

Prediction of Lymph Node Metastasis in Breast Cancer by Gene Expression and Clinicopathological Models: Development and Validation within a Population-Based Cohort



Looket Dihge^{1,2}, Johan Vallon-Christersson³, Cecilia Hegardt³, Lao H. Saal³, Jari Häkkinen³, Christer Larsson⁴, Anna Ehinger³, Niklas Loman^{3,5}, Martin Malmberg⁵, Pär-Ola Bendahl³, Åke Borg³, Johan Staaf³, and Lisa Rydén^{1,6}

Abstract

Purpose: More than 70% of patients with breast cancer present with node-negative disease, yet all undergo surgical axillary staging. We aimed to define predictors of nodal metastasis using clinicopathological characteristics (CLINICAL), gene expression data (GEX), and mixed features (MIXED) and to identify patients at low risk of metastasis who might be spared sentinel lymph node biopsy (SLNB).

Experimental Design: Breast tumors ($n = 3,023$) from the population-based Sweden Cancerome Analysis Network–Breast initiative were profiled by RNA sequencing and linked to clinicopathologic characteristics. Seven machine-learning models present the discriminative ability of N0/N+ in development ($n = 2,278$) and independent validation cohorts ($n = 745$) stratified as ER⁺HER2⁻, HER2⁺, and TNBC. Possible SLNB reduction rates are proposed by applying CLINICAL and MIXED predictors.

Results: In the validation cohort, the MIXED predictor showed the highest area under ROC curves to assess nodal metastasis; AUC = 0.72. For the subgroups, the AUCs for MIXED, CLINICAL, and GEX predictors ranged from 0.66 to 0.72, 0.65 to 0.73, and 0.58 to 0.67, respectively. Enriched proliferation metagene and luminal B features were noticed in node-positive ER⁺HER2⁻ and HER2⁺ tumors, while upregulated basal-like features were observed in node-negative TNBC tumors. The SLNB reduction rates in patients with ER⁺HER2⁻ tumors were 6% to 7% higher for the MIXED predictor compared with the CLINICAL predictor accepting false negative rates of 5% to 10%.

Conclusions: Although CLINICAL and MIXED predictors of nodal metastasis had comparable accuracy, the MIXED predictor identified more node-negative patients. This translational approach holds promise for development of classifiers to reduce the rates of SLNB for patients at low risk of nodal involvement.

Introduction

The benefit of extensive surgical excisions of axillary lymph nodes in primary breast cancer is being questioned on the basis of the advances in adjuvant therapy and the results of randomized

trials on axillary management in sentinel lymph node-positive disease (1–5). Although these trials mark a growing interest in minimizing axillary surgery, sentinel lymph node biopsy (SLNB) is still performed in all clinically node-negative patients despite the notion that the majority of them have disease-free axilla (6, 7). For these patients, the invasive procedure has no therapeutic benefit and better tools to predict the tumor's metastatic capacity are needed for correct guidance on surgical and systemic treatments.

Five intrinsic molecular subtypes of breast cancer have been identified by gene expression profiles, for distinguishing phenotypes and prognosis (8–10) and for guiding decisions on adjuvant treatment (11, 12). Although the prognostic value of gene profiling has been validated in node-negative disease (13), its prognostic utility has been extended to node-positive cases (14, 15). Despite the accumulated evidence of genomic profiles as prognostic indicators, nodal status retains its importance as an independent prognostic marker (16). Accurate preoperative assessment of axillary status and improved understanding of the biological processes associated with nodal metastasis are thus important for guidance on breast cancer treatment. Histopathological features, such as tumor size and lymphovascular invasion (LVI), have proved essential for assessing risk of nodal involvement and have been integrated into nomograms and risk models, together

¹Department of Clinical Sciences Lund, Division of Surgery, Lund University, Lund, Sweden. ²Department of Plastic and Reconstructive Surgery, Skåne University Hospital, Malmö, Sweden. ³Department of Clinical Sciences Lund, Division of Oncology and Pathology, Lund University, Lund, Sweden. ⁴Department of Laboratory Medicine, Division of Translational Cancer Research, Lund University, Lund, Sweden. ⁵Department of Oncology, Skåne University Hospital, Lund, Sweden. ⁶Department of Surgery, Skåne University Hospital, Lund, Sweden.

Note: Supplementary data for this article are available at Clinical Cancer Research Online (<http://clincancerres.aacrjournals.org/>).

J. Staaf and L. Rydén contributed equally to this article and shared senior authorship.

Corresponding Author: Looket Dihge, Division of Surgery, Department of Clinical Sciences Lund, Lund University, Skåne University Hospital, SE-22185 Lund, Sweden. Phone: 46-7025-56835; Fax: 46-4614-7298; E-mail: looket.dihge@med.lu.se

Clin Cancer Res 2019;25:6368–81

doi: 10.1158/1078-0432.CCR-19-0075

©2019 American Association for Cancer Research.

Translational Relevance

Although axillary nodal status is an important prognostic factor in breast cancer, the node-positive rate in newly diagnosed patients has decreased by early detection. Still, surgical staging is routinely performed in all clinically node-negative patients. This study demonstrates and independently validates the feasibility of noninvasive machine learning–based prediction models of nodal metastasis including RNA sequencing, alone or in combination with clinicopathologic factors in a population-based cohort. The accuracy of the mixed models showed an AUC 0.72 and highlighted that proliferation-associated genes were upregulated among node-positive ER⁺HER2⁻ and HER2⁺ tumors, while enriched basal-like features were observed in node-negative TNBC tumors. MIXED and CLINICAL predictors had similar accuracy, although the MIXED predictor identified more node negative patients and proposed a SLNB reduction rate of 20–30% in patients with ER⁺HER2⁻ disease. GEX predictors had inferior performance underscoring that the complexity of lymphatic dissemination is not only related to gene expression of the tumor.

with multifocality and age among other features (7, 17). The first nomogram was developed in 2007 at the Memorial-Sloan Kettering Cancer Center (MSKCC) with an AUC of 0.75 and has been validated in independent cohorts, including European and Chinese populations (18–20), with AUCs ranging from 0.67 to 0.78. However, risk assessment of metastatic axillary spread based on these clinicopathological variables is considered imperfect and consequently SLNB remains the standard axillary staging procedure (18, 21, 22). Previous studies of gene expression patterns from the primary breast tumor for prediction of nodal metastasis have shown inconsistent results, ranging from being nearly-perfect to almost inadequate in small, selected cohorts (23–28). The classifiers developed to date are not sufficient to support relevant clinical conclusions (29). A remaining question is thus whether gene expression profiling alone or in combination with clinical variables may surpass the performance of existing methods, for instance, nomogram methods that report AUCs between 0.70 and 0.75, in performance (18, 30).

In the current study, we aimed to evaluate the potential of clinicopathologic and gene expression-based predictors (single and mixed) of lymph node metastasis in a population-based contemporary breast cancer cohort ($n = 3,023$) analyzed by RNA sequencing (RNA-seq) to identify patients least likely to benefit from surgical axillary staging, and consequently reduce the rate of unbeneficial SLNB. Furthermore, we aimed to evaluate biological processes related to nodal metastasis and estimate the prognostic impact of these predictors.

Materials and Methods

Ethics approval

The study was approved by the Regional Ethical Review Board in Lund, Sweden (Registration numbers 2009/658, 2009/659, 2010/383, 2012/58, 2012/379, 2013/12, 2013/459, 2014/521, 2014/681, 2015/277, 2016/541, 2016/944, 2018/267) and was

conducted in accordance with the Declaration of Helsinki. All patients were included after obtaining written informed consent.

Patient cohort

The Swedish National Quality Registry for Breast Cancer (31) identified 5,235 patients undergoing primary surgery in the Southern Healthcare Region from September 1, 2010, to March 31, 2015. Of these, 4,633 patients were enrolled in the Sweden Cancerome Analysis Network–Breast (SCAN-B). SCAN-B employs a prospective, population-based, observational cohort study design (ClinicalTrials.gov ID: NCT02306096; refs. 32, 33) and the option for inclusion was proposed to all patients with primary breast cancer in the catchment area. High-quality tumor RNA-seq profiles were obtained for 3,023 patients (Fig. 1A). Information on age, tumor size, nodal status, mode of detection, multifocality, histologic grade, LVI, status of estrogen receptor (ER), progesterone receptor (PR), and HER2, adjuvant therapies, and overall survival (OS) were retrieved from the Swedish National Quality Registry for Breast Cancer (31) and Statistics Sweden (34). Ki-67 values were excluded due to incomplete data. Exclusion criteria for patients included neoadjuvant therapy and unknown or inconsistent pathologic nodal status from cancer registry data (Fig. 1A). Tumors were classified into four surrogate molecular subtypes based on the ER, PR, and HER2 status registered in the Swedish National Quality Registry for Breast Cancer: (i) all cases, (ii) ER⁺HER2⁻, (iii) HER2⁺, and (iv) triple-negative breast cancer (TNBC). Patients with micro- or macrometastases on SLNB were defined as lymph node-positive and were recommended completion axillary lymph node dissection (ALND) according to the Swedish National Guidelines for Breast Cancer at that time (35). The median follow-up for patients without an event was 4.53 years (0.16–6.78).

Gene expression analyses

Gene expression profiling was performed using RNA-seq as previously described (32, 33) and reported elsewhere (GSE96058; ref. 36) with complementary material in Supplementary Table S1. In brief, fresh breast tumor specimens, taken by a breast pathologist, were stored cooled in microtubes with the preservative solution (RNAlater) before delivery to the central SCAN-B laboratory in Lund, Sweden for tissue sample treatment and RNA-seq. Sequencing was performed to a depth of approximately 30 million paired-end reads (Illumina HiSeq2000 or NextSeq500 systems) using a dUTP protocol. The quality of RNA-seq data were assessed; aligned reads < 10 million, duplication level > 55%, or RNA integrity number (RIN) < 6 were excluded. A gene was represented by a single combined FPKM value of the matching transcript. Only refseq annotated genes ($n = 19,102$) were used for further analysis. An offset of +1 was added to all FPKM values followed by log₂ transformation. Intrinsic breast cancer subtype classification was performed by applying absolute intrinsic molecular subtyping (AIMS; ref. 37) to assign the PAM50 subtypes. AIMS subtype classification can be accomplished by a gene expression profile for each sample without the requirement of a normalized dataset. Further details on performed calculation of biological metagene scores (38) can be found in Supplementary Information, Appendix S1. Pathway analysis of differentially expressed genes was performed using the topGO R package as described (Supplementary Information, Appendix S1).

