# The Humoral Immune System Has a Key Prognostic Impact in Node-Negative Breast Cancer

Marcus Schmidt,[1] Daniel Böhm,[1] Christian von Törne,[2] Eric Steiner,[1] Alexander Puhl,[1] Henryk Pilch,[3] Hans-Anton Lehr,[5] Jan G. Hengstler,[4] Heinz Kölbl,[1] and Mathias Gehrmann[2]

[1]Department of Obstetrics and Gynecology, Medical School, Johannes Gutenberg University, Mainz, Germany; [2]Siemens Medical Solutions Diagnostics GmbH, Cologne, Germany; [3]Department of Obstetrics and Gynecology, and [4]Center for Toxicology, Institute of Legal Medicine and Rudolf-Boehm Institute of Pharmacology and Toxicology, University of Leipzig, Leipzig, Germany; and [5]Department of Pathology, University of Lausanne, Lausanne, Switzerland

## Abstract

Estrogen receptor (ER) expression and proliferative activity are established prognostic factors in breast cancer. In a search for additional prognostic motifs, we analyzed the gene expression patterns of 200 tumors of patients who were not treated by systemic therapy after surgery using a discovery approach. After performing hierarchical cluster analysis, we identified coregulated genes related to the biological process of proliferation, steroid hormone receptor expression, as well as B-cell and T-cell infiltration. We calculated metagenes as a surrogate for all genes contained within a particular cluster and visualized the relative expression in relation to time to metastasis with principal component analysis. Distinct patterns led to the hypothesis of a prognostic role of the immune system in tumors with high expression of proliferation-associated genes. In multivariate Cox regression analysis, the proliferation metagene showed a significant association with metastasis-free survival of the whole discovery cohort [hazard ratio (HR), 2.20; 95% confidence interval (95% CI), 1.40–3.46]. The B-cell metagene showed additional independent prognostic information in carcinomas with high proliferative activity (HR, 0.66; 95% CI, 0.46–0.97). A prognostic influence of the B-cell metagene was independently confirmed by multivariate analysis in a first validation cohort enriched for high-grade tumors ($n$ = 286; HR, 0.78; 95% CI, 0.62–0.98) and a second validation cohort enriched for younger patients ($n$ = 302; HR, 0.83; 95% CI, 0.7–0.97). Thus, we could show in three cohorts of untreated, node-negative breast cancer patients that the humoral immune system plays a pivotal role in metastasis-free survival of carcinomas of the breast. [Cancer Res 2008;68(13):5405–13]

## Introduction

Estrogen receptor (ER) expression and proliferative activity of breast carcinomas have long been established as prognostic markers. Patients with ER-positive carcinomas have a better prognosis than those with ER-negative carcinomas (1) and rapidly proliferating carcinomas have an adverse prognosis (2). Knowledge about the molecular mechanisms involved in the processes of estrogen-dependent tumor growth and proliferative activity has led to the successful development of therapeutic concepts (i.e., antiendocrine and cytotoxic chemotherapy).

The recent advent of gene expression profiling has allowed researchers to venture into the heterogeneous nature of breast cancer (3–5), and several research groups have identified sets of genes differentially expressed in breast carcinomas that progress to metastasize within 5 years (6–9). Although there is hardly any overlap between the identified list of genes, it is becoming increasingly apparent that most prognostic and predictive classification algorithms rely heavily on ERα-regulated genes, as well as genes involved in the cell cycle (10–14). In particular, tumors scored as ER positive by immunohistochemistry can be subdivided into good outcome and bad outcome groups by proliferation-associated genes (15, 16). In a recent meta-analysis, an immune response gene motif has been shown to distinguish between patients with good and bad outcome in ER-negative tumors (17). In addition, it has been found that high expression of lymphocyte-associated genes confers a good prognosis in node-negative ERBB2-positive breast cancer (18).

Peritumoral lymphocytic infiltration has long been suggested to influence clinical outcome (19). In particular, medullary breast cancer is characterized by both prominent lymphocytic infiltrates and a relatively good prognosis, despite a lack of ER expression and poor histologic grade (20). To systematically evaluate the prognostic effect of the immune system, we performed gene profiling in a discovery cohort study of node-negative, untreated breast carcinomas.

