

# A Novel 18-Marker Panel Predicting Clinical Outcome in Breast Cancer

Jana Biermann<sup>1</sup>, Szilárd Nemes<sup>2</sup>, Toshima Z. Parris<sup>1</sup>, Hanna Engqvist<sup>1</sup>, Elisabeth Werner Rönnerman<sup>1,3</sup>, Eva Forssell-Aronsson<sup>4</sup>, Gunnar Steineck<sup>1</sup>, Per Karlsson<sup>1</sup>, and Khalil Helou<sup>1</sup>



## Abstract

**Background:** Gene expression profiling has made considerable contributions to our understanding of cancer biology and clinical care. This study describes a novel gene expression signature for breast cancer-specific survival that was validated using external datasets.

**Methods:** Gene expression signatures for invasive breast carcinomas (mainly luminal B subtype) corresponding to 136 patients were analyzed using Cox regression, and the effect of each gene on disease-specific survival (DSS) was estimated. Iterative Bayesian model averaging was applied on multivariable Cox regression models resulting in an 18-marker panel, which was validated using three external validation datasets. The 18 genes were analyzed for common pathways and functions using the Ingenuity Pathway Analysis software. This study complied with the REMARK criteria.

**Results:** The 18-gene multivariable model showed a high predictive power for DSS in the training and validation

cohort and a clear stratification between high- and low-risk patients. The differentially expressed genes were predominantly involved in biological processes such as cell cycle, DNA replication, recombination, and repair. Furthermore, the majority of the 18 genes were found to play a pivotal role in cancer.

**Conclusions:** Our findings demonstrated that the 18 molecular markers were strong predictors of breast cancer-specific mortality. The stable time-dependent area under the ROC curve function (AUC(t)) and high C-indices in the training and validation cohorts were further improved by fitting a combined model consisting of the 18-marker panel and established clinical markers.

**Impact:** Our work supports the applicability of this 18-marker panel to improve clinical outcome prediction for breast cancer patients. *Cancer Epidemiol Biomarkers Prev*; 26(11); 1619–28. ©2017 AACR.

## Introduction

Microarray-based gene expression profiling has demonstrated the heterogeneous nature of breast cancer, which is currently known to be composed of several distinct biological and clinical subtypes (1–4). Unfortunately, the genetic complexity of breast cancer poses a major obstacle in the development of improved treatment regimens. Therefore, the challenge lies in the identification and selection of relevant genes that contribute to breast tumorigenesis via complex gene interactions and their effect on treatment efficacy and patient survival.

Individual biomarkers usually have very little statistical power. Therefore, the current approach is to identify novel molecular signatures consisting of several individual genes, as a gene signature should (i) be robust by including key gene regulators in important signaling pathways (redundancy), (ii) reflect the complex heterogeneity of the specific cancer, and (iii) subsequently offer a better prediction of clinical outcome than current prognostic markers (5). The established clinical markers for breast cancer are primarily based on patient- and tumor-related factors, such as patient age at diagnosis, histologic grade, number of positive axillary lymph nodes, pathologic tumor size, and the status of molecular tumor markers [HER2, progesterone receptor (PR), and estrogen receptor (ER)]. Molecular techniques have the potential to refine breast cancer classification and improve treatment procedures (6, 7). Patients would benefit from a more accurate tool to predict clinical outcome, such as sets of markers that may lead to treatment tailoring as well as the development of new therapeutic agents.

Strategies to develop novel prognostic models have frequently followed four basic steps: selection of different covariates (clinical parameters and/or gene expression data) based on an underlying statistical method (Cox regression), development of a prognostic index (linear predictor) that stratifies the patients into different risk groups, assessment of model performance [concordance index (C-index), time-dependent area under the ROC curve function (AUC(t)), etc.], and validation (8). The majority of prognostic models for cancer have been developed with Cox regression resulting in a prognostic index, which can (i) consist

<sup>1</sup>Department of Oncology, Institute of Clinical Sciences, Sahlgrenska Cancer Center, Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden. <sup>2</sup>Swedish Hip Arthroplasty Register, Gothenburg, Sweden. <sup>3</sup>Department of Clinical Pathology and Genetics, Sahlgrenska University Hospital, Gothenburg, Sweden. <sup>4</sup>Department of Radiation Physics, Institute of Clinical Sciences, Sahlgrenska Cancer Center, Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden.

**Note:** Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (<http://cebp.aacrjournals.org/>).

P. Karlsson and K. Helou contributed equally to this article.

**Corresponding Author:** Jana Biermann, Department of Oncology, Sahlgrenska Cancer Center, University of Gothenburg, Medicinaregatan 1G, Gothenburg 41390, Sweden. Phone: (+46) 31 7866751; E-mail: jana.biermann@gu.se

**doi:** 10.1158/1055-9965.EPI-17-0606

©2017 American Association for Cancer Research.

of the same covariates and Cox coefficients in the training and validation cohorts or (ii) use the same covariates but not the same Cox coefficients for both cohorts. The most common measure of discrimination is the C-index along with Kaplan–Meier estimates to visualize the stratification. Validation using external datasets provides a stronger validation than internal validation (e.g., bootstrapping).

In the current study, gene expression profiling was used to identify a novel 18-marker prognostic gene signature for breast cancer by using Cox regression and patient stratification based on the linear predictor. The proposed gene signature was then validated using three external validation cohorts.

## Materials and Methods

### Patients and clinicopathological data

In total, 136 primary invasive breast carcinomas were selected from previously analyzed patient cohorts mainly consisting of luminal B subtype tumors (Table 1; refs. 9–11). The patients were diagnosed in Western Sweden between 1991 and 1999, and the fresh-frozen tumor samples were stored in the tumor bank at the Sahlgrenska University Hospital Oncology Lab (Gothenburg, Sweden). clinicopathological information was obtained from Regional Cancer Centre West (Gothenburg, Sweden). The dataset was stratified into the molecular breast cancer subtypes (normal-like, basal-like, luminal subtype A, luminal subtype B, and HER2/ER<sup>-</sup>) and genomic grade index (low, high) using estrogen receptor-positive tumors, as described elsewhere (12–14). The study was approved by the Regional Ethical Review Board in Gothenburg, Sweden.

### Gene expression microarray

Illumina HumanHT-12 gene expression profiles for the 136 tumors were evaluated as described previously (10). In brief, data preprocessing and quantile normalization were applied to the raw signal intensities using the web-based BioArray Software Environment system (15) provided by Swegen Genomics DNA Microarray Resource Center (SCIBLU). Further data processing was performed in Nexus Expression 2.0 (BioDiscovery) using log<sub>2</sub>-transformed, normalized expression values and a variance filter. The microarray data have been previously validated using qRT-PCR with Spearman correlation coefficients (two-tailed) and showed a strong linear relationship between the Illumina HumanHT-12 BeadChip and qRT-PCR results ( $r_s = 0.97$ ;  $P < 0.01$ ; ref. 10). The gene expression data discussed in this publication are accessible through the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GSE20462, GSE97177).