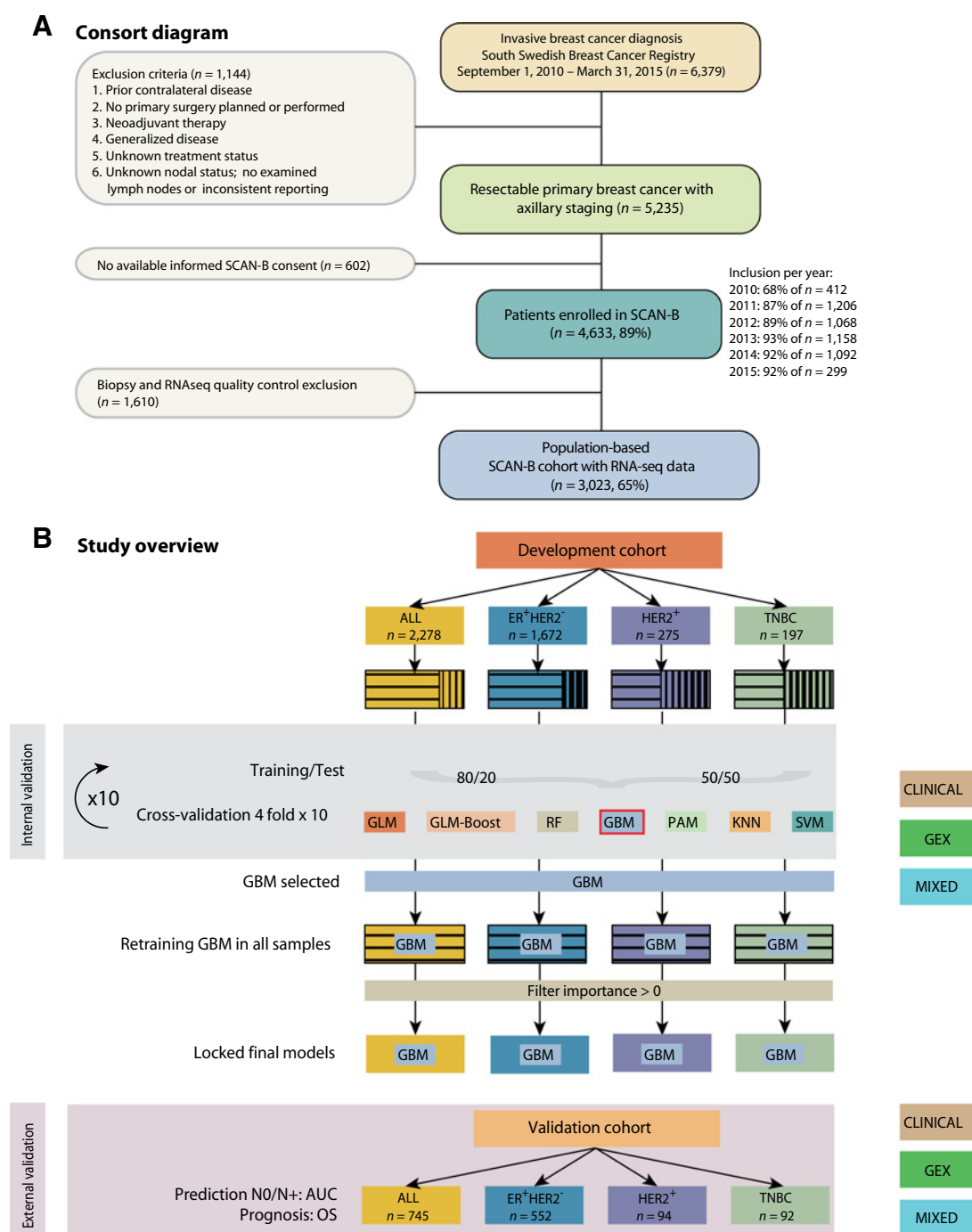


Figure 1.

Definition of a population-based breast cancer RNA-seq cohort and study overview. **A**, CONSORT diagram showing patients from the catchment area in southern Sweden included in the current Sweden Cancerome Analysis Network–Breast (SCAN-B) cohort. **B**, Seven machine-learning models (GLM, GLM-Boost, RF, GBM, PAM, KNN, and SVM) were applied to evaluate the predictive performance of nodal metastasis for the entire development cohort and for the surrogate molecular subtypes. For each of the four evaluation groups (All, ER⁺HER2⁻, HER2⁺, TNBC), three classifiers were trained, one based on clinicopathologic features alone (CLINICAL), one based on RNA-seq data alone (GEX), and one with both clinicopathologic parameters and RNA-seq data (MIXED), in a total of 7 × 4 × 3 separate models. The larger evaluation groups (All, ER⁺HER2⁻) were split into 80 percent training and 20 percent test groups, while the smaller evaluation groups (HER2⁺, TNBC) were split into 50 percent training and 50 percent test groups. Internal validation was performed by 10 iterations of 4-fold cross-validations, repeated 10 times. Mean area under the receiver operating characteristic (ROC) curve (AUC) identified generalized boosted regression models (GBM) with the highest predictive performance. The final GBM-based classification models were trained on the entire development cohort prior to validation in an independent dataset. GBM, generalized boosted regression models; GLM, generalized linear model; GLM-Boost, boosted generalized linear model; KNN, k-nearest neighbors algorithm; OS, overall survival; PAM, partition around medoids; RF, random forest; SVM, support vectormachine.

Machine learning models for prediction of lymph node status

The 3,023 patients were divided into development ($n = 2278$) and validation ($n = 745$) cohorts. The validation cohort comprised of patients diagnosed in 2011 (>5-year follow-up), while the remaining cases were assigned to the development cohort. In the development cohort, seven machine-learning models were applied to each of the four evaluation groups. For each evaluation group and model type, three classifiers were trained based on (i) clinicopathologic features alone (CLINICAL), (ii) RNA-seq data alone (GEX), and (iii) mix of clinicopathologic features and RNA-seq data (MIXED), totaling $4 \times 7 \times 3$ models. To identify the machine learning model with the best performance, the evaluation groups of the development cohort were divided into internal training and test sets by splitting the data into either 80% training and 20% test, or 50% training and 50% test (Fig. 1B). The training:test ratios were chosen based on the differences in sizes of the evaluation groups, which was due to the population-based nature of the SCAN-B cohort. Internal validation was performed by 10 iterations of 4-fold cross-validations, repeated 10 times (Fig. 1B). Assessment of model performance (the area under the ROC curve, AUC) identified generalized boosted regression models (GBM) with the highest predictive performance across evaluation groups and model types (CLINICAL, GEX, MIXED; Supplementary Fig. S1). A final GBM model (CLINICAL, GEX, MIXED) was then retrained using all samples in the development cohort for each evaluation group, removing features with no estimated importance (totaling 4×3 final GBM models). The inclusion of all samples in the development cohort for retraining the final locked models was to provide as many training cases as possible for GBM, as well as to avoid selectively choosing one model from the iterative internal validation process. Genes were eligible for training the final GEX and MIXED models if they were among the 5,000 most varying for each evaluation group and had a t test derived P value for difference between N0 and N+ samples <10% after Benjamin-Hochberg correction for multiple testing. Applying this approach, 74 to 1,633 eligible genes, depending on classifier and evaluation group, were included in the final training procedure (Supplementary Fig. S2). The final models were applied to the independent validation cohort for each evaluation group. Methodological details of the training procedure, feature selection, and deployed models are provided in the Supplementary Information, Appendix S1. Final data object models, including prediction scripts, are available on request from the authors.

Statistical analysis

A cutoff to discriminate nodal-negativity was set to correspond to maximized negative predictive value (NPV). This level of cutoff was intended to identify individuals with a very low probability of axillary disease where surgical axillary staging by SLNB would not be supported by the models. True positive (TP), true negative (TN), false positive (FP), and false negative (FN) results were assessed. The false negative rate (FNR) was calculated as the number of false negative cases divided by the number of all cases with axillary nodal metastasis ($FNR = FN/FN+TP$). This portrays the paucity of the models in predicting nodal spread as a ratio relative to all pathology-verified axillary nodal disease burden. Possible SLNB reduction rate = $(TN+FN)/(TN+FN+TP+FP)$ was assessed at a cut-off equivalent maximized NPV and at alternative cut-offs corresponding to the maximum FNRs of 5% and 10% reflecting accepted FNR of the SLNB procedure.

Survival curves were estimated by the Kaplan-Meier method and compared using the log-rank test. OS was the endpoint for survival analyses. HRs were estimated by univariable or multivariable Cox regression. SPSS Statistics for Windows version 24 (IBM Corp.) was used for descriptive statistics with Pearson χ^2 test. Machine learning and survival analyses were performed in R (39) using the Caret package and the `coxph` R function. The developing of the predictive models and the reporting of the findings were in accordance with an EQUATOR Guideline for reporting machine learning predictive models (40) and STARD 2015 checklist (41).

Results

Characteristics of the study population

The clinicopathologic characteristics for all patients in the catchment region and among those enrolled in SCAN-B, both in the development ($n = 2,278$) and validation cohorts ($n = 745$) were comparable (Table 1; Supplementary Table S1, page 2). However, tissue for RNA-seq was not always obtainable for tumors ≤ 10 mm. SLNB alone, without completion of ALND, was performed in 62% of patients. Axillary metastasis, N+, was found in 36% and 34% of the patients in the development and validation cohorts, respectively. Information on the metastatic size was incomplete; micrometastasis was the largest deposit in 6% of node-positive patients with obtainable records. There were significant differences in the proportion of lymph node metastasis across AIMS-PAM50 molecular subtypes within the development ($P < 0.001$) and validation cohorts ($P = 0.009$; Supplementary Table S2).