## Materials and Methods

**Patient characteristics and tissue specimens.** The population-based cohort study consists of 200 consecutive lymph node-negative breast cancer patients treated at the Department of Obstetrics and Gynecology of the Johannes Gutenberg University Mainz between 1988 and 1998. Patients were all treated with surgery and did not receive any systemic therapy in the adjuvant setting (Table 1). The established prognostic factors (histologic grade, tumor size, age at diagnosis, and steroid receptor status) were collected from the original pathology reports of the gynecologic pathology division within our department. Patients were treated with either modified radical mastectomy ($n$ = 75) or breast-conserving surgery followed by irradiation ($n$ = 125) and were without evidence of regional lymph node and distant metastasis at the time of surgery. The median age of the patients at surgery was 60 y (range, 34–89 y). The median time of follow up was 92 mo. For all tumors, samples were snap frozen and stored at $-80°$C. Tumor cell content exceeded 40% in all samples. Approximately 50 mg of frozen breast tumor tissue were crushed in liquid nitrogen. RLT buffer was added and the homogenate was centrifuged through a QIAshredder column (Qiagen). From the eluate, total RNA was isolated with the RNeasy Kit (Qiagen) according to the manufacturer's instructions. RNA yield was determined

**Table 1.** Patient characteristics of the Mainz (*n* = 200), Rotterdam (*n* = 286), and TRANSBIG (*n* = 302) data sets

| | Mainz<br>No. of patients (%) | Rotterdam<br>No. of patients (%) | TRANSBIG<br>No. of patients (%) |
|---|---|---|---|
| Tumor size | | | |
| $T_1$ | 111 (56) | 146 (51) | 163 (54) |
| $T_2$ | 81 (40) | 132 (46) | 138 (45.7) |
| $T_{3/4}$ | 8 (4) | 8 (3) | 1 (0.3) |
| Tumor grade | | | |
| Well differentiated | 41 (21) | 7 (2) | 60 (20) |
| Moderately differentiated | 110 (55) | 42 (15) | 117 (39) |
| Poor/undifferentiated | 45 (23) | 148 (52) | 106 (35) |
| Unknown | 4 (2) | 89 (31) | 19 (6) |
| ER | EIA | DCC or EIA | |
| Negative | 44 (22) | 77 (27) | 95 (31) |
| Positive | 156 (78) | 209 (73) | 201 (67) |
| Unknown | | | 6 (2) |
| PR | EIA | DCC or EIA | |
| Negative | 70 (35) | 111 (39) | 6 (2) |
| Positive | 130 (65) | 165 (58) | 56 (19) |
| Unknown | | 10 (3) | 240 (79) |
| Age, y | | | |
| Mean | 60 (SD, 12) | 54 (SD, 12) | 49 (SD, 9) |
| ≤40 | 10 (5) | 36 (13) | 53 (18) |
| 41–55 | 64 (32) | 129 (45) | 175 (58) |
| 56–70 | 83 (42) | 89 (31) | 71 (23) |
| ≥70 | 43 (22) | 32 (11) | 3 (1) |
| Metastasis within 5 y | | | |
| Yes | 28 (14) | 93 (33) | 56 (18) |
| No | 154 (77) | 183 (64) | 198 (66) |
| Censored | 18 (9) | 10 (3) | 27 (9) |
| Metastasis after 5 y | 18 (9) | | 21 (7) |

NOTE: The Mainz collection was population based whereas the Rotterdam cohort was selected for a case-control study [Wang et al. (7)]. Note the differences in tumor grade and patient age between the cohorts.

Abbreviation: PR, progesterone receptor.

by UV absorbance, and RNA quality was assessed by analysis of rRNA band integrity on an Agilent 2100 Bioanalyzer RNA 6000 LabChip kit (Agilent Technologies). The study was approved by the ethical review board of the medical association of Rhineland-Palatinate [no. 837.139.05 (4797)].

**Gene expression profiling.** The Affymetrix HG-U133A array and GeneChip System was used to quantify the relative transcript abundance in the breast cancer tissues. Starting from 5-μg total RNA, labeled cRNA was prepared using the Roche Microarray cDNA Synthesis, Microarray RNA Target Synthesis (T7) and Microarray Target Purification Kit according to the manufacturer's instructions. Raw.cel file data were processed by MAS 5.0 software. In the analysis settings, the global scaling procedure was chosen, which multiplied the output signal intensities of each array to a mean target intensity (TGT) of 500. Samples with suboptimal average signal intensities (i.e., scaling factors >25) or glyceraldehyde-3-phosphate dehydrogenase 3′/5′ ratios >5 were relabeled and rehybridized on new arrays. Raw.cel file, MAS 5.0 processed and patient data have been deposited in National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO)[6] and are accessible through GEO Series accession no. GSE11121.