### Selection of prognostic panel

The prognostic model was developed in three steps using the gene expression microarray data. First, Cox regression-based modeling was used to reduce the number of genes that can predict favorable and unfavorable disease course. A series of univariable Cox regression models were fitted for each of the 48,796 probes, and the effect of each gene on disease-specific survival (DSS) was estimated. The univariable hazard ratio (HR) and associated  $P$  values were calculated. Second, the number of statistically significant probes was reduced using Bonferroni correction for multiple testing. Finally, iterative Bayesian model averaging (BMA) was applied on multivariable Cox regression models with the gene

expression signatures as predictors and survival status as the dependent variable to select a panel of genes with good trade-off between estimation and predictive power (16, 17). For each subset, the BMA algorithm retained variables with posterior probability  $> 0.5$ . The removed variables were replaced by new variables until the procedure iterated through the entire dataset. In the final iteration, genes with posterior probabilities  $> 0.5$  were selected and the rest were removed.

### Training cohort

The 136 samples from patients with complete follow-up information were used to develop the predictive model. The training cohort was constructed using 79 of the 136 tumors with complete clinical information for established clinicopathological features (age, number of positive axillary lymph nodes, histologic grade, tumor size, ER, PR, and HER2 status). Survival time was defined as the time from initial diagnosis to breast cancer-related death for DSS and the time from initial diagnosis to death from any cause for overall survival (OS). In the validation, the 18-marker signature was tested using 79 of the 136 breast tumors (training cohort), as 57 samples were excluded due to missing data for the established clinicopathological features.

### Validation cohorts

External validation was performed using two different publicly available breast cancer datasets profiled with the Affymetrix Human Genome U133 Set. The GSE1456 dataset consisted of 159 breast tumors, of which 128 had complete information for molecular subtype and histologic grade (18). Survival data were available for DSS, OS, and recurrence-free survival (RFS; time from surgical lesion removal to detection of tumor recurrence). The second dataset (GSE4922) comprised 249 tumors (Uppsala cohort), of which 237 tumors had complete information for age, ER status, tumor size, axillary lymph node status, and histologic grade (19). For this dataset, survival data were available for disease-free survival (DFS), which was defined as time from initial diagnosis to first relapse or breast cancer-related death. Of the 18 identified prognostic genes, 11 genes (*ACAA1*, *ADGRG6*, *BORCS6*, *CCNA2*, *CDKN2A*, *HJURP*, *HSPA14*, *KIAA0494*, *NEIL3*, *STAM*, and *TRIP13*) were found on the U133A Affymetrix chip and seven genes (*CDCA5*, *FAM91A1*, *LRRCC1*, *MTURN*, *PRR11*, *SKA2*, and *SNX8*) were found on U133B Affymetrix chip. To validate all 18 markers, the U133A and U133B sets were first normalized separately, and then, the log<sub>2</sub>-transformed values for both sets were merged for each cohort.

The third dataset [The Cancer Genome Atlas (TCGA) Breast Invasive Carcinoma dataset; ref. 20] consisted of mRNA sequencing (mRNA-seq) data for 900 primary breast tumors, of which 720 tumors had clinical information for the number of positive axillary lymph nodes, tumor size, age, ER, and PR status. Survival time was given as OS. The clinical characteristics of the validation cohorts stratified by the linear predictor can be found in Supplementary Table S1.

### Statistical analysis

Statistical analyses were performed using a 0.05  $P$  value cutoff in R (v3.3.2). The 18-marker panel was identified using the R-packages {BMA} (v3.18.7; ref. 21) and {iterativeBMAurv} (v1.32.0; ref. 16). Data processing of the training and validation datasets was performed separately using the R-packages {survival} (v2.40-1) to fit Cox proportional hazards models, {survminer}

**Table 1.** clinicopathological characteristics of the 136 breast cancer patients with confounding factors stratified by linear predictor (risk group)

	No. of patients (%)			P
	Total (n = 136)	High-risk group (n = 65)	Low-risk group (n = 71)	
Breast cancer-related death	67 (49.3)	55 (84.6)	12 (16.9)	<0.001
Cause of death				<0.001
Breast cancer	67 (49.3)	55 (84.6)	12 (16.9)	
Other	11 (8.1)	3 (4.6)	8 (11.3)	
Alive	58 (42.6)	7 (10.8)	51 (71.8)	
Follow-up time (days)	2874.7 (1762.7)	1718.6 (1344.8)	3933.1 (1400.5)	<0.001
Age (years)	58.9 (13.8)	57.6 (15.1)	60.1 (12.5)	0.305
Tumor size (mm)	32.2 (25.6)	34.6 (17.6)	29.9 (31.3)	0.304
Pathologic tumor status				0.504
pT1	36 (26.5)	14 (21.5)	22 (31.0)	
pT2	60 (44.1)	29 (44.6)	31 (43.7)	
pT3	30 (22.1)	17 (26.2)	13 (18.3)	
pT4c	1 (0.7)	0 (0.0)	1 (1.4)	
pT4d	4 (2.9)	3 (4.6)	1 (1.4)	
Not available	5 (3.7)	2 (3.1)	3 (4.2)	
Histology				0.827
Ductal	95 (69.9)	45 (69.2)	50 (70.4)	
Lobular	13 (9.6)	5 (7.7)	8 (11.3)	
Other	20 (14.7)	11 (16.9)	9 (12.7)	
Not available	8 (5.9)	4 (6.2)	4 (5.6)	
ER				0.003
Negative	28 (20.6)	21 (32.3)	7 (9.9)	
Positive	107 (78.7)	43 (66.2)	64 (90.1)	
Not available	1 (0.7)	1 (1.5)	0 (0.0)	
PR				<0.001
Negative	57 (41.9)	40 (61.5)	17 (23.9)	
Positive	78 (57.4)	24 (36.9)	54 (76.1)	
Not available	1 (0.7)	1 (1.5)	0 (0.0)	
HER2				0.846
Negative	119 (87.5)	56 (86.2)	63 (88.7)	
Positive	17 (12.5)	9 (13.8)	8 (11.3)	
Surgery				0.584
Lumpectomy	44 (32.4)	19 (29.2)	25 (35.2)	
Mastectomy	69 (50.7)	33 (50.8)	36 (50.7)	
Not available	23 (16.9)	13 (20.0)	10 (14.1)	
Endocrine treatment				0.091
No	31 (22.8)	19 (29.2)	12 (16.9)	
Yes	21 (15.4)	12 (18.5)	9 (12.7)	
Not available	84 (61.8)	34 (52.3)	50 (70.4)	
Radiotherapy				0.654
No	71 (52.2)	33 (50.8)	38 (53.5)	
Yes	40 (29.4)	18 (27.7)	22 (31.0)	
Not available	25 (18.4)	14 (21.5)	11 (15.5)	
Chemotherapy				0.885
No	76 (55.9)	36 (55.4)	40 (56.3)	
Yes	35 (25.7)	16 (24.6)	19 (26.8)	
Not available	25 (18.4)	13 (20.0)	12 (16.9)	
Molecular subtypes				0.016
Basal-like	15 (11.0)	13 (20.0)	2 (2.8)	
HER2/ER <sup>-</sup>	15 (11.0)	7 (10.8)	8 (11.3)	
Luminal subtype A	2 (1.5)	1 (1.5)	1 (1.4)	
Luminal subtype B	104 (76.5)	44 (67.7)	60 (84.5)	
Axillary lymph node status				0.17
pN0	63 (46.3)	34 (52.3)	29 (40.8)	
pN1	61 (44.9)	28 (43.1)	33 (46.5)	
Not available	12 (8.8)	3 (4.6)	9 (12.7)	
Lymph node ratio	0.22 (0.31)	0.28 (0.36)	0.16 (0.25)	0.041
SBR grade				0.003
I	15 (11.0)	1 (1.5)	14 (19.7)	
II	49 (36.0)	24 (36.9)	25 (35.2)	
III	20 (14.7)	14 (21.5)	6 (8.5)	
Not available	52 (38.2)	26 (40.0)	26 (36.6)	
Genomic grade index				<0.001
High	64 (47.1)	37 (56.9)	27 (38.0)	
Low	44 (32.4)	7 (10.8)	37 (52.1)	
Not available	28 (20.6)	21 (32.3)	7 (9.9)	