Predictive performances in the development and validation cohorts

Seven different machine-learning methods were evaluated and GBM was selected on the basis of best performance (Fig. 1B). GBM, in the implementation used by us, closely follows Friedman gradient boosting machine (42) and involves a stochastic gradient boosting strategy for logistic regression using a Bernoulli distribution (due to binary outcome of N0/N+). Gradient boosting combines multiple simple models "weak learners" into a single "strong" composite model through an iterative process. Figure 2A summarizes performances of the GBM models for prediction of nodal metastasis in the complete development cohort, with MIXED predictors achieving highest performance in all evaluation groups. Figure 2B shows predictive performance in the validation cohort. The overall levels of accuracy in the independent validation cohort suggest model robustness. However, a slight drop in performance was observed for HER2⁺ and TNBC validation groups for predictors including gene expression features, which could indicate possible overfitting in the training procedure as well as a reflection of small subgroups. Although the best predictive performance for the total validation cohort ($n = 745$) was achieved by a MIXED predictor [AUC = 0.72, 95% confidence interval (CI) = 0.68–0.76], the CLINICAL predictor showed comparable performance (AUC = 0.71; CI = 0.67–0.75). For the ER⁺HER2⁻ validation group, the CLINICAL model achieved an AUC of 0.73 (CI = 0.68–0.77), while the MIXED predictor achieved an AUC = 0.72 (CI = 0.68–0.77). For the HER2⁺ validation group, the CLINICAL predictor showed an AUC = 0.70 (CI = 0.58–0.81), and the MIXED model showed an AUC = 0.66 (CI = 0.55–0.78). In the TNBC validation group, the CLINICAL and MIXED predictors achieved AUCs of 0.65 (CI = 0.53–0.76) and 0.68 (CI = 0.56–0.81), respectively. Discriminatory

Table 1. Clinicopathologic characteristics and adjuvant treatment regimen within the development and validation cohorts

	All		Development cohort			Validation cohort			
	All n = 3,023	All n = 2,278	ER ⁺ HER2 ⁻ n = 1,672	HER2 ⁺ n = 275	TNBC n = 197	All n = 745	ER ⁺ HER2 ⁻ n = 552	HER2 ⁺ n = 94	TNBC n = 92
Age, years ^a	64	64	65	60	63	64	65	59	63
Tumor size (pT) ^b									
≤20 mm (pT1)	1,967 (65)	1,506 (66)	1,145 (69)	153 (56)	109 (56)	461 (62)	369 (67)	44 (47)	44 (48)
>20–≤50 mm (pT2)	980 (33)	717 (32)	492 (30)	114 (42)	80 (41)	263 (35)	172 (31)	46 (49)	42 (46)
>50 mm (pT3)	67 (2)	46 (2)	30 (2)	7 (3)	7 (4)	21 (3)	11 (2)	4 (4)	6 (7)
Missing	9	9	5	1	1	0	0	0	0
Multifocality									
Absent	1,317 (71)	736 (62)	567 (63)	93 (58)	58 (73)	581 (87)	421 (86)	78 (89)	77 (89)
Present	531 (29)	443 (38)	331 (37)	68 (42)	21 (27)	88 (13)	68 (14)	10 (11)	10 (12)
Missing	1,175	1,099	774	114	118	76	63	6	5
Histological type									
Ductal	2,442 (81)	1,840 (81)	1,317 (79)	255 (93)	168 (85)	602 (81)	426 (77)	92 (98)	78 (85)
Lobular	372 (12)	283 (12)	253 (15)	11 (4)	3 (2)	89 (12)	85 (15)	1 (1)	2 (2)
Other	209 (7)	155 (7)	102 (6)	9 (3)	26 (13)	54 (7)	41 (7)	1 (1)	12 (13)
Missing	0	0	0	0	0	0	0	0	0
Histological grade									
I	472 (16)	356 (16)	328 (20)	4 (2)	2 (1)	116 (16)	114 (21)	0 (0)	0 (0)
II	1,442 (48)	1,105 (49)	944 (57)	65 (24)	25 (13)	337 (46)	298 (54)	26 (28)	12 (13)
III	1,099 (37)	811 (36)	397 (24)	205 (75)	169 (86)	288 (39)	138 (25)	68 (72)	78 (87)
Missing	10	6	3	1	1	4	2	0	2
Lymphovascular invasion									
Absent	2,155 (81)	1,658 (81)	1,247 (83)	165 (69)	141 (77)	497 (81)	374 (83)	58 (73)	60 (74)
Present	502 (19)	385 (19)	2,253 (17)	76 (32)	42 (23)	117 (19)	75 (17)	21 (27)	21 (26)
Missing	366	235	172	34	14	131	103	15	11
ER status									
Positive	2,567 (86)	1,942 (86)	1,672 (100)	186 (68)	0 (0)	625 (84)	552 (100)	69 (73)	0 (0)
Negative	424 (14)	305 (14)	0 (0)	86 (32)	197 (100)	119 (16)	0 (0)	25 (27)	92 (100)
Missing	32	31	0	3	0	1	0	0	0
PR status									
Positive	2,205 (74)	1,672 (75)	1,461 (88)	133 (49)	0 (0)	533 (72)	480 (87)	48 (51)	0(0)
Negative	784 (26)	573 (26)	209 (13)	139 (51)	197 (100)	211 (29)	72 (13)	46 (49)	92 (100)
Missing	34	33	2	3	0	1	0	0	0
HER2 status									
Positive	369 (13)	275 (13)	0 (0)	275 (100)	0 (0)	94 (13)	0(0)	94 (100)	0 (0)
Negative	2,548 (87)	1,901 (87)	1,672 (100)	0 (0)	197 (100)	647 (87)	552 (100)	0 (0)	92 (100)
Missing	106	102	0	0	0	4	0	0	0
Regional lymph node status									
NO	1,940 (64)	1,448 (64)	1,071 (64)	152 (55)	135 (69)	492 (66)	368 (67)	58 (62)	62 (67)
N+	1,083 (36)	830 (36)	601 (36)	123 (45)	62 (31)	253 (34)	184 (33)	36 (38)	30 (31)
Adjuvant therapy									
Yes	2,664 (90)	2,003 (90)	1,495 (91)	258 (96)	142 (74)	661 (89)	499 (91)	89 (95)	66 (73)
None	311 (11)	233 (10)	151 (9)	11 (4)	50 (26)	78 (11)	48 (9)	5 (5)	25 (28)
Unknown	48	42	26	6	5	6	5	0	1
Type of adjuvant therapy ^c									
Endocrine therapy alone	1,518 (57)	1,101 (55)	1,007 (67)	32 (12)	2 (1)	417 (63)	398 (80)	14 (16)	1(1.5)
Chemotherapy alone	222 (8)	156 (8)	5 (0.3)	1 (0.4)	137 (70)	66 (10)	3 (0.6)	0 (0)	63 (95)
Endocrine therapy + chemotherapy	620 (23)	519 (26)	480 (32)	4 (1.5)	1 (0.5)	101 (15)	96 (19)	1 (1.1)	1 (1.5)
Any combination with HER2-directed therapy	303 (11)	226 (11)	3 (0.2)	220 (85)	2 (1)	77 (12)	2 (0.4)	74 (83)	1 (1.5)
Mode of detection ^d									
No. of patients in the target mammography screening population (ages 40–74 years) ^d	2,346	1,771	1,333	207	130	575	425	78	65
Mammography screening	1,460 (63)	1,127 (64)	891 (68)	109 (53)	62 (48)	333 (58)	269 (64)	37 (47)	23 (36)
Symptomatic presentation	864 (37)	625 (36)	428 (32)	96 (47)	67 (52)	239 (42)	154 (36)	41 (53)	41 (64)
Missing	22	19	14	2	1	3	2	0	1

NOTE: Data on clinicopathologic characteristics and obtained adjuvant treatment regimen are derived from The Swedish National Quality Registry for Breast Cancer. Values in parentheses are column percentages given for categorical variables unless otherwise indicated. Percentage have been rounded and may not total 100. Abbreviations: N+, axillary lymph node metastasis; NO, no regional lymph node metastasis.

^aValues are median.

^bAccording to the TNM classification for breast cancer, seventh edition.

^cPercentages are calculated from the total number of patients who received adjuvant treatment.

^dPercentages are calculated from the total number of patients within the age range (40–74 years) for mammography within the National Breast Cancer Screening Programme.