**Previously published microarray data sets.** Two breast cancer Affymetrix HG-U133A microarray data sets including patient outcome

information were downloaded from the NCBI GEO data repository.[6] The first data set (GSE2034) represents 180 lymph node–negative relapse free patients and 106 lymph node–negative patients that developed a distant metastasis. None of these patients had received systemic neoadjuvant or adjuvant therapy (Rotterdam cohort). The.txt file data were recalculated to a TGT of 500. The second data set consists of 302 samples from breast cancer patients that remained untreated in the adjuvant setting after surgery (TRANSBIG cohort). GSM numbers of samples ( from GSE6532 and GSE7390) used for analysis are listed in Supplementary Table S1. Raw.cel file data were processed by MAS 5.0 using a TGT of 500.

**Analysis of microarray data.** For our Mainz data set, selection of "informative" genes was made using (*a*) quality control criteria as provided by the Affymetrix software, (*b*) the absolute median signal intensity, and (*c*) the coefficient of variation of a gene within our data set. Genes passing the quality control filter of having a "present" call in at least 10 samples, median signal intensity >75, and a coefficient of variation >60% within our data set were considered to be informative and used for subsequent analysis (Supplementary Table S2). For unsupervised analysis, we performed average linkage hierarchical clustering on all informative genes and samples using Pearson correlation as implemented in GeneSpring 7.0 software (Agilent Technologies). Principal component analysis (PCA) was done using GeneSpring 7.0. Clinical information was visualized as categorical or continuous variables, and relative gene expression was visualized on a relative scale from red, indicating high expression, to green, indicating low expression. Gene groups were defined after manual selection of nodes of the

---

[6] http://www.ncbi.nlm.nih.gov/geo/

gene dendrogram as suggested by the occurrence of cluster regions within the heat map. Gene Ontology (GO) annotations of genes contained within a cluster were tested by Fisher's exact test for overrepresentation of specific GO terms using the Expressionist software (GeneData AG; Supplementary Table S3). A metagene was calculated as representative of all genes contained within one gene cluster based on the median of the normalized expression values within the respective cluster. Per-gene normalization within the validation cohorts was done using median values obtained in the discovery cohort (Supplementary Table S4). High and low proliferation were defined using the proliferation metagene values. Because no assumptions about the shape of the metagene value distribution were made, an unsupervised clustering approach was chosen to determine a natural cutoff point between high and low proliferation. More specifically, a standard one-dimensional Lloyd quantization algorithm was used with the objective to build two distinct classes. The clustering was carried out using the Matlab software (The MathWorks), and a cutoff value of 1.037 was identified in the discovery cohort.

**Survival analysis.** Kaplan-Meier analysis was done using GraphPad Prism 7.0 software. Identification of a suitable cutoff for the B-cell metagene in the fast-proliferating discovery subcohort study was done using samples from patients who developed a distant metastasis within 5 y ($n = 23$) or remained disease-free for at least 5 y ($n = 57$) by receiver operating characteristic curve analysis. The resulting area under the receiver operating characteristic curve was 0.6987 [95% confidence interval (95% CI), 0.585–0.813; $P = 0.006$] with 95.7% sensitivity and 40.4% specificity, with 2.0 as cutoff. The value was used in the Rotterdam and TRANSBIG cohort studies without further adjustment. Metastasis-free survival (metastasis-free interval) was computed from the date of diagnosis to the date of diagnosis of distant metastasis. Survival curves were compared with the log-rank test. Univariate as well as multivariate $P$ values for the respective risk factors in the survival model were obtained by a Cox proportional hazards model as implemented in Matlab. All tests were done at a significance level of $\alpha = 0.05$. All $P$ values are two-sided. The Mainz finding cohort was used in a discovery-driven approach for hypothesis generation. Because no multiple test correction was done, we included $P$ values for this cohort for illustrative purposes only. Two validation cohorts (Rotterdam and TRANSBIG) were used for independent statistical testing of the hypothesis.