(Continued on the following page)

Downloaded from http://aacrjournals.org/cebp/article-pdf/26/11/1619/1941207/1619.pdf by guest on 13 April 2024

**Table 1.** clinicopathological characteristics of the 136 breast cancer patients with confounding factors stratified by linear predictor (risk group) (Cont'd)

	No. of patients (%)			<b>P</b>
	<b>Total (n = 136)</b>	<b>High-risk group (n = 65)</b>	<b>Low-risk group (n = 71)</b>	
Oncotype Dx				<b>&lt;0.001</b>
High-risk group	69 (50.7)	16 (24.6)	53 (74.6)	
Low-risk group	67 (49.3)	49 (75.4)	18 (25.4)	
DNA index	1.09 (0.27)	1.14 (0.30)	1.06 (0.23)	0.095
Ploidy status				0.338
Aneuploid	20 (14.7)	13 (20.0)	7 (9.9)	
Diploid	104 (76.5)	46 (70.8)	58 (81.7)	
Multiploid	3 (2.2)	1 (1.5)	2 (2.8)	
Not available	9 (6.6)	5 (7.7)	4 (5.6)	

NOTE: Categorical variables show the total number of patients and percentage in parentheses, while continuous variables are given as the mean and SD. Statistically significant variables ( $P < 0.05$ ) are displayed in bold text.

(v0.2.2) to generate Kaplan–Meier plots, and {risksetROC} (v1.0.4) to generate AUC(t) plots.

Univariable and multivariable Cox proportional hazard models were individually fitted for each cohort using the  $\log_2$ -transformed expression values for the 18 genes as continuous variables. The high- and low-risk groups were assigned according to the linear predictor  $\eta$  (eta), which represents the product of the covariate vector  $x$  and the parameter vector  $\beta$ . Patients with  $\eta > 0$  were classified as high-risk patients and those with  $\eta < 0$  as low-risk patients. If  $\eta = 0$ , the patient cannot be classified in the high- or low-risk groups, because an HR of 1 means the covariate has no effect on the model.

The AUC was based on survival data and complete clinical data (for established and complete models) with  $\eta$  as a marker. As complete clinical information is needed for the combined models (18 markers and established markers), only patients with complete information were used for the training and validation cohorts. The C-index is a scalar that depicts the global accuracy of a fitted survival model as an overall summary and can be seen as the generalization of the AUC over time [range, 0.5 (random prediction) to 1 (perfect discrimination); refs. 22, 23].

The commercially available Oncotype Dx is an RT-PCR-based 21-gene signature that includes 16 cancer-related genes and 5 reference genes for normalization (24). A multivariable model was fitted on the basis of gene expression microarray data for the training cohort using the 16 cancer-related genes. The C-indices and AUC(t) functions were obtained as described above. Internal validation was performed using bootstrapping with 1,000 iterations. This study complied with the guidelines for reporting recommendations for tumor marker prognostic studies (REMARK; Supplementary Table S2; refs. 25, 26).

### Ingenuity pathway analysis

The 18 candidate genes were analyzed using the Ingenuity Pathway Analysis (IPA) software (Ingenuity Systems) using Fisher exact test ( $P$  value cutoff of 0.05). The Diseases and Bio Functions tool was used to detect diseases and disorders associated with specific genes and molecular and cellular functions.

## Results

### Multivariable predictive modeling identified an 18-marker signature associated with breast cancer-specific death

Univariable Cox regression analysis identified 9,159 transcripts in the training cohort ( $n = 136$ ) that were significantly associated with DSS. After adjusting for multiple testing using Bonferroni

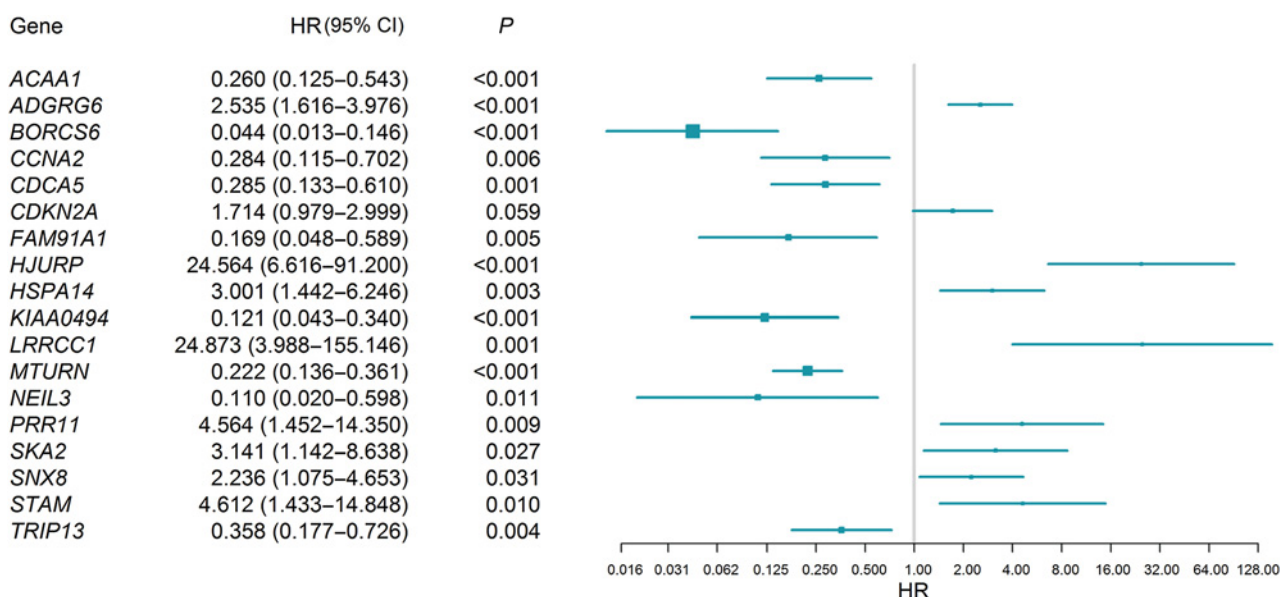
correction, the number of significant genes was reduced to 186. The 186 genes were further reduced to 18 using iterative BMA, resulting in a robust model with predictive power above 0.8 as measured by time-dependent AUC curves (Supplementary Table S3 for detailed gene functions).