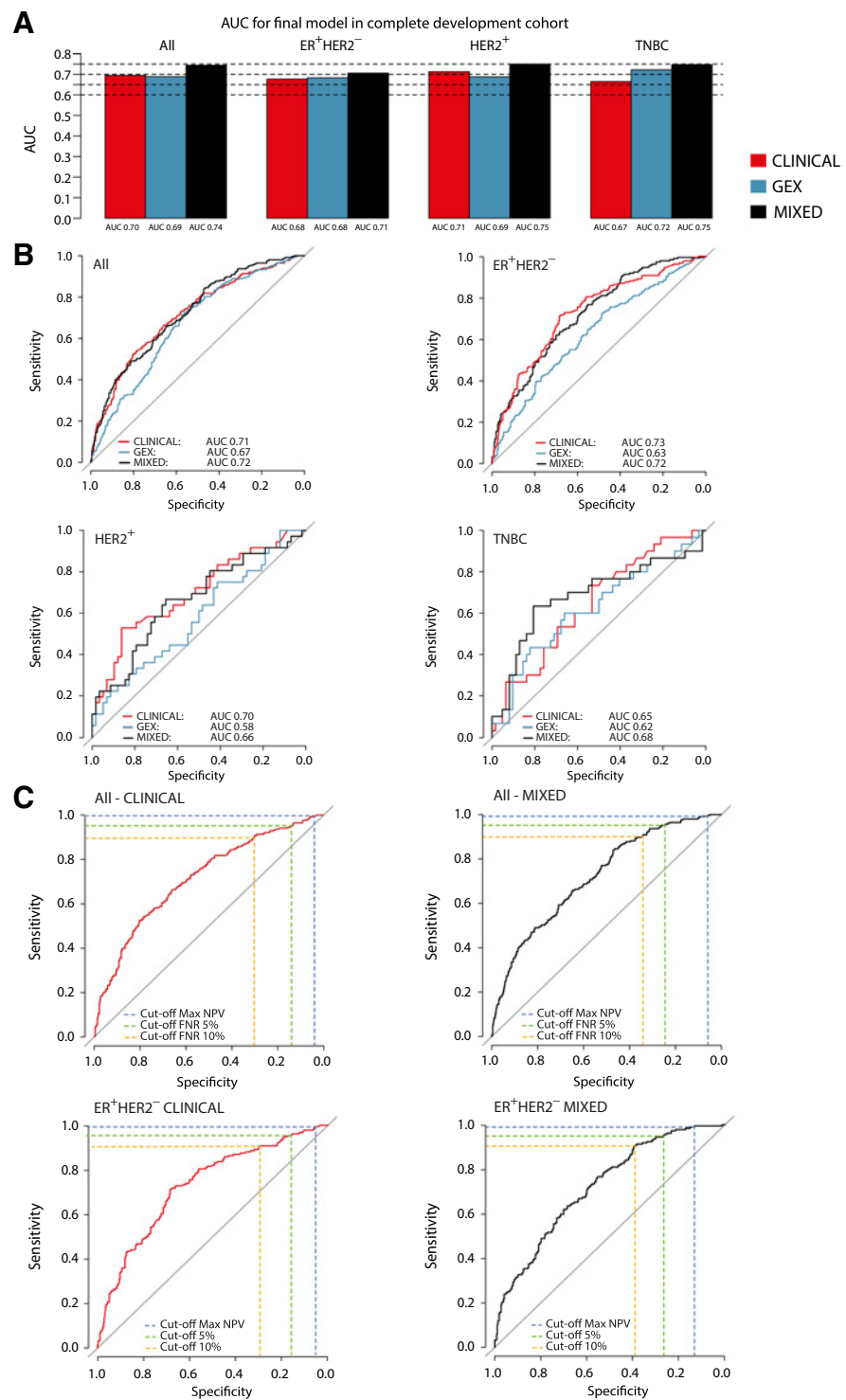


Figure 2.

Performance of developed predictors in predicting nodal status. **A**, Performance of the three GBM-based predictor models in predicting nodal metastasis presented as area under the receiver operating characteristic (ROC) curve (AUC) in the entire development cohort and stratified in subgroups. **B**, AUC curves estimating the predictive performance for nodal status classification in the independent validation cohort for all tumors and subgroups. **C**, AUC curves predicting nodal status in the validation cohort for all tumors (ALL) and for the ER⁺HER2⁻ subgroup using CLINICAL and MIXED predictors with cut-offs to discriminate nodal-negativity corresponding to maximized NPV and to the maximum FNRs of 5% and 10%, reflecting accepted FNR of the SLNB procedure. CLINICAL, predictor based on clinicopathologic features alone; GEX, predictor based on RNA-seq data alone; MIXED, predictor based on both clinicopathologic parameters and RNA-seq data.

performances of the GEX predictors were inferior to those of the CLINICAL and MIXED models, with AUCs ranging from 0.58 to 0.67. Although discriminatory performances were highly similar between the CLINICAL and MIXED predictors in general, Cohen kappa coefficients were 0.54, 0.56, 0.61, and 0.31 for the overall validation cohort, ER⁺HER2⁻, HER2⁺, and the TNBC subgroups, respectively (Supplementary Fig. S3A). These values indicate a fair to

moderate agreement between the two models. In comparison, better agreement was observed between the GEX and MIXED models in TNBC tumors (kappa coefficient = 0.66), and a modest agreement for the remaining groups (Supplementary Fig. S3B). When comparing the agreement between predictors (i.e., proportion of cases predicted similarly by two predictors), we found that 75% to 85% of cases were predicted similarly for the two different predictor set

comparisons across evaluation groups (Supplementary Fig. S3C). Finally, when stratifying N+ disease (the training label) to N1 (1–3 metastatic nodes) and N2 (≥4 metastatic nodes), we noted that most discrepant classification between CLINICAL, MIXED, and GEX predictors and the pathologic class occurred in the N1 group across all evaluation groups (Supplementary Fig. S3D). This observation indicates that N1 tumors have a heterogeneous

phenotype in terms of the investigated clinicopathologic variables and gene expression patterns.

Implication for sentinel lymph node reduction rate

Table 2 displays possible SLNB reduction rates after applying the prediction models to assign disease-free axilla (N0 vs. N+) in the SCAN-B validation cohort (both for the complete SCAN-B

Table 2. SLNB reduction rates using the models to predict disease-free axilla within the SCAN-B validation 2011 cohort; possible SLNB reduction rate corresponding to cut-offs at maximum negative predictive value, false negative rate 5% and 10%, respectively

All		NO vs. N+		
Clinicopathological		n = 745		
Cutoff Max NPV = 0.95	TP	TN	FP	FN
No.	252	20	472	1
SLNB reduction rate	(TN+FN)/(TP +TN +FP +FN) = 2.82%			
False negative rate	FN/(TP+FN) = 0.40%			
Cutoff NPV = 0.90	TP	TN	FP	FN
No.	241	69	423	12
SLNB reduction rate	(TN+FN)/(TP +TN +FP +FN) = 10.87%			
False negative rate	FN/(TP+FN) = 5%			
Cutoff NPV = 0.86	TP	TN	FP	FN
No.	228	149	343	25
SLNB reduction rate	(TN+FN)/(TP +TN +FP +FN) = 23.35%			
False negative rate	FN/(TP+FN) = 10%			
All		NO vs. N+		
MIXED		n = 745		
Cutoff Max NPV = 0.97	TP	TN	FP	FN
No.	252	30	462	1
SLNB reduction rate	(TN+FN)/(TP +TN +FP +FN) = 4.16%			
False negative rate	FN/(TP+FN) = 0.40%			
Cutoff NPV = 0.91	TP	TN	FP	FN
No.	241	127	365	12
SLNB reduction rate	(TN+FN)/(TP +TN +FP +FN) = 18.66%			
False negative rate	FN/(TP+FN) = 5%			
Cutoff NPV = 0.87	TP	TN	FP	FN
No.	228	171	321	25
SLNB reduction rate	(TN+FN)/(TP +TN +FP +FN) = 26.61%			
False negative rate	FN/(TP+FN) = 10%			
ER+HER2		NO vs. N+		
Clinicopathological		n = 552		
Cutoff Max NPV = 0.95	TP	TN	FP	FN
No.	183	19	349	1
SLNB reduction rate	(TN+FN)/(TP +TN +FP +FN) = 3.62%			
False negative rate	FN/(TP+FN) = 0.54%			
Cutoff NPV = 0.88	TP	TN	FP	FN
No.	176	59	309	8
SLNB reduction rate	(TN+FN)/(TP +TN +FP +FN) = 12.14%			
False negative rate	FN/(TP+FN) = 5%			
Cutoff NPV = 0.87	TP	TN	FP	FN
No.	167	110	258	17
SLNB reduction rate	(TN+FN)/(TP +TN +FP +FN) = 23.01%			
False negative rate	FN/(TP+FN) = 10%			
ER+HER2		NO vs. N+		
MIXED		n = 552		
Cutoff Max NPV = 0.98	TP	TN	FP	FN
No.	183	49	319	1
SLNB reduction rate	(TN+FN)/(TP +TN +FP +FN) = 9.06%			
False negative rate	FN/(TP+FN) = 0.54%			
Cutoff NPV = 0.92	TP	TN	FP	FN
No.	175	99	269	9
SLNB reduction rate	(TN+FN)/(TP +TN +FP +FN) = 19.57%			
False negative rate	FN/(TP+FN) = 5%			
Cutoff NPV = 0.89	TP	TN	FP	FN
No.	166	146	222	18
SLNB reduction rate	(TN+FN)/(TP +TN +FP +FN) = 29.71%			
False negative rate	FN/(TP+FN) = 10%			

Abbreviations: FN, false negative; FP, false positive; Max NPV, maximum negative predictive value; N+, axillary lymph node metastasis; No, no regional lymph node metastasis; SLNB, sentinel lymph node biopsy; TN, true negative; TP, true positive.

Downloaded from <http://aacrjournals.org/clinccancerres/article-pdf/25/21/6374/2054186/6374.pdf> by guest on 11 November 2024

2011 validation dataset and the largest evaluation group comprising ER⁺HER2⁻ tumors within the validation dataset) corresponding to cutoffs at the maximum negative predictive value (FNR 0.4%) and FNR 5% and 10% (Fig. 2C). In the overall validation cohort, a SLNB reduction rate of 2.8% would be attained by applying CLINICAL predictors at a cut-off corresponding to maximized NPV 95% (FNR 0.40%) to discriminate disease-free axilla. Allowing a maximum of 5% to 10% FNR in discriminating N0 by CLINICAL predictors, possible SLNB reduction rates of 11% to 23% were displayed. However, if MIXED predictors were applied, SLNB reduction rates of 19% to 30% would be yield at a maximum of 5% to 10% FNR (corresponding to the accepted level of FNR for the SLNB procedure). In general, MIXED predictors yielded 6% to 7% further reduction rates compared with CLINICAL models at 5% to 10% FNR.

Principal clinicopathologic variables included in the predictors

Variables in the final models are listed in Supplementary Table S3 for each predictor type and evaluation group. Among the CLINICAL predictors, the most significant variables associated with axillary metastasis were tumor size, LVI, age, and multifocality. Tumor size and LVI remained the top two variables associated with nodal disease for MIXED predictors in the validation cohort for all cases, ER⁺HER2⁻ cases, and HER2⁺ cases, but ranked seventh and sixth for the TNBC group, respectively.

Biological processes related to the predictors

Biological processes related to predictor classification were assessed by metagene expression analysis in four nodal status subgroups defined by combinations of pathologic anatomic diagnosis of the nodal status, N0/N+ (PAD-/PAD+), and model predicted axillary status, N0/N+ (Pred-/Pred+; Fig. 3A-C). For the ER⁺HER2⁻ cases, differences in expression of proliferation-related genes (checkpoint metagene) were noted among Pred-/Pred+ samples, regardless of the pathologic status (Fig. 3A).