## Results

**Hierarchical cluster analysis and biological motives in breast cancer tissues.** Primary tumor tissues from 200 patients with invasive breast carcinoma were analyzed by gene expression profiling using HG-U133A oligonucleotide arrays. All patients were node negative and did not receive systemic chemotherapy or endocrine therapy after surgery (Table 1). To identify coregulated genes representing distinct biological processes or cell types, we performed a two-dimensional hierarchical cluster analysis using 2,579 genes selected for variable expression and quality within our data set (Supplementary Table S2). This discovery-driven analysis approach grouped the samples, as well as the genes, according to overall similarity in relative gene expression, thereby visualizing dominant clusters of coregulated genes (Fig. 1). Inspection of the annotation (gene name, locus, GO, etc.) of the probe sets contained in the individual clusters indicates either the underlying biological process represented by these genes or their cell type–specific origin. GO categories overrepresented in each cluster were identified by Fisher's exact test (Supplementary Tables S2 and S3). Similar clusters as well as marker genes typically contained in these clusters have been described by several other groups (3, 6). The clusters can be assigned as (*a*) basal-like, (*b*) T-cell, (*c*) B-cell, (*d*) IFN, (*e*) proliferation, (*f*) ER (luminal), (*g*) chromosome 17 (ERBB2), (*h*) stromal, (*i*) normal-like, (*j*) Jun-Fos cluster, or (*k*) transcription cluster. Because ER-

coregulated genes had a dominant effect on overall gene expression, the samples were grouped according to their ER status as displayed in the first row of the sample parameter bar below the heat map. Likewise, a correlation between tumor grade and expression of proliferation genes can be deduced from the heat map as well as the sample parameter bar (*second row*). However, further interrelations between gene expression and clinical or histopathologic features of the corresponding tumors did not become apparent with hierarchical clustering as compared with the visualization method. In particular, the correlation between gene expression and the occurrence of a metastasis (*third row*) remains elusive.

**Characterization of metagene expression by PCA.** To obtain a clearer view on the molecular heterogeneity of node-negative breast cancer, we applied PCA. The largest amount of variation within a data set that can be captured and displayed in three dimensions is a plot made by the first three principal components (PC1–PC3), which, for our data set, accounted for 28% of total variance. We wanted to investigate how the relative expression of prognosis-related genes contributed to the separation of samples within the PCA plot. Therefore, we constructed metagenes for the T-cell, B-cell, proliferation, and ER clusters by calculating, for each sample, the median of the normalized expression of all genes contained in each respective cluster. In this way, we defocused from individual genes to metagenes representing the median expression of all genes within a particular cluster.

In our population-based cohort study, samples were separated on the first principal component (PC1) predominantly according to expression of *ESR1* and *ESR1* coregulated genes. Accordingly, samples with the highest ER metagene expression clustered on the bottom left, and those with the lowest values on the bottom right (Fig. 2*A*). Variable expression was seen in the intermediate area, which broadly scattered on PC2. Visualization of the proliferation metagene revealed a gradient, with samples in the top left having the lowest and samples in the bottom right having the highest expression (Fig. 2*B*). A similar gradient was formed by individual well-known cell cycle–associated genes such as *MKI67, CCNE2*, and others (Supplementary Fig. S1*a* and *b*). As expected, a high correlation existed between proliferation genes and tumor grade (Supplementary Fig. S1*e*). Therefore, we conclude that high expression of proliferation genes is a surrogate for tumors with high proliferative activity. Interestingly, cancers of medullar histology clustered in a region of high proliferation and very low *ESR1* expression (Fig. 3*A*).

When time to distant metastasis was visualized, it became apparent that most patients with early metastasis were located in a region characterized by middle to high values of PC1 and low values of PC2 (Fig. 3*B*). These samples were characterized by intermediate to low ER metagene expression and concurrent high proliferation (i.e., proliferation metagene expression). Thus, two different tumor types are less prone to metastasize: one characterized by very high ER metagene expression and the other by intermediate ER metagene and simultaneous low expression of the proliferation metagene. However, in a region of samples with relatively high proliferation and low ER metagene levels, a sparsity of samples with distant metastasis was observed as well. Remarkably, this region was characterized by high T-cell and B-cell metagene expression, as shown in Fig. 2*C* and *D*, indicating that a lymphoid infiltration in these tumor tissues might be associated with a good prognosis. The T-cell metagene