The HRs of the 18 genes of the final multivariable model ( $n = 136$ ) showed that nine of the 18 genes (*ACAA1*, *BORCS6*, *CCNA2*, *CDCA5*, *FAM91A1*, *KIAA0494*, *MTURN*, *NEIL3*, and *TRIP13*) were negatively associated with breast cancer-specific mortality and thus had a favorable effect on the survival (Fig. 1). The remaining nine genes (*ADGRG6*, *CDKN2A*, *HJURP*, *HSPA14*, *LRRCC1*, *PRR11*, *SKA2*, *SNX8*, and *STAM*) were positively associated with the event and therefore had an unfavorable effect on survival. Seventeen of the 18 markers were significant in the multivariable model, where only the *CDKN2A* gene was nonsignificant ( $P = 0.059$ ) but was kept in the model due to its strong contribution to the predictive power of the multivariable model.

### Similar trends in univariable Cox regression models

Univariable Cox proportional hazard models were fitted for each of the 18 genes to test the prognostic potential (Table 2). All 18 markers were significant for both DSS and OS in the training cohort ( $n = 79$ ;  $P < 0.05$ ). The Cox coefficients showed the same tendencies in the complete 136-patient cohort. In the validation, we focus on the 79-patient subset to be able to put these results into a context with the AUC(t) functions. Supplementary Table S4 gives an overview of clinical characteristics and confounders of the risk groups for the 79-patient subset. Cox regression coefficients are directly related to hazard rates, where positive coefficients represent unfavorable prognosis ( $HR > 1$ ) and negative coefficients exert protective effects ( $HR < 1$ ).

In the microarray-based validation cohorts, the univariable Cox regression models revealed four markers (*CCNA2*, *CDCA5*, *HJURP*, and *TRIP13*) significantly associated with DFS (GSE4922) as well as DSS, RFS, and OS (GSE1456; Table 2). Positive Cox coefficients were consistently observed for the four markers in the training and microarray-based validation cohorts, thereby indicating an association with unfavorable prognosis. Further coefficient consistency throughout all cohorts could be observed for the *ADGRG6*, *BORCS6*, *FAM91A1*, *HSPA14*, *MTURN*, and *NEIL3* genes, although all of these genes were not statistically significant. Taken together, 10 of the 18 markers had the same Cox regression trends in all the cohorts. Furthermore, the results for the two microarray-based validation cohorts were generally more similar than the TCGA mRNA-seq data.



**Figure 1.** Forest plot with hazard ratio (HR) for the 18 genes of the final multivariable model in the training cohort. HRs above one indicate that a gene is positively associated with the event probability and thus negatively with survival time. The box size is based on precision, and the x-axis has a logarithmic scale [a bigger box size represents a more precise confidence interval (95% CI)].

**Kaplan–Meier estimators based on linear predictors showed significant stratification in all cohorts**

Multivariable Cox proportional hazard models based on the linear predictor were fitted for the training and validation cohorts leading to a stratification of the patients into a low- and high-risk group (Fig. 2). In proportional hazard models, the HR is the exponentiated Cox coefficient and provides an estimate of the ratio of the hazard rate in the high-risk group versus the low-risk group over time, where HR = 1 means no difference in the survival rates for the two risk groups. In the training cohort, the HR for DSS in the 79-patient subset was 0.069 and indicated a 93% decrease in mortality risk in the low-risk group compared with the high-risk group, which reciprocally means that at any time 15 times as many patients died in the high-risk group compared with the low-risk group. The HR in the complete training cohort ( $n = 136$ ; Fig. 2B) was similarly low at 0.089. In the validation cohorts, the HR ranged from 0.182 for RFS in the GSE1456 cohort to 0.417 for the TCGA dataset. The HR for DSS (GSE1456) was 0.184 and specified an 82% decrease in mortality risk in the low-risk group compared with the high-risk group, indicating that 5 times as many patients died in the high-risk group compared with the low-risk group.

Survival analysis showed that the 18-marker signature was a proficient predictor of survival in the training and validation cohorts with the most striking HRs for DSS and RFS in the GSE1456 cohort (Fig. 2). All corresponding log-rank test  $P$  values for the training and validation cohorts were significant ( $P < 0.05$ ).

**The 18-marker signature improved outcome prediction**

Subsequently, multivariable models were fitted based on (i) only the 18 markers; (ii) only the established clinicopathological markers; and (iii) a combined model of the 18 markers and the

established markers (Table 3). The C-index gives a global overview of how well the model discriminates between favorable and unfavorable outcome. A C-index above 0.7 typically indicates a good model, while a C-index above 0.8 is considered a strong model. In the training cohort, the 18-marker model performed well for DSS (C-index = 0.913) and OS (C-index = 0.896). Combining the 18-marker panel with the established markers improved outcome prediction to 0.930 for DSS and 0.929 for OS. The C-indices for the established markers model never surpassed 0.8. In the validation cohorts, the 18-marker model performed well for DSS, with a C-index of 0.803. In the combined model, the C-index increased to 0.829. The C-indices for the other models were not higher than 0.8.

**AUC(t) functions demonstrated highest predictive power for the combined multivariable model in all cohorts**

The AUC(t) functions of the multivariable models were developed to give an indication of how well the markers could differentiate between the two prognosis groups (Fig. 3). The advantage of AUC(t) functions is the portrayal of the discriminative ability of the model over time as opposed to the global C-index that considers the rank of patients according to their survival times (22).

In the training cohort ( $n = 79$ ), the AUC(t) function confirmed the accuracy of the 18-marker model over time with a constant slow decline (Fig. 3A). The AUC(t) functions of the established markers (patient age at diagnosis, histologic grade, number of positive axillary lymph nodes, pathologic tumor size, ER, PR, and HER2 status) started at a very high level ( $\sim 0.95$ ) but quickly dropped to 0.85 within the first 300 days and declined further to about 0.75 with time. The combined model showed the strongest predictive power.