PAM50 classification revealed heterogeneity in the four nodal status subgroups in the ER⁺HER2⁻ evaluation group (Fig. 3D). Nevertheless, distribution of the PAM50 subtypes was comparable in a pair-wise pattern defined by predictor nodal status (Pred-/Pred+) rather than by pathologic status (PAD-/PAD+) and involved higher proportions of the luminal B intrinsic subtype in the predicted N+ cases. Similar patterns were observed in the HER2⁺ evaluation group; samples with model-predicted nodal metastasis showed higher expression of proliferation-related genes (Fig. 3B) and intrinsic molecular subtypes associated with a more aggressive tumor type (Fig. 3E). For the TNBC evaluation group, low expression of basal-like markers was detected in samples with predicted nodal metastasis by the GEX and MIXED models but not by the CLINICAL model (Fig. 3C).

Unsupervised hierarchical analysis of the validation cohort was performed to investigate the relationship between global gene expression patterns and the distribution of nodal status, clinicopathologic variables, and breast cancer subgroups. Fig. 3G shows the transcriptional landscape in the context of N0/N+ status. Heterogeneous distribution was observed for N0/N+ status in relation to key clinicopathologic variables and breast cancer subgroups, irrespective of whether these subgroups were defined by surrogate subtype markers or by molecular PAM50 classification. Functional analysis based on gene ontologies (GO) for derived GEX and MIXED predictors are available in Supplementary Fig. S2, demonstrating strong association with proliferation

related biological processes especially in the ER⁺HER2⁻ group, but also in the HER2⁺ group.

Prognostic value of derived nodal predictors

Assessment of prognosis based on nodal status was performed in patients with ER⁺HER2⁻ disease ($n = 398$) who received adjuvant endocrine therapy only (the largest subgroup). Other treatment groups, such as chemotherapy plus endocrine therapy, did not include enough patients to allow statistical analysis and evaluation at this time. Figure 4A presents OS for the cohort stratified by the combinations of pathologic and predicted axillary status (PAD-Pred-, PAD+Pred-, PAD-Pred+, PAD+Pred+). On the basis of the available follow-up data, positive pathology-based lymph node status did not predict a significantly worse OS in Kaplan-Meier analysis (log-rank $P = 0.13$, Fig. 4A) or univariate Cox analysis (HR = 1.58; 95% CI = 0.87-2.86; Fig. 4B), although the HR indicates a trend toward significance. In contrast, predicted nodal status stratified patients into better (PAD-Pred- and PAD+Pred-) or worse (PAD-Pred+ and PAD+Pred+) outcomes irrespective of pathologic status, for all models. In univariable analysis, factors predicting OS included age (years), tumor size (mm), histologic grade 3 versus grade 1 and grade 2, node positivity versus node negativity predicted by any of the models, and combinations of the pathologic and predicted axillary status (PAD+Pred- or PAD-Pred+ or PAD+Pred+ versus PAD-Pred-) across models. Next, we tested whether N0/N+ status by PAD added prognostic information in multivariable Cox regression models with the predictions from GEX, CLINICAL, or MIXED as the only covariates included in the model, respectively. However, in none of the three analyses N0/N+ by PAD reached statistical significance ($P > 0.05$), while the GEX, CLINICAL, and MIXED predictions did ($P < 0.05$). It should be noted that our derived models were not optimized for outcome; still they capture a signal showing prognostic value irrespective of the pathologically defined nodal status. Finally, the prognostic importance of the PAD+Pred+ versus the PAD-Pred- remained significant (HR = 2.13; CI = 1.06-4.26; $P = 0.03$) in a Cox multivariable analysis based on the GEX predictor when adjusted for tumor size (mm) and age (years) (Fig. 4B and C), but was not significant (HR = 2.04; CI = 0.93-4.49; $P = 0.07$) when adjusting for histologic grade. CLINICAL and MIXED predictors were not tested in the multivariable analysis as they included the adjusted covariates.

Discussion

This study presents data from the largest prospectively collected and population-based cohort to date on genomic predictors for nodal metastasis, and comprehensively evaluates and independently validates mRNA-seq and clinicopathologic features as predictors of nodal metastasis in ER⁺HER2⁻, HER2⁺, and TNBC tumors. Today (43), a precise noninvasive option for nodal staging would circumvent the adverse effects of axillary surgical staging since the majority of breast cancer patients present with low-risk tumors, node-negative disease, and excellent overall prognosis. In the overall validation cohort, the predictor using RNA-seq features and clinicopathological characteristics (MIXED) showed the highest accuracy in predicting nodal metastasis, attaining AUC = 0.72. Thus, the ability to classify nodal status using current predictors and endpoint definition was not perfect. Nevertheless, the nodal metastatic features recognized by these predictors implied an added prognostic impact compared to pathological nodal status

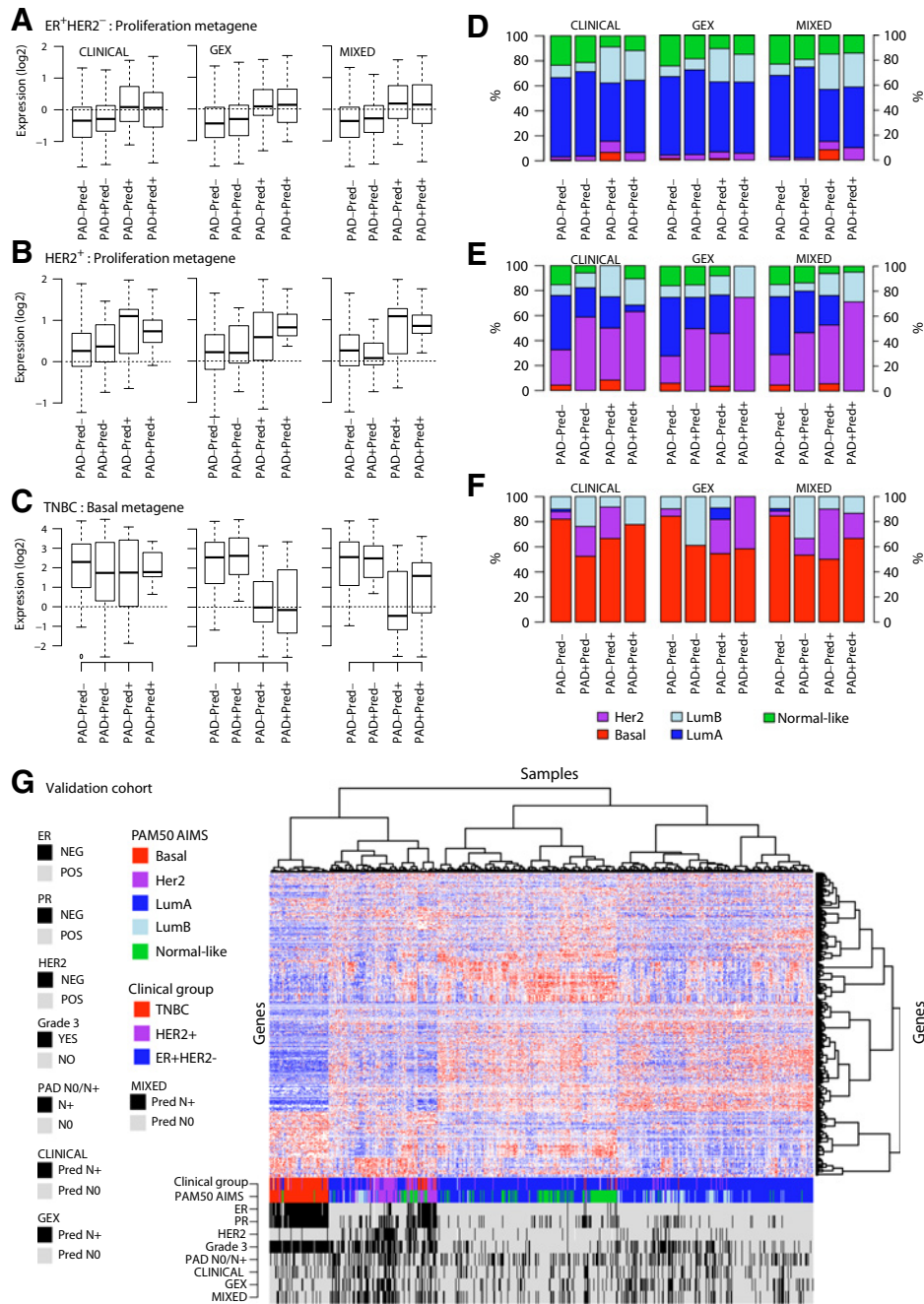


Figure 3. Associations of derived nodal predictors with transcriptional patterns and subtypes in breast cancer. Four nodal status subgroups were evaluated, stratified by combinations of pathological nodal status, NO/N+ (PAD−/PAD+) and model-predicted axillary status, NO/N+ (Pred−/Pred+) in the validation cohort. The boxplots in **A–C** show expression of metagenes, and the barplots in **D–F** show PAM50 classification for the CLINICAL (left), GEX (center), and MIXED (right) models within the evaluation groups. **A** and **D**, Expression of the checkpoint proliferation metagene and PAM50 classification within ER+HER2- evaluation group. **B** and **E**, Expression of the checkpoint proliferation metagene and PAM50 classification within HER2+ evaluation group. **C** and **F**, Expression of a basal metagene, and distribution of PAM50 subtypes within the TNBC evaluation group. **G**, Unsupervised hierarchical clustering of the top 500 differentially expressed genes in the entire validation cohort reveals the heterogeneous distribution of pathology and predicted nodal status in relation to clinicopathological variables, PAM50 subtypes, and evaluation subgroups as defined by surrogate definitions. Clustering was performed using Pearson correlation and Ward's linkage. In the heatmap, expression ranges from low (blue, log₂ ratio = −3) to high (red, log₂ ratio = 3) for genes (rows) across samples (columns). Basal, basal-like; CLINICAL, predictor based on clinicopathological features alone; GEX, predictor based on RNA-seq data alone; HER2, HER2-enriched; LumA, luminal A; LumB, luminal B; normal, normal-like; MIXED, predictor based on both clinicopathological parameters and RNA-seq data; NO, no regional lymph node metastasis; N+, axillary lymph node metastasis; PAD, pathology-defined nodal status; PAM50 AIMS, Absolute Assignment of breast cancer intrinsic molecular subtype by PAM50 assay; Pred, model-predicted nodal status.

