Clinical Oncology clinical practice guideline focused update. *J Clin Oncol*. 2017;35(24):2838-2847.
18. Soliman H, Shah V, Srkalovic G, et al. MammaPrint guides treatment decisions in breast cancer: results of the IMPACt trial. *BMC Cancer*. 2020;20(1):81.
19. Wuerstlein R, Kates R, Gluz O, et al. Strong impact of MammaPrint and BluePrint on treatment decisions in luminal early breast cancer: results of the WSG-PRIMe study. *Breast Cancer Res Treat*. 2019;175(2):389-399.
20. Tsai M, Lo S, Audeh W, et al. Association of 70-gene signature assay findings with Physicians' treatment guidance for patients with early breast cancer classified as intermediate risk by the 21-gene assay. *JAMA Oncol*. 2018;4(1):e173470.
21. Rutqvist LE, Johansson H. Long-term follow-up of the randomized Stockholm trial on adjuvant tamoxifen among postmenopausal patients with early stage breast cancer. *Acta Oncol*. 2007;46(2):133-145.
22. Fornander T, Rutqvist LE, Cedermark B, et al. Adjuvant tamoxifen in early breast cancer: occurrence of new primary cancers. *Lancet (London, England)*. 1989;1(8630):117-120.
23. Jerevall PL, Ma XJ, Li H, et al. Prognostic utility of HOXB13:IL17BR and molecular grade index in early-stage breast cancer patients from the Stockholm trial. *Br J Cancer*. 2011;104(11):1762-1769.
24. Svenska bröstcancergruppen (Swedish Breast Cancer Group). Nationella riktlinjer för behandling av bröstcancer; 2012.
25. Glas AM, Floore A, Delahaye LJ, et al. Converting a breast cancer microarray signature into a high-throughput diagnostic test. *BMC Genomics*. 2006;7:278.
26. Parker JS, Mullins M, Cheang MC, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol*. 2009;27(8):1160-1167.
27. Tobin NP, Harrell JC, Lovrot J, et al. Molecular subtype and tumor characteristics of breast cancer metastases as assessed by gene expression significantly influence patient post-relapse survival. *Ann Oncol*. 2015;26(1):81-88.
28. Desmedt C, Haibe-Kains B, Wirapati P, et al. Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes. *Clin Cancer Res*. 2008;14(16):5158-5165.
29. Ignatiadis M, Singhal SK, Desmedt C, et al. Gene modules and response to neoadjuvant chemotherapy in breast cancer subtypes: a pooled analysis. *J Clin Oncol*. 2012;30(16):1996-2004.
30. Gendoo DMA, Ratanasirigulchai N, Schröder MS, Paré L, Parker JS, Prat A, Haibe-Kains B. Genefu: an R/Bioconductor package for computation of gene expression-based signatures in breast cancer. *Bioinformatics*. 2016;32(7):1097-1099. https://doi.org/10.1093/bioinformatics/btv693

31. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Syst*. 2015;1(6):417-425. https://doi.org/10.1016/j.cels.2015.12.004

32. The Gene Ontology Consortium. The gene ontology resource: 20 years and still GOing strong. *Nucleic Acids Res*. 2019;47(D1): D330-d338.

33. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*. 2005;102(43):15545-15550.

34. Pawitan YPA. *OCplus: operating characteristics plus sample size and local fdr for microarray experiments. R Package Version 1600*. Vienna, Austria: R Foundation for Statistical Computing; 2019.

35. Duffy MJ, McGowan PM, Harbeck N, et al. uPA and PAI-1 as biomarkers in breast cancer: validated for clinical use in level-of-evidence-1 studies. *Breast Cancer Res*. 2014;16(4):428.

36. Ohnstad HO, Borgen E, Falk RS, et al. Prognostic value of PAM50 and risk of recurrence score in patients with early-stage breast cancer with long-term follow-up. *Breast Cancer Res*. 2017;19(1):120.

37. Nordenskjöld A, Fohlin H, Fornander T, Löfdahl B, Skoog L, Stål O. Progesterone receptor positivity is a predictor of long-term benefit from adjuvant tamoxifen treatment of estrogen receptor positive breast cancer. *Breast Cancer Res Treat*. 2016;160(2):313-322.

38. Colozza M, Azambuja E, Cardoso F, Sotiriou C, Larsimont D, Piccart MJ. Proliferative markers as prognostic and predictive tools in early breast cancer: where are we now? *Ann Oncol*. 2005;16(11): 1723-1739.

39. Paplomata E, O'Regan R. The PI3K/AKT/mTOR pathway in breast cancer: targets, trials and biomarkers. *Ther Adv Med Oncol*. 2014;6(4): 154-166.

40. Saal LH, Johansson P, Holm K, et al. Poor prognosis in carcinoma is associated with a gene expression signature of aberrant PTEN tumor suppressor pathway activity. *Proc Natl Acad Sci USA*. 2007;104(18): 7564-7569.

41. Loi S, Haibe-Kains B, Majjaj S, et al. PIK3CA mutations associated with gene signature of low mTORC1 signaling and better outcomes in estrogen receptor-positive breast cancer. *Proc Natl Acad Sci USA*. 2010;107(22):10208-10213.

42. Jovanovic J, Rønneberg JA, Tost J, Kristensen V. The epigenetics of breast cancer. *Mol Oncol*. 2010;4(3):242-254.

43. Teschendorff AE, Miremadi A, Pinder SE, Ellis IO, Caldas C. An immune response gene expression module identifies a good prognosis subtype in estrogen receptor negative breast cancer. *Genome Biol*. 2007;8(8):R157.

44. Creighton CJ, Casa A, Lazard Z, et al. Insulin-like growth factor-I activates gene transcription programs strongly associated with poor breast cancer prognosis. *J Clin Oncol*. 2008;26(25):4078-4085.

45. Christopoulos PF, Msaouel P, Koutsilieris M. The role of the insulin-like growth factor-1 system in breast cancer. *Mol Cancer*. 2015;14:43.

46. Savci-Heijink CD, Halfwerk H, Hooijer GKJ, et al. Epithelial-to-mesenchymal transition status of primary breast carcinomas and its correlation with metastatic behavior. *Breast Cancer Res Treat*. 2019; 174(3):649-659.

47. Bharadwaj S, Prasad GL. Tropomyosin-1, a novel suppressor of cellular transformation is downregulated by promoter methylation in cancer cells. *Cancer Lett*. 2002;183(2):205-213.

48. Dickinson RE, Dallol A, Bieche I, et al. Epigenetic inactivation of SLIT3 and SLIT1 genes in human cancers. *Br J Cancer*. 2004;91(12):2071-2078.

49. Dong M, How T, Kirkbride KC, et al. The type III TGF-beta receptor suppresses breast cancer progression. *J Clin Invest*. 2007;117(1):206-217.

50. Jonsson M, Dejmek J, Bendahl PO, et al. Loss of Wnt-5a protein is associated with early relapse in invasive ductal breast carcinomas. *Cancer Res*. 2002;62(2):409-416.

51. Vanaja DK, Ballman KV, Morlan BW, et al. PDLIM4 repression by hypermethylation as a potential biomarker for prostate cancer. *Clin Cancer Res*. 2006;12(4):1128-1136.

52. Engelberg JA, Retallack H, Balassanian R, et al. "Score the Core" web-based pathologist training tool improves the accuracy of breast cancer IHC4 scoring. *Hum Pathol*. 2015;46(11):1694-1704.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.