Downloaded from <http://aacrjournals.org/cebp/article-pdf/26/11/1619/1941207/1619.pdf> by guest on 13 April 2024

**Table 2.** Univariable Cox proportional hazards models in the training (left) and validation cohorts (right) displaying Cox regression coefficients and P values for each of the 18 genes

18 Genes	Training cohort				Validation Cohorts											
	GSE20462, GSE97177 (n = 79)		TCGA data (n = 720)		GSE4922 (n = 237)			DSS			RFS			OS		
	Coefficient	P	Coefficient	P	Coefficient	P	Coefficient	P	Coefficient	P	Coefficient	P	Coefficient	P	Coefficient	P
ACAA1	-1.479	<b>0.001</b>	-1.432	<b>0.001</b>	-0.680	<b>&lt;0.001</b>	0.381	0.489	1.263	0.105	1.166	0.078	0.865	0.227		
ADGRG6	0.859	<b>&lt;0.001</b>	0.842	<b>&lt;0.001</b>	0.130	<b>0.025</b>	0.295	<b>0.041</b>	0.351	0.185	0.272	0.192	0.247	0.269		
BORCS6	-2.472	<b>&lt;0.001</b>	-2.455	<b>&lt;0.001</b>	-0.421	<b>0.010</b>	-0.072	0.653	-0.347	0.245	-0.306	0.227	-0.260	0.332		
CCNA2	0.842	<b>&lt;0.001</b>	0.870	<b>&lt;0.001</b>	0.128	0.145	0.469	<b>0.006</b>	1.086	<b>0.001</b>	0.863	<b>&lt;0.001</b>	0.803	<b>0.003</b>		
CDCA5	0.796	<b>&lt;0.001</b>	0.823	<b>&lt;0.001</b>	0.117	0.204	0.155	<b>0.040</b>	0.513	<b>0.001</b>	0.493	<b>&lt;0.001</b>	0.362	<b>0.005</b>		
CDKN2A	1.213	<b>&lt;0.001</b>	1.118	<b>&lt;0.001</b>	-0.016	0.810	-0.389	0.105	0.332	0.189	0.255	0.253	0.226	0.352		
FAM91A1	2.752	<b>&lt;0.001</b>	2.775	<b>&lt;0.001</b>	0.437	<b>0.002</b>	0.319	0.245	0.287	0.540	0.386	0.314	0.216	0.603		
HJURP	1.571	<b>&lt;0.001</b>	1.596	<b>&lt;0.001</b>	0.060	0.478	0.594	<b>&lt;0.001</b>	0.687	<b>0.002</b>	0.578	<b>0.001</b>	0.497	<b>0.010</b>		
HSPA14	1.624	<b>0.001</b>	1.662	<b>&lt;0.001</b>	0.392	<b>0.022</b>	0.196	0.423	1.136	<b>0.007</b>	1.074	<b>0.003</b>	0.735	0.063		
KIAA0494	-2.503	<b>&lt;0.001</b>	-2.316	<b>&lt;0.001</b>	0.232	0.334	-0.909	0.078	-2.973	<b>&lt;0.001</b>	-2.309	<b>&lt;0.001</b>	-2.225	<b>0.001</b>		
LRRCC1	4.030	<b>0.004</b>	3.734	<b>0.006</b>	0.332	<b>0.009</b>	0.048	0.834	-0.269	0.539	0.052	0.883	-0.412	0.290		
MTURN	-0.925	<b>&lt;0.001</b>	-0.899	<b>&lt;0.001</b>	-0.119	0.328	-0.224	0.295	-0.565	<b>0.005</b>	-0.442	<b>0.015</b>	-0.390	<b>0.046</b>		
NEIL3	1.411	<b>0.001</b>	1.416	<b>0.001</b>	0.074	0.342	0.849	<b>0.003</b>	0.664	0.084	0.600	0.057	0.625	0.066		
PRR17	1.687	<b>&lt;0.001</b>	1.701	<b>&lt;0.001</b>	0.149	<b>0.022</b>	-0.097	0.822	-0.551	0.364	0.258	0.597	-0.460	0.387		
SKA2	1.344	<b>0.002</b>	1.492	<b>&lt;0.001</b>	0.057	0.697	0.366	0.051	0.239	0.506	0.353	0.227	-0.005	0.988		
SNX8	2.127	<b>&lt;0.001</b>	2.051	<b>&lt;0.001</b>	-0.339	0.093	0.186	0.433	0.622	0.055	0.508	0.067	0.525	0.073		
STAM	1.614	<b>0.025</b>	1.956	<b>0.005</b>	0.542	<b>0.024</b>	0.452	0.296	0.470	0.498	0.713	0.207	-0.190	0.772		
TRIP13	0.875	<b>&lt;0.001</b>	1.020	<b>&lt;0.001</b>	0.105	0.247	0.685	<b>&lt;0.001</b>	0.823	<b>0.005</b>	0.759	<b>0.002</b>	0.668	<b>0.012</b>		

NOTE: Statistically significant variables (P < 0.05) are displayed in bold text.

The AUC(t) functions for the validation cohorts were all very stable over time. The established markers included patient age at diagnosis, number of positive axillary lymph nodes, pathologic tumor size, ER, and PR status for all 720 patients in the TCGA mRNA-seq validation cohort; patient age at diagnosis, histologic grade, number of positive axillary lymph nodes, tumor size, and ER status for all 237 patients of the GSE4922 validation cohort (Supplementary Fig. S1); and histologic grade and subtype for all 128 patients of the GSE1456 validation cohort (Fig. 3B). In all validation cohorts and clinical endpoints, the 18-marker model performed better than the established marker model. The combined model showed the highest predictive power for DSS in the GSE1456 cohort, providing the strongest validation for the 18-marker panel.