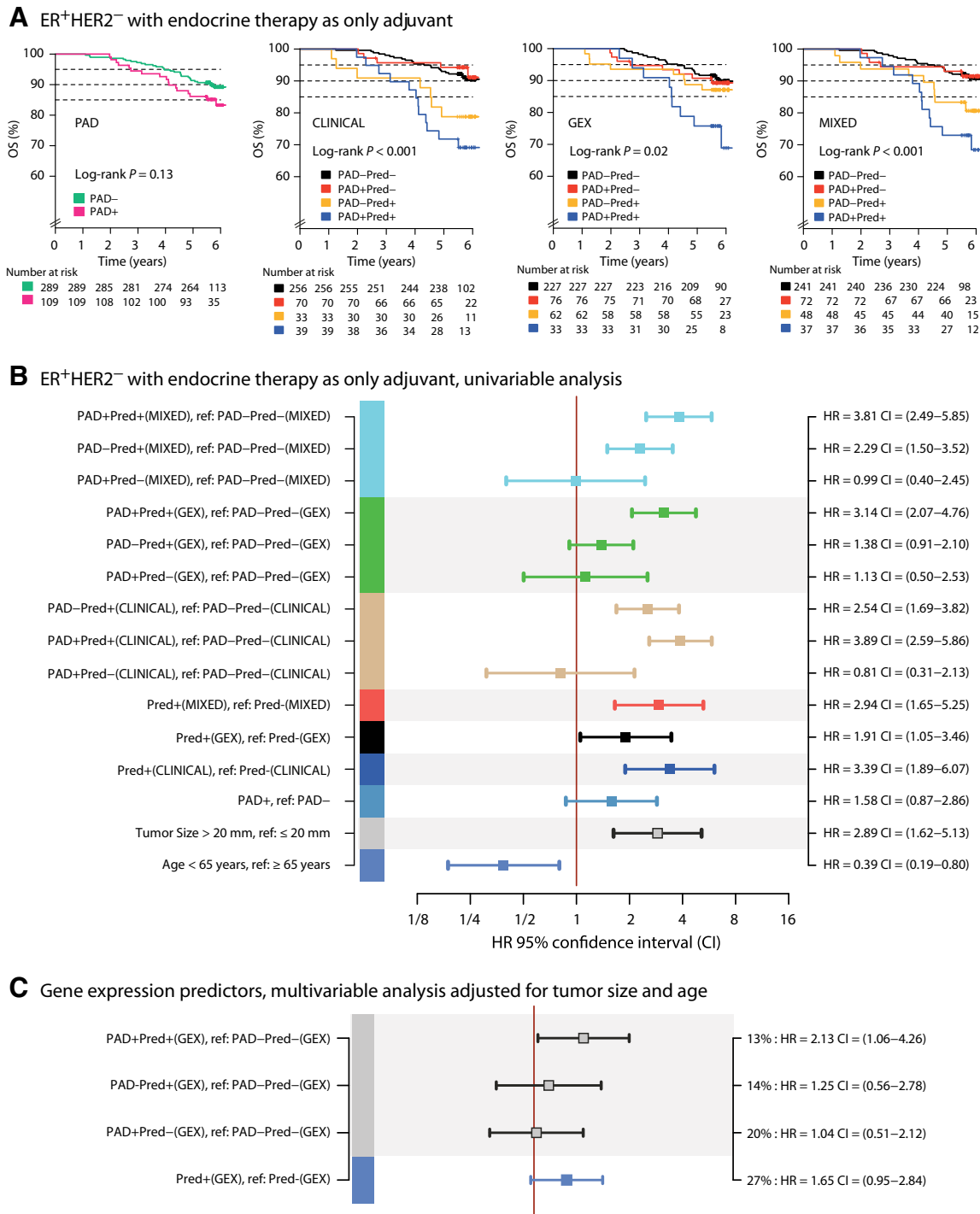


Figure 4. Association of derived nodal predictors with outcome in ER⁺HER2⁻ disease treated with adjuvant endocrine therapy. Shown here is the estimation of prognosis in the largest evaluation group in the validation set, which included patients with ER⁺HER2⁻ tumors receiving endocrine therapy as the only adjuvant treatment. **A**, Kaplan-Meier curves and log-rank *P* values of overall survival (OS) are displayed for nodal status subgroups, stratified by pathological nodal status, NO/N+ (PAD-/PAD+) alone, and in combination with predicted axillary status, NO/N+ (Pred-/Pred+), in the CLINICAL, GEX, or MIXED models. **B**, Forest plot displaying HRs of univariable Cox regression for age, tumor size, and node-positivity versus node-negativity, as predicted by any of the three models, and the combinations of the pathologic and predicted axillary status. **C**, Multivariable analysis adjusted for tumor size (mm) and age (years) for the GEX predictor. Percentages indicate respective group size compared with analyzed samples in the analysis. CI, confidence interval; CLINICAL, predictor based on clinicopathologic features alone; GEX, predictor based on RNA-seq data alone; MIXED, predictor based on both clinicopathologic parameters and RNA-seq data; PAD, pathology-defined nodal status; Pred, model-predicted nodal status; ref, reference group.

alone and the presented SLNB reduction rate was higher for the MIXED predictor compared to the CLINICAL predictor. Notably, the AUCs attained in our models are similar to those of prediction models previously published for nodal metastasis (AUC 0.67–0.78), which included only clinicopathologic data (18–21, 30), but did not report on possible SLNB reduction rate.

Although the addition of gene expression data to existing clinicopathologic variables did not show a clear superiority in predicting nodal status, it should be noted that machine-learning approaches identify the most dominant trait for each predictor type and capture different metastatic features. Even though the discriminatory performances (AUC) of the CLINICAL and MIXED predictors were identical, there was only a moderate agreement between them in the validation cohort (Supplementary Fig. S3). Previous studies investigating prediction of nodal metastasis in breast cancer have reported diverse performances of GEX-based predictors, with AUCs between chance level to near perfect separation (23–28), likely to be due to differences in patient characteristics, cohort sizes, definition of nodal disease, gene expression analysis platforms, and feature-selection strategies. In other malignancies, nodal prediction accuracies between 50% to 60% have been reported based on TCGA data (28). Our predictors containing gene expression data (GEX and MIXED models) revealed AUCs of 0.58 to 0.72 during validation and were consistent with a mid-range performance as defined in previous publications (23–28). The current results are, however, validated in an unselected contemporary large primary breast cancer cohort. Still, although the small proportion of HER2⁺ and TNBC tumors reflect the breast cancer population in Sweden (31), results from these evaluation groups must be interpreted with caution.

One problem with developing mixed gene expression classifiers for the classification question at hand is the scarcity of larger cohorts including gene expression- and relevant clinicopathologic data that is commonly used in, for example, nomogram methods (such as lymphovascular invasion, mode of detection, multifocality and tumor location; refs. 18, 30). To further validate the MIXED classifier for ER⁺HER2⁻ tumors, we used the TCGA dataset (44) comprising 503 primary breast tumors. Despite considerable imputation of missing values for the important clinicopathologic parameters and caution must be exercised when interpreting these results, we found an AUC of 0.68 compared with 0.72 in our SCAN-B 2011 validation cohort. This is in line with the decline in performance reported for the MSKCC nomogram in an independent cohort (Supplementary information, Appendix S1; ref. 20). Corresponding possible SLNB reduction rates for patients with ER⁺HER2⁻ tumors within the TCGA cohort would be 12% to 18% if applying the MIXED predictor and accepting false negative rates of 5% to 10%. Based only on the AUC values for N0/N+ prediction by MIXED and CLINICAL models, differences appear too small to justify the added use of RNAseq. However, when considering the SLNB reduction rate as well, there is a clear indication that the MIXED model is able to identify more node negative patients than the CLINICAL model (exemplified by 7% absolute improvement in reduction rate for patients with ER⁺HER2⁻ tumors while accepting a false negative rate of 5%). With declining prices of GEX assays and the potential to combine different prognostic or predictive gene signatures in one transcriptional analysis, even a minor improvement of SLNB reduction rate in favor of the MIXED predictor can potentially be translated into cost effectiveness in the future.

Although the discrimination of nodal metastasis is challenging due to the heterogeneity of breast cancer, AUC analysis allows for different cutoffs to be assigned, determined by the prevalent clinical setting (45). For the purpose of identifying patients with very low risk of nodal metastasis to omit surgical axillary staging (SLNB), cutoffs were set to achieve the highest possible negative predictive value (Max NPV) in the prediction of N0. Moreover, additional cut-offs were set to reflect accepted FNR of the SLNB procedure (FNR 5%–10%) in itself, which is the golden standard for axillary staging. Allowing FNR 5% to 10% for N0 prediction, CLINICAL predictors and MIXED predictors would yield SLNB reduction rates of 11% to 23% (FNR 5%) and 19% to 30% (FNR 10%), respectively, supporting that the MIXED predictors identified more node-negative patients compared with the CLINICAL predictors. The importance of gene expression profiles for adjuvant treatment recommendation is today readily accepted, but not included as a tool for selection of optimal surgical strategy (11, 12, 15). If prospectively validated, integration of gene signatures in the preoperative diagnostic workup might reduce the rates of SLNB for patients at low risk of nodal involvement.