contains information from genes like T-cell receptor *TRA@, TRB@*, as well as several other genes preferentially expressed in T cells (Supplementary Table S2). In contrast, the B-cell metagene is primarily formed by immunoglobulin heavy- and light-chain genes of several immunoglobulin classes such as *IGKC, IGHG*, and *IGHM* (Supplementary Table S2). Both metagenes form another gradient within the samples in the PCA plot with an axis from the bottom left to the top right. The complete absence of lymphoid infiltrates in the group of highest ER metagene expression results in a kind of sandwich situation in which a good prognosis coincides with either very high or virtually no lymphoid infiltration, whereas a particular group with intermediate lymphoid infiltration has a high risk of recurrence (Figs. 2*C* and *D* and 3*B*). Interestingly, *IGHM* is expressed in tumors of low and high expression of proliferation-associated genes, whereas *IGHG* is expressed predominantly in high-proliferating tumors (Supplementary Fig. S1*c* and *d*).

**Prognostic relevance of lymphoid infiltration in three patient cohorts.** When univariate Cox regression analysis was done for the ER, proliferation, T-cell, and B-cell metagenes, only proliferation ($P < 0.001$) and ER ($P = 0.028$) were significant. The B-cell metagene showed only a trend for significance ($P = 0.095$) and the T-cell metagene was not associated with prognosis at all (Table 2*A*). In multivariate Cox regression analysis including all metagenes (B-cell, T-cell, proliferation, and ER metagenes), only proliferation remained significant ($P < 0.001$). Apparently, the immune system does not play a relevant role throughout all breast cancer subtypes in this patient cohort. Because the PCA already suggested that a protective role of the immune system might be confined to fast-proliferating and usually highly aggressive tumors, we performed Lloyd quantization on the proliferation metagene values to split the discovery set ($n = 200$) into patients with low ($n = 105$) and high ($n = 95$) proliferation (two clusters). Lloyd's method was chosen because it makes no assumption about the underlying distribution and tries to find a natural separation point for the given number of clusters. Univariate Cox regression analysis identified an association of the B-cell ($P = 0.018$) and T-cell ($P = 0.041$) metagenes with
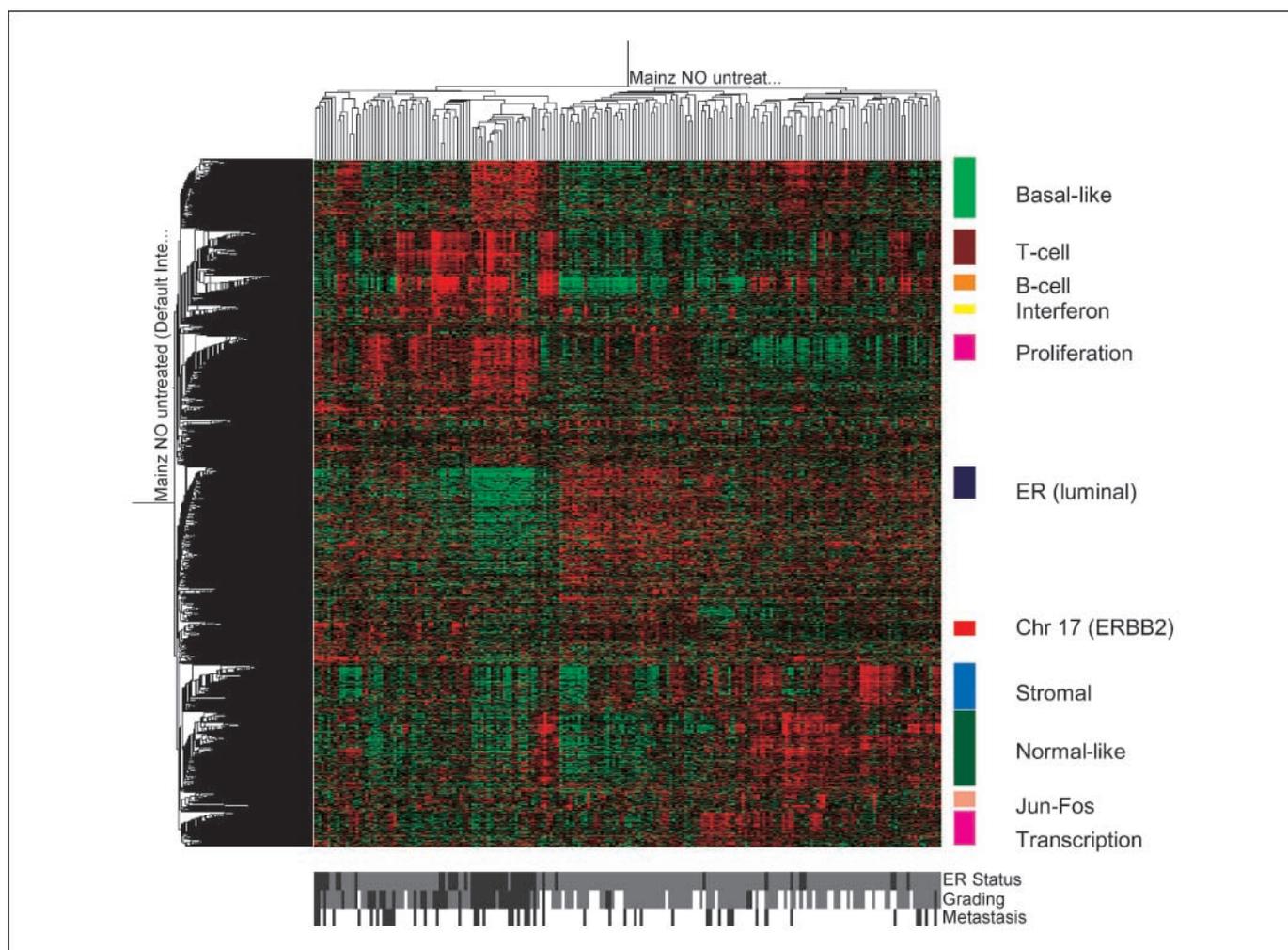