### Internal validation and comparison of predictive power to Oncotype Dx-based gene signature

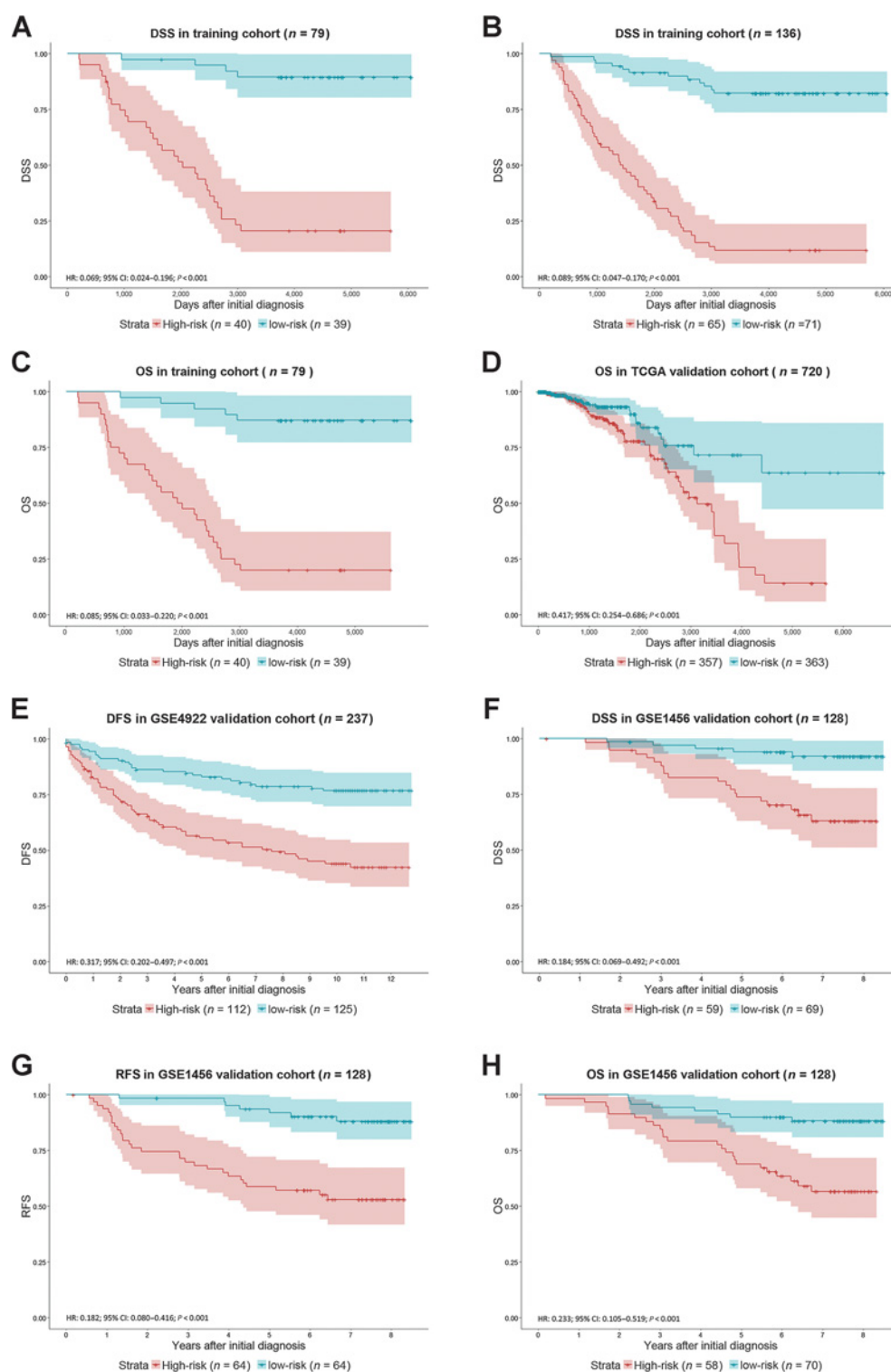
Bootstrapping of the training cohort (n = 136) with the 18-marker panel gave a mean C-index of 0.907 [confidence interval (CI), 0.871–0.945], which is close to the previously generated C-index of 0.913. Application of the Oncotype Dx-based 16-gene signature on the complete training cohort gave a C-index of 0.765 (CI, 0.706–0.816; Supplementary Fig. S2A). In ER-positive tumors (n = 107), the 18-marker panel had a C-index of 0.917 (CI, 0.873–0.957), while the Oncotype Dx-based 16-marker panel displayed a C-index of 0.785 (CI, 0.719–0.852; Supplementary Fig. S2B).

### Pathway analysis and disease association

IPA analysis of associated diseases and disorders showed that 15 of 18 genes (cancer) and 16 of 18 genes (organismal injury and abnormalities) play a crucial role in carcinogenesis (Supplementary Table S5). The molecular and cellular functions demonstrated that several of the 18 genes are associated with cell cycle (CCNA2, CDCA5, CDKN2A, HJURP, PRR11, SKA2, and TRIP13), cellular assembly and organization (BORCS6, CCNA2, CDCA5, CDKN2A, HJURP, SKA2, SNX8, and STAM), and DNA replication, recombination, and repair (CCNA2, CDCA5, CDKN2A, HJURP, NEIL3, and SKA2). Interestingly, network analysis showed that 11 of the 18 genes interact either directly or indirectly with one another, and the majority of the genes associated with DNA and cell-cycle regulation molecules are located inside the nucleus (Supplementary Fig. S3).

## Discussion

In the current study, we describe a robust prognostic gene signature that can effectively stratify breast cancer patients into low- and high-risk prognosis groups in the training and independent validation cohorts (27), which was confirmed by the stable AUC(t) function. Statistical literature makes a clear distinction between models developed for prediction and those developed for explanation (28). Medical research oftentimes focuses on an explanation, for example, clinically relevant and interpretable HRs. Here, we sacrificed to a certain extent the interpretability of the hazard rates to improve the predictive power of the model. As expected, the observed predictive power receded with time. However, the predictive power (measured by the AUC(t) function) remained above 0.8, even after 8 years. The strength of the proposed model is that the selection of the genes was not based on previous knowledge but solely on statistics. We are confident



**Figure 2.**

Kaplan–Meier analysis of the 18-marker signature in the training and validation cohorts. The x-axes depict days/years after initial diagnosis and the y-axes depict survival. **A** and **B**, Estimators of the probability of DSS according to risk assessment in the 79-patient subset training cohort ( $n = 79$ ) and the complete training cohort ( $n = 136$ ). **C**, Estimators of the probability of OS according to risk assessment in the 79-patient subset training cohort ( $n = 79$ ). **D**, Estimators of the probability of OS in the TCGA mRNA-seq validation cohort ( $n = 720$ ). **E**, Estimators of the probability of DFS in the GSE4922 validation cohort ( $n = 237$ ). **F–H**, Estimators of the probability of disease-specific, recurrence-free, and overall survival in the GSE1456 validation cohort ( $n = 128$ ). HRs, 95% confidence intervals (95% CI), and  $P$  values were calculated using the Cox proportional hazards regression and log-rank test, respectively.

**Table 3.** C-indices for the different multivariable models (18 markers, established markers and combined)

Cohort	18 markers	Established markers	Combined model
Training			
DSS	0.913	0.767	0.930
OS	0.896	0.778	0.929
Validation			
TCGA			
OS	0.665	0.631	0.699
GSE4922			
DFS	0.686	0.625	0.719
GSE1456			
DSS	0.803	0.707	0.829
RFS	0.754	0.678	0.769
OS	0.748	0.655	0.767

NOTE: Combining the 18 markers with the established markers generally improved outcome prediction.

that if the statistical stringency had been changed, the 18 genes would still have been included in the model. However, the proposed panel offered a good trade-off between parsimony and predictive capacity. In addition to the known involvement of several of the genes in pivotal cancer processes, the promising results of the prognostic 18-marker panel emphasize the biological potential of the marker panel in breast carcinoma.