In line with reports on clinicopathologic variables associated with nodal metastasis (18, 21, 46, 47), tumor size and LVI were key variables in our CLINICAL and MIXED models for the ER⁺HER2⁻ and HER2⁺ groups, but not for TNBC tumors. Interestingly, TNBC is characterized by a low risk of nodal metastasis and displays a non-linear association between tumor size and the risk of nodal engagement (48, 49). We observed low expression of basal-like markers for TNBC tumors with predicted nodal metastasis by the GEX and MIXED models, but not for the CLINICAL model. In accordance with TNBC showing the lowest prediction overlap between CLINICAL and MIXED models (Supplementary Fig. S3A), five genes ranked higher in the MIXED TNBC model (*MAGEA4*, *FUT3*, *MIA*, *LPAR2*, *LOXL4*) than did LVI and tumor size. Interestingly, *LOXL4* expression has been shown to be associated with tumor growth and lung metastasis in TNBC (50). The fair to moderate overlap between predictions by the CLINICAL, MIXED, and GEX models derived from the same samples indicates that these models capture different aspects of the N0/N+ disease, exaggerated in TNBC tumors when including gene expression patterns. Consequently, a great molecular heterogeneity must exist in each evaluation group given the chosen nodal classification endpoint.

To test the hypothesis of heterogeneity within the N0/N+ endpoint, we performed an unsupervised analysis of the validation cohort, demonstrating transcriptional heterogeneity of axillary status across the molecular breast cancer subtypes (Fig. 3G). Specifically, the N0/N+ status may not significantly reflect transcriptional variations in breast cancer and consequently presents a challenge for expression-based predictors. To substantiate this claim, we used a machine-learning procedure that included a feature selection step to identify genes differentially expressed between N0 and N+ in each evaluation group, with adjustment for multiple testing. For the MIXED classifiers this analysis identified 1268 differentially expressed genes in ER+HER2- tumors, 74 genes in HER2⁺, and 130 genes in TNBC (Supplementary Fig. S2). The low numbers of differentially expressed genes in the HER2⁺ and TNBC evaluation groups did not support a strong gene expression signal. Additionally, functional pathway analysis of the predictor gene sets showed heterogeneity in enriched terms, consistent with the observed transcriptional heterogeneity of the N0/N+ status. This finding is not surprising due to the limited

number of genes retained after the retraining step (Fig. 1B) and the feature selection/optimization process of machine learning methods. However, for the GEX and MIXED predictor gene lists prior to the retraining step that excluded genes with weight ≤ 0 for the final locked models, we found a notable enrichment of proliferation-associated gene ontology terms in the ER⁺HER2⁻ and HER2⁺ subgroups (Supplementary Fig. S2). This is consistent with (i) the differential expression of the proliferation metagene, stratifying tumors based on predictor classifications rather than on pathologic status, and (ii) the variations in proportion of luminal B classified tumors for the same groups (Fig. 3), which could explain the observed prognostic disparities (Fig. 4).

The recognition of proliferation-related prognostic transcriptional programs, (51–55) especially in luminal breast cancer, prompted us to investigate the predictors' prognostic performance (for which they were not trained) in the largest subgroup in the validation cohort: ER⁺HER2⁻ tumors treated with adjuvant endocrine therapy only. In line with the expression pattern of the proliferation metagene (Fig. 3), the poorest outcome was observed for patients with predicted nodal disease involvement (Pred+), with worst OS for those with both pathologic and predicted nodal positivity (PAD+Pred+; Fig. 4A and B). This suggests that at least for patients with ER⁺HER2⁻ tumors PAD does not add prognostic value when nodes are predicted by the models to be negative. The weak association between pathology-based lymph node status and outcome is likely due to the short follow-up time and use of OS as clinical endpoint. In multivariable analysis adjusted for tumor size and age, the GEX PAD+Pred+ group remained significant. However, this did not prevail after further adjustment for grade. Therefore, the prognostic impact obtained from the predictors' proliferative features is related to the prognostic importance of histological grading. The limited cases within other evaluation groups precluded evaluation.

The current study has a number of limitations. First, clinicopathologic characteristics were obtained from a national registry, with incomplete data for some variables that could otherwise be included in predictor training or be used in defining evaluation groups, for example, surrogate molecular subtype or Ki67 scores. Moreover, despite analyzing >3,000 cases, the number of HER2⁺ and TNBC tumors in our population-based cohort was limited and precluded assessment in more defined subgroups (e.g., HER2⁺ER⁻). Consequently, it may be hypothesized that a more refined molecular subgrouping of breast cancer by means of intrinsic signatures could improve the performance of gene expression-based classifiers. Second, our outcome analyses are limited by the type of endpoint available with complete coverage (OS) and the relatively short follow-up time of the prospective SCAN-B cohort. Finally, although we exhaustively evaluated seven different machine-learning models and different feature-selection approaches, we acknowledge that additional optimization and combinations of clinicopathological and transcriptional data could improve the results. Although the latter may be technically achievable, a highly relevant issue in moving gene expression predictors towards the clinic is the ability to predict single samples preferably without the need of reference sets, representing a limitation of our current models. Here, reports of methods deriving gene rule-based predictors (37, 56) may represent an alternative machine-learning approach, provided sufficient performance and adequate training numbers (57).

In summary, this study provides a first estimate of the performance of predictors including RNA-seq data as well as clinicopathologic data of axillary nodal status in a population-based setting and proposes possible SLNB reduction rates by applying predictors to discriminate N0. CLINICAL and MIXED predictors of nodal metastasis had comparable accuracy in line with previously published predictors, while more node-negative patients were identified by the MIXED predictors. The GEX models had inferior predictive performance, underscoring that the complexity of lymphatic spread dissemination is not only related to gene expression of the primary tumor. To fully harness the power of transcriptional profiling for nodal prediction, a more precise stratification of breast cancer subgroups also appears relevant. Given such improvements, predictors such as the MIXED classifier, combining clinicopathologic variables and gene expression data may develop into clinically relevant lymph node predictor with putative prognostic value.

Disclosure of Potential Conflicts of Interest

A. Ehinger reports receiving speakers bureau honoraria from Roche, Amgen, and Novartis. Å. Borg reports receiving speakers bureau honoraria from AstraZeneca and Roche. No potential conflicts of interest were disclosed by the other authors.

Authors' Contributions

Conception and design: L. Dihge, P.-O. Bendahl, Å. Borg, J. Staaf, L. Rydén
Development of methodology: L. Dihge, J. Vallon-Christersson, J. Staaf, L. Rydén

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): J. Vallon-Christersson, C. Hegardt, L.H. Saal, C. Larsson, A. Ehinger, N. Loman, M. Malmberg, L. Rydén

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): L. Dihge, J. Vallon-Christersson, L.H. Saal, J. Häkkinen, M. Malmberg, P.-O. Bendahl, J. Staaf, L. Rydén

Writing, review, and/or revision of the manuscript: L. Dihge, J. Vallon-Christersson, C. Hegardt, L.H. Saal, J. Häkkinen, C. Larsson, A. Ehinger, N. Loman, M. Malmberg, P.-O. Bendahl, Å. Borg, J. Staaf, L. Rydén

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): J. Vallon-Christersson, C. Hegardt, J. Häkkinen, M. Malmberg, J. Staaf, L. Rydén

Study supervision: J. Staaf, L. Rydén

Acknowledgments

The authors would like to acknowledge patients and clinicians participating in the SCAN-B study, all personnel at the central SCAN-B laboratory at the Division of Oncology and Pathology, Lund University, the South Swedish Breast Cancer Group (SSBCG), the Swedish National Breast Cancer Quality Registry (NKBC), The Regional Biobank Centre (RBC) Syd and the Regional Cancer Centre South (RCC Syd). This study was supported by grants from The Swedish Cancer Society (CAN 2016/659, CAN 2018/685), The Skåne County Councils Research and Development Foundation (2016/623611), The Governmental Funding of Clinical Research within the National Health Service (ALF) (2014/434901, 2018/40612, 2019/40304), The Swedish Breast Cancer Association (2017/1213), Mrs. Berta Kamprad Foundation (FBKS 2018-3-166), Mats Paulsson Foundation for Research, Innovation and Community Development (IACD 2017), The Crafoord Foundation (20180543), and The Erling-Persson Family Foundation (2018/1008). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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 January 10, 2019; revised May 24, 2019; accepted July 22, 2019; published first July 24, 2019.