**Figure 1.** Hierarchical clustering of 2,579 probe sets (*rows*) and 200 breast cancer specimens (*columns*). The color represents variation in gene expression relative to the median level of expression across all samples. *Red,* high expression; *green,* low expression. The first row depicted below the heat map indicates the ER status as determined by immunohistochemistry (*black,* ER-negative; *gray,* ER-positive). The second row displays tumor grade (*white,* grade 1; *gray,* grade 2; *black,* grade 3). *Third row, black,* all patients who developed a distant metastasis regardless of time. Clusters of coregulated genes were selected by visual inspection using the heat map and gene dendrogram. Cluster names were based on prominent gene annotations (location, functions, etc.) contained within the respective cluster (Supplementary Tables S2 and S3).

**Figure 2.** Metagene expression visualized in the Mainz cohort study by PCA. Relative expression of the ER (*A*), proliferation (*B*), T-cell (*C*), and B-cell (*D*) metagenes in 200 samples distributed in the three-dimensional space of PC1 to PC3. *Red,* high expression; *yellow,* intermediate expression; *green,* low expression. *Arrows,* directions of the "metagene axis" from low values to high values.
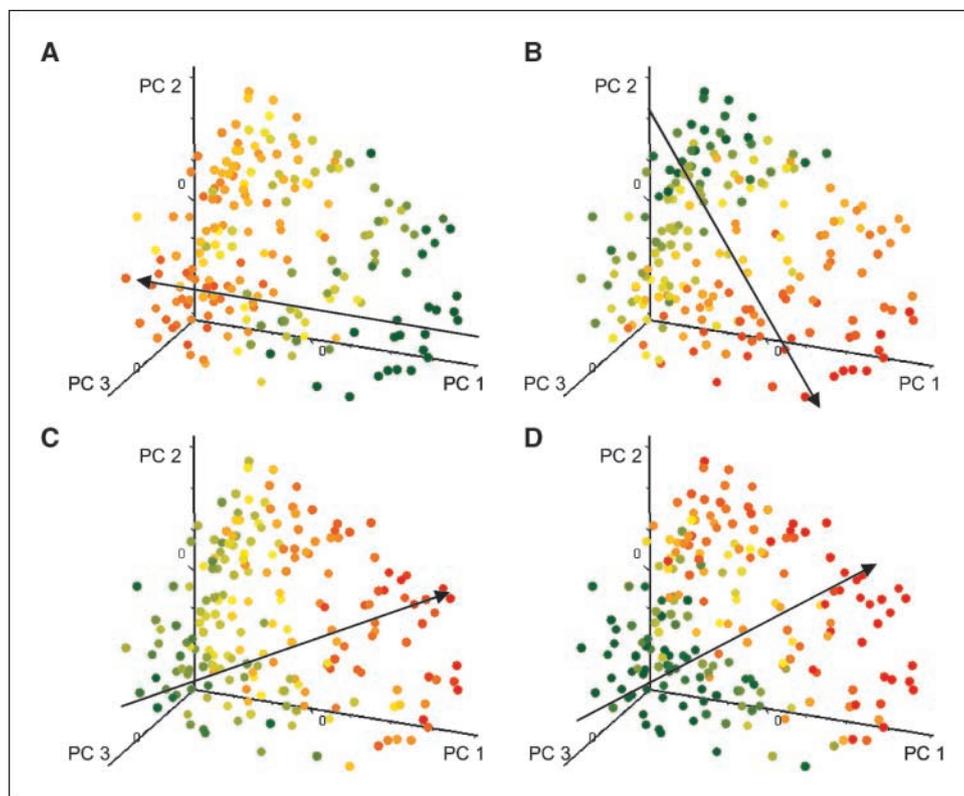
metastasis-free survival in the subgroup of patients having tumors with high expression of proliferation genes. Proliferation was still significantly associated with survival ($P$ = 0.001; Table 2*A*). Subsequently, we performed a multivariate Cox analysis, which showed that the B-cell metagene was independent from the other metagenes ($P$ = 0.034; Table 2*A*). Clinical variables such as ER status, progesterone receptor status, grading, age, and tumor size were not associated with survival in highly proliferating tumors (Supplementary Table S5). Multivariate regression analysis showed that the B-cell metagene was also independent from these clinical factors. Exclusion of medullary cancers from analysis still yielded a significant association for the B-cell metagene by univariate ($P$ = 0.038) and a borderline association by multivariate Cox regression (Supplementary Table S6). From these findings we hypothesized that a protective effect of the immune system might be of particular relevance to fast-proliferating tumors. To test a possible association of the B-cell metagene with survival in independent patient cohorts, we