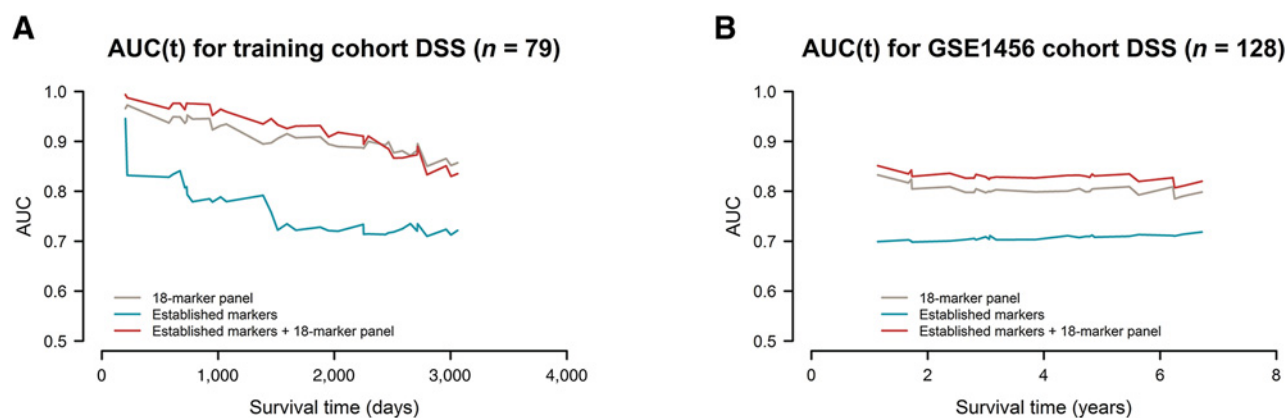
The lack of complete clinical information for the established markers in the training cohort presented a limitation of the study as the results for the multivariable model were generated using the 79-patient subset. Furthermore, the training and microarray-based validation cohorts originated from Swedish Cancer Registry studies. The vast majority of publicly available gene expression microarray datasets rarely contain matching clinical data leading to an overrepresentation of Swedish Cancer Registry studies due to detailed clinical data from extensively monitored patients. However, future studies should include other populations to examine the general clinical utility of our findings. In addition, the microarray platform needs to also be taken into consideration (Illumina

Human HT-12 Whole-Genome Expression BeadChip and Affymetrix Human Genome U133 Set). The probes on the two microarray platforms did not, in most cases, represent the same gene sequence, and the TCGA mRNA-seq dataset represented a different type of experiment. Despite using diverse validation datasets, true biological effects would most likely be detectable irrespective of the platform or type of experiment used to analyze gene expression. However, this comes with the downside that the training and validation cohorts were not directly comparable.

The external validation proved the feasibility of the 18-marker panel as a predictive model for breast cancer clinical outcome. To increase the power of the validation, several independent validation cohorts were evaluated with available gene expression microarray data and corresponding clinical data. The GSE1456 validation cohort was considered to be most relevant and similar to the training cohort because the 18-marker panel was generated using microarray data and DSS as the clinical endpoint. In the univariable survival analysis, 15 of the 18 genes showed the same tendency in the Cox regression coefficients for DSS in the training and the GSE1456 validation cohort. Among these 15 genes, seven were significant in both cohorts (*CCNA2*, *CDCA5*, *HJURP*, *HSPA14*, *KIAA0494*, *MTURN*, and *TRIP13*), providing a strong validation of the model.

The Kaplan–Meier estimators of the training and validation cohorts showed significant differences between the high- and low-risk groups. The low HR for the multivariable model (GSE1456 cohort for DSS) provided a strong validation of the predictive model. The significance of the other validation groups, which represent different types of survival endpoints (especially the TCGA cohort, which also represents a different type of experimental setup), highlighted the biological relevance of the genetically defined subgroups of patients.

The 18-marker panel had a higher predictive power compared with the established clinicopathological marker model in all cohorts. The predictive power further increased when combined with the established marker model in all cohorts despite different survival endpoints, which emphasized the clinical relevance of the

**Figure 3.**

AUC(t) functions of multivariable models. The lines represent the time-dependent area under the ROC curve (AUC(t)) for the 18-marker panel (gray), the established markers (blue) and the combined model with the 18-marker panel, and the established markers (red). The estimated performance of the combined model was better than that of the individual models in all cohorts and clinical endpoints. The predictive ability of all models was relatively stable over time. **A**, Estimated performance of the training cohort for DSS. Established clinical variables contain patient age at diagnosis, histologic grade, number of positive axillary lymph nodes, pathologic tumor size, ER, PR, and HER2 status for all 79 patients. **B**, Estimated performance of GSE1456 validation cohort for DSS. Established clinical variables contain histologic grade and subtype for all 128 patients.



panel. The relatively poor predictive power of the established marker model in the GSE1456 cohort can be explained by limited clinical information. The strong performance of the 18-marker model in the GSE1456 cohort, which represents a more naturally distributed cohort regarding the molecular subtypes as compared with the luminal B–overrepresented training cohort, proves that the model performs well for different subtypes.

The internal validation of the 18-marker panel using bootstrapping additionally confirmed the predictive power of the model. The 18-marker panel showed an evidently higher predictive power than the Oncotype Dx–based marker panel, which resembles the clinically relevant Oncotype Dx test. However, a clear limitation of this comparison was that Oncotype Dx is supposed to be used with the RT-PCR assay rates of 21 genes, where 5 of 21 genes represent reference genes for normalization. Here, we applied a multivariable model based on 16 of 21 genes using gene expression microarray data to gain an overview of the novelty and clinical benefit of the 18-marker signature.

The IPA analysis gave a first insight into the complex molecular networks behind this novel marker panel. The molecular and cellular functions proposed by IPA contained mainly functions in the cell cycle, cellular assembly and organization, and DNA replication, recombination, and repair, which serve as tools for carcinogenesis and explain that the majority of the 18 markers were significantly connected to cancer. Several of the markers have previously been associated with poor prognosis in breast cancer [*CCNA2* (29), *HJURP* (30, 31), *NEIL3* (32), and *TRIP13* (33, 34)], as well as the inclusion of the *ADGRG6* gene in the 70-gene MammaPrint signature (1).

In summary, the novel 18-marker panel proved to be a robust classification model with high predictive power. The predictive power was stable over time and gave the best prediction in combination with established markers for DSS. Use of the 18-marker panel in conjunction with clinical parameters can help to further stratify patients into risk groups, serve as a predictive tool for clinical outcome, and tailor treatment selection resulting in a more aggressive therapy for high-risk patients (reduce under-

treatment) and less aggressive therapy for low-risk patients (reduce overtreatment). Further research is needed to understand the interplay between these genes and thereby be able to develop better treatment alternatives for high-risk breast cancer patients.

### Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

### Authors' Contributions

**Conception and design:** J. Biermann, S. Nemes, T.Z. Parris, E. Forssell-Aronsson, G. Steineck, P. Karlsson, K. Helou

**Development of methodology:** J. Biermann, S. Nemes, T.Z. Parris

**Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.):** J. Biermann, T.Z. Parris, P. Karlsson

**Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis):** J. Biermann, S. Nemes, T.Z. Parris, H. Engqvist, G. Steineck, K. Helou

**Writing, review, and/or revision of the manuscript:** J. Biermann, S. Nemes, T.Z. Parris, H. Engqvist, E.W. Rönnerman, E. Forssell-Aronsson, G. Steineck, K. Helou

**Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases):** J. Biermann, T.Z. Parris, E.W. Rönnerman, P. Karlsson

**Study supervision:** T.Z. Parris, E. Forssell-Aronsson, P. Karlsson, K. Helou

### Acknowledgments

We are grateful to BILS (Bioinformatics Infrastructure for Life Sciences) and NBIS (National Bioinformatics Infrastructure Sweden) for their bioinformatics support.