References

- Giuliano AE, Hunt KK, Ballman KV, Beitsch PD, Whitworth PW, Blumentcranz PW, et al. Axillary dissection vs no axillary dissection in women with invasive breast cancer and sentinel node metastasis: a randomized clinical trial. *JAMA* 2011;305:569–75.
- Galimberti V, Cole BF, Zurrada S, Viale G, Luini A, Veronesi P, et al. Axillary dissection versus no axillary dissection in patients with sentinel-node micrometastases (IBCSG 23–01): a phase 3 randomised controlled trial. *Lancet Oncol* 2013;14:297–305.
- Donker M, van Tienhoven G, Straver ME, Meijnen P, van de Velde CJ, Mansel RE, et al. Radiotherapy or surgery of the axilla after a positive sentinel node in breast cancer (EORTC 10981–22023 AMAROS): a randomised, multicentre, open-label, phase 3 non-inferiority trial. *Lancet Oncol* 2014;15:1303–10.
- Gentilini O, Veronesi U. Abandoning sentinel lymph node biopsy in early breast cancer? A new trial in progress at the European Institute of Oncology of Milan (SOUND: Sentinel node vs Observation after axillary Ultra-Sound). *Breast* 2012;21:678–81.
- Giuliano AE, Ballman KV, McCall L, Beitsch PD, Brennan MB, Kelemen PR, et al. Effect of axillary dissection vs no axillary dissection on 10-year overall survival among women with invasive breast cancer and sentinel node metastasis: the ACOSOG Z0011 (Alliance) randomized clinical trial. *JAMA* 2017;318:918–26.
- Chen SL, Hoehne FM, Giuliano AE. The prognostic significance of micrometastases in breast cancer: a SEER population-based analysis. *Ann Surg Oncol* 2007;14:3378–84.
- Carter CL, Allen C, Henson DE. Relation of tumor size, lymph node status, and survival in 24,740 breast cancer cases. *Cancer* 1989;63:181–7.
- 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.
- Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 2001;98:10869–74.
- O'Brien KM, Cole SR, Tse CK, Perou CM, Carey LA, Foulkes WD, et al. Intrinsic breast tumor subtypes, race, and long-term survival in the Carolina Breast Cancer Study. *Clin Cancer Res* 2010;16:6100–10.
- Harris LN, Ismaila N, McShane LM, Andre F, Collyar DE, Gonzalez-Angulo AM, 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. *J Clin Oncol* 2016;34:1134–50.
- Krop I, Ismaila N, Andre F, Bast RC, Barlow W, Collyar DE, 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;35:2838–47.
- Sparano JA, Gray RJ, Makower DF, Pritchard KI, Albain KS, Hayes DF, et al. Prospective validation of a 21-gene expression assay in breast cancer. *N Engl J Med* 2015;373:2005–14.
- Gluz O, Nitz UA, Christgen M, Kates RE, Shak S, Clemens M, et al. West German Study Group phase III PlanB trial: first prospective outcome data for the 21-Gene Recurrence Score assay and concordance of prognostic markers by central and local pathology assessment. *J Clin Oncol* 2016;34:2341–9.
- Cardoso F, van't Veer LJ, Bogaerts J, Slaets L, Viale G, Delaloge S, et al. 70-Gene signature as an aid to treatment decisions in early-stage breast cancer. *N Engl J Med* 2016;375:717–29.
- Ribnikar D, Cardoso F. Tailoring chemotherapy in early-stage breast cancer: based on tumor biology or tumor burden? *Am Soc Clin Oncol Educ Book* 2016;35:e31–8.
- Gajdos C, Tartert PI, Bleiweiss JJ. Lymphatic invasion, tumor size, and age are independent predictors of axillary lymph node metastases in women with T1 breast cancers. *Ann Surg* 1999;230:692–6.
- Bevilacqua JL, Kattan MW, Fey JV, Cody HS III, Borgen PI, Van Zee KJ. Doctor, what are my chances of having a positive sentinel node? A validated nomogram for risk estimation. *J Clin Oncol* 2007;25:3670–9.
- Klar M, Foeldi M, Markert S, Gitsch G, Stickeler E, Watermann D. Good prediction of the likelihood for sentinel lymph node metastasis by using the MSKCC nomogram in a German breast cancer population. *Ann Surg Oncol* 2009;16:1136–42.
- van la Parra RF, Francissen CM, Peer PG, Ernst MF, de Roos WK, Van Zee KJ, et al. Assessment of the Memorial Sloan-Kettering Cancer Center nomogram to predict sentinel lymph node metastases in a Dutch breast cancer population. *Eur J Cancer* 2013;49:564–71.
- Viale G, Zurrada S, Maiorano E, Mazzarol G, Pruneri G, Paganelli G, et al. Predicting the status of axillary sentinel lymph nodes in 4351 patients with invasive breast carcinoma treated in a single institution. *Cancer* 2005;103:492–500.
- Chen JY, Chen JJ, Yang BL, Liu ZB, Huang XY, Liu GY, et al. Predicting sentinel lymph node metastasis in a Chinese breast cancer population: assessment of an existing nomogram and a new predictive nomogram. *Breast Cancer Res Treat* 2012;135:839–48.
- West M, Blanchette C, Dressman H, Huang E, Ishida S, Spang R, et al. Predicting the clinical status of human breast cancer by using gene expression profiles. *Proc Natl Acad Sci U S A* 2001;98:11462–7.
- Huang E, Cheng SH, Dressman H, Pittman J, Tsou MH, Hornig CF, et al. Gene expression predictors of breast cancer outcomes. *Lancet* 2003;361:1590–6.
- Smeets A, Daemen A, Vanden Bempt I, Gevaert O, Claes B, Wildiers H, et al. Prediction of lymph node involvement in breast cancer from primary tumor tissue using gene expression profiling and miRNAs. *Breast Cancer Res Treat* 2011;129:767–76.
- Shriver CD, Hueman MT, Ellsworth RE. Molecular signatures of lymph node status by intrinsic subtype: gene expression analysis of primary breast tumors from patients with and without metastatic lymph nodes. *J Exp Clin Cancer Res* 2014;33:116.
- Nakauchi C, Naoi Y, Shimazu K, Tsunashima R, Nishio M, Maruyama N, et al. Development of a prediction model for lymph node metastasis in luminal A subtype breast cancer: the possibility to omit sentinel lymph node biopsy. *Cancer Lett* 2014;353:52–8.
- Tarabichi M, Saiselet M, Trésallet C, Hoang C, Larsimont D, Andry G, et al. Revisiting the transcriptional analysis of primary tumours and associated nodal metastases with enhanced biological and statistical controls: application to thyroid cancer. *Br J Cancer* 2015;112:1665–74.
- Simon R. Roadmap for developing and validating therapeutically relevant genomic classifiers. *J Clin Oncol* 2005;23:7332–41.
- Dihge L, Bendahl PO, Ryden L. Nomograms for preoperative prediction of axillary nodal status in breast cancer. *Br J Surg* 2017;104:1494–505.
- Regionala cancercentrum i samverkan, Nationellt kvalitetsregister för bröstcancer och bröstrekonstruktion. 2010–2015. National Quality Registry for Breast Cancer. [cited 2017 Jan 3]. Available from: <https://www.cancercentrum.se/samverkan/cancerdiagnoser/brost/kvalitetsregister/>.
- Saal LH, Vallon-Christersson J, Hakkinen J, Hegardt C, Grabau D, Winter C, et al. The Sweden Cancerome Analysis Network - Breast (SCAN-B) initiative: a large-scale multicenter infrastructure towards implementation of breast cancer genomic analyses in the clinical routine. *Genome Med* 2015;7:20.
- Ryden L, Loman N, Larsson C, Hegardt C, Vallon-Christersson J, Malmberg M, et al. Minimizing inequality in access to precision medicine in breast cancer by real-time population-based molecular analysis in the SCAN-B initiative. *Br J Surg* 2018;105:e158–e68.
- Statistiska centralbyrån. 2017. Statistics Sweden. Available from: <http://www.scb.se/en/>.
- Regionala cancercentrum i samverkan. 2017. Swedish National Guidelines for Breast Cancer. [cited 2017 Dec 29]. Available from: <https://www.cancercentrum.se/samverkan/cancerdiagnoser/brost/vardprogram/gallandevardprogram/>.
- Brueffer C, Vallon-Christersson J, Grabau D, Ehinger A, Häkkinen J, Hegardt C, et al. Clinical value of RNA sequencing-based classifiers for prediction of the five conventional breast cancer biomarkers: a report from the population-based multicenter Sweden Cancerome Analysis Network Breast initiative. *JCO Precision Oncology* 2018;2:1–18.
- Paquet ER, Hallett MT. Absolute assignment of breast cancer intrinsic molecular subtype. *J Natl Cancer Inst* 2015;107.
- Fredlund E, Staaf J, Rantala JK, Kallioniemi O, Borg A, Ringner M. The gene expression landscape of breast cancer is shaped by tumor protein p53 status and epithelial-mesenchymal transition. *Breast Cancer Res* 2012;14:R113.

39. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2016.
40. Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. *J Med Internet Res* 2016;18:e323.
41. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ* 2015;351:h5527.
42. Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat* 2001;29:1189–232.
43. Miller KD, Siegel RL, Lin CC, Mariotto AB, Kramer JL, Rowland JH, et al. Cancer treatment and survivorship statistics, 2016. *CA Cancer J Clin* 2016;66:271–89.
44. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* 2012;490:61–70.
45. Florkowski CM. Sensitivity, specificity, receiver-operating characteristic (ROC) curves and likelihood ratios: communicating the performance of diagnostic tests. *Clin Biochem Rev* 2008;29 Suppl 1:S83–7.
46. Meretoja TJ, Heikkilä PS, Mansfield AS, Cserni G, Ambrozay E, Boross G, et al. A predictive tool to estimate the risk of axillary metastases in breast cancer patients with negative axillary ultrasound. *Ann Surg Oncol* 2014;21:2229–36.
47. Reyal F, Rouzier R, Depont-Hazelzet B, Bollet MA, Pierga JY, Alran S, et al. The molecular subtype classification is a determinant of sentinel node positivity in early breast carcinoma. *PLoS One* 2011;6:e20297.
48. Hernandez-Aya LF, Chavez-Macgregor M, Lei X, Meric-Bernstam F, Buchholz TA, Hsu L, et al. Nodal status and clinical outcomes in a large cohort of patients with triple-negative breast cancer. *J Clin Oncol* 2011;29:2628–34.
49. Wang XX, Jiang YZ, Li JJ, Song CG, Shao ZM. Effect of nodal status on clinical outcomes of triple-negative breast cancer: a population-based study using the SEER 18 database. *Oncotarget* 2016;7:46636–45.
50. Choi SK, Kim HS, Jin T, Moon WK. LOXL4 knockdown enhances tumor growth and lung metastasis through collagen-dependent extracellular matrix changes in triple-negative breast cancer. *Oncotarget* 2017;8:11977–89.
51. Perreard L, Fan C, Quackenbush JF, Mullins M, Gauthier NP, Nelson E, et al. Classification and risk stratification of invasive breast carcinomas using a real-time quantitative RT-PCR assay. *Breast Cancer Res* 2006;8:R23.
52. Wirapati P, Sotiriou C, Kunkel S, Farmer P, Pradervand S, Haibe-Kains B, et al. Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. *Breast Cancer Res* 2008;10:R65.
53. Gingras I, Desmedt C, Ignatiadis M, Sotiriou C. CCR 20th Anniversary commentary: gene-expression signature in breast cancer—where did it start and where are we now? *Clin Cancer Res* 2015;21:4743–6.
54. Gatz ML, Silva GO, Parker JS, Fan C, Perou CM. An integrated genomics approach identifies drivers of proliferation in luminal-subtype human breast cancer. *Nat Genet* 2014;46:1051–9.
55. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 2011;144:646–74.
56. Afsari B, Fertig EJ, Geman D, Marchionni L. switchBox: an R package for k-Top Scoring Pairs classifier development. *Bioinformatics* 2015;31:273–4.
57. Cirenajwis H, Lauss M, Planck M, Vallon-Christersson J, Staaf J. Performance of gene expression-based single sample predictors for assessment of clinicopathological subgroups and molecular subtypes in cancers: a case comparison study in non-small cell lung cancer. *Brief Bioinform* 2019 Feb 4 [Epub ahead of print].