### Grant Support

This work was supported by grants from the Swedish Cancer Society (CAN 2012/406; CAN 2015/311; to K. Helou), the King Gustav V Jubilee Clinic Cancer Research Foundation (2016:65; to K. Helou), and the LUA/ALF-agreement in West of Sweden Health Care Region (to P. Karlsson).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received June 29, 2017; revised August 23, 2017; accepted August 28, 2017; published OnlineFirst September 6, 2017.

### References

- van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002;415:530–6.
- Reis-Filho JS, Pusztai L. Gene expression profiling in breast cancer: classification, prognostication, and prediction. *Lancet* 2011;378:1812–23.
- Prat A, Ellis MJ, Perou CM. Practical implications of gene-expression-based assays for breast oncologists. *Nat Rev Clin Oncol* 2012;9:48–57.
- Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. *Nature* 2000;406:747–52.
- Györfy B, Hatzis C, Sanft T, Hofstatter E, Aktas B, Pusztai L. Multigene prognostic tests in breast cancer: past, present, future. *Breast Cancer Res* 2015;17:11.
- Schnitt SJ. Classification and prognosis of invasive breast cancer: from morphology to molecular taxonomy. *Mod Pathol* 2010;23:S60–S4.
- Duffy MJ. Tumor markers in clinical practice: a review focusing on common solid cancers. *Med Principles Pract* 2013;22:4–11.
- Mallett S, Royston P, Waters R, Dutton S, Altman DG. Reporting performance of prognostic models in cancer: a review. *BMC Med* 2010;8:21.
- Mollerstrom E, Delle U, Danielsson A, Parris T, Olsson B, Karlsson P, et al. High-resolution genomic profiling to predict 10-year overall survival in node-negative breast cancer. *Cancer Genet Cytogenet* 2010;198:79–89.
- Parris TZ, Danielsson A, Nemes S, Kovacs A, Delle U, Fallenius G, et al. Clinical implications of gene dosage and gene expression patterns in diploid breast carcinoma. *Clin Cancer Res* 2010;16:3860–74.
- Parris TZ, Kovacs A, Hajizadeh S, Nemes S, Semaan M, Levin M, et al. Frequent MYC coamplification and DNA hypomethylation of multiple genes on 8q in 8p11-p12-amplified breast carcinomas. *Oncogenesis* 2014;3:e95.
- Hu H, Li J, Plank A, Wang H, Daggard G. Comparative study of classification methods for microarray data analysis. Proceedings of the Fifth Australasian Conference on Data Mining and Analytics; 2006 Nov 29–30; Sydney, Australia. Darlinghurst, Australia: Australian Computer Society; 2006. Available from: <http://dl.acm.org/citation.cfm?id=1273813>.
- Loi S, Haibe-Kains B, Desmedt C, Lallemand F, Tutt AM, Gillet C, et al. Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. *J Clin Oncol* 2007;25:1239–46.
- Sotiriou C, Wirapati P, Loi S, Harris A, Fox S, Smeds J, et al. Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J Natl Cancer Inst* 2006;98:262–72.
- BASE. BASE - BioArray Software Environment. Available from: <http://base.thep.lu.se>.
- Annest A, Bumgarner R, Raftery A, Yeung KY. Iterative Bayesian Model Averaging: a method for the application of survival analysis to high-dimensional microarray data. *BMC Bioinformatics* 2009;10:72.
- Yeung KY, Bumgarner RE, Raftery AE. Bayesian model averaging: development of an improved multi-class, gene selection and classification tool for microarray data. *Bioinformatics* 2005;21:2394–402.

18. Pawitan Y, Bjöhle J, Amler L, Borg A-L, Egyhazi S, Hall P, et al. Gene expression profiling spares early breast cancer patients from adjuvant therapy: derived and validated in two population-based cohorts. *Breast Cancer Res* 2005;7:R953–R64.
19. Ivshina AV, George J, Senko O, Mow B, Putti TC, Smeds J, et al. Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer. *Cancer Res* 2006;66:10292–301.
20. TCGA. The Cancer Genome Atlas (TCGA). Available from: <https://cancer.genome.nih.gov/>.
21. Adrian Raftery JH, Volinsky C, Painter I, Yeung KY. BMA: Bayesian Model Averaging. R Package Version 3.18.72017.
22. Saha-Chaudhuri P, Heagerty PJ. Non-parametric estimation of a time-dependent predictive accuracy curve. *Biostatistics (Oxford, England)* 2013;14:42–59.
23. Harrell FE Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15:361–87.
24. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 2004;351:2817–26.
25. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. Reporting recommendations for tumor marker prognostic studies (REMARK). *J Natl Cancer Inst* 2005;97:1180–4.
26. Altman DG, McShane LM, Sauerbrei W, Taube SE. Reporting recommendations for tumor marker prognostic studies (REMARK): explanation and elaboration. *PLoS Med* 2012;9:e1001216.
27. Moons KM, Altman DG, Reitsma JB, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (tripod): explanation and elaboration. *Ann Int Med* 2015;162:W1–W73.
28. Shmueli G. To explain or to predict? *Stat Sci* 2010;25:289–310.
29. Gao T, Han Y, Yu L, Ao S, Li Z, Ji J. CCNA2 is a prognostic biomarker for ER+ breast cancer and tamoxifen resistance. *PLoS ONE* 2014;9:e91771.
30. Hu Z, Huang G, Sadanandam A, Gu S, Lenburg ME, Pai M, et al. The expression level of HJURP has an independent prognostic impact and predicts the sensitivity to radiotherapy in breast cancer. *Breast Cancer Res* 2010;12:R18–R.
31. Montes de Oca R, Gurard-Levin ZA, Berger F, Rehman H, Martel E, Corpet A, et al. The histone chaperone HJURP is a new independent prognostic marker for luminal A breast carcinoma. *Mol Oncol* 2015;9:657–74.
32. Shinmura K, Kato H, Kawanishi Y, Igarashi H, Goto M, Tao H, et al. Abnormal expressions of DNA glycosylase genes NEIL1, NEIL2, and NEIL3 are associated with somatic mutation loads in human cancer. *Oxidative Med Cell Longevity* 2016;2016:1546392.
33. Maurizio E, Wisniewski JR, Ciani Y, Amato A, Arnoldo L, Penzo C, et al. Translating proteomic into functional data: an high mobility group A1 (HMGA1) proteomic signature has prognostic value in breast cancer. *Mol Cell Proteomics* 2016;15:109–23.
34. Wang K, Sturt-Gillespie B, Hittle JC, Macdonald D, Chan GK, Yen TJ, et al. Thyroid hormone receptor interacting protein 13 (TRIP13) AAA-ATPase is a novel mitotic checkpoint-silencing protein. *J Biol Chem* 2014;289:23928–37.