analyzed two publicly available expression data sets of 286 and 302 node-negative, untreated breast cancer patients, respectively, profiled by the same platform as our samples (7, 14, 15). The patient characteristics of these cohorts are summarized in Table 1. Obvious differences between the cohorts exist for patient age and tumor grade. Nevertheless, all key features observed in the PCA plot of our discovery cohort study were reproduced in the validation data sets (Supplementary Figs. S2 and S3). In particular, the orientation of all analyzed metagenes (B-cell, T-cell, ER, and proliferation) and the position of patients suffering early metastasis were similar. In univariate analysis of the Rotterdam validation set of tumors ($n$ = 286), the B-cell ($P$ = 0.009), proliferation ($P$ = 0.002), and ER ($P$ = 0.042) metagenes showed a

significant association with survival. Under multivariate analysis, the B-cell and ER metagenes remained significant (Table 2*B*). When the Rotterdam data set was stratified for proliferation, a larger proportion of tumors was above the cutoff defined in the Mainz cohort distinguishing high- and low-proliferating tumors ($n$ = 184 versus $n$ = 102). Thus, the enrichment of grade 2 and 3 tumors in the Rotterdam cohort is reflected by a shift toward higher proliferation metagene values. In the high proliferation subgroup, only the T-cell ($P$ = 0.006) and B-cell ($P$ = 0.003) metagenes (i.e., the immune axis) were associated with survival according to univariate analysis. In multivariate analysis of highly proliferating tumors, only the B-cell metagene was independently associated with survival ($P$ = 0.03), confirming the hypothesis obtained from the discovery set of tumors. In the second validation set, TRANSBIG, the B-cell and proliferation metagenes showed in the univariate analysis of the whole cohort a significant ($P$ = 0.019, respectively $P$ = 0.049) association with survival. Only the B-cell metagene correlated with survival by multivariate analysis in both the whole cohort and the subcohort of fast-proliferating tumors. The distribution of samples according to proliferation and B-cell metagene expression revealed a kind of sandwich situation in which good outcome coincided with low to intermediate B-cell metagene expression in slowly proliferating tumors and with high B-cell metagene expression in fast-proliferating tumors (Supplementary Fig. S4). The dominant effect of the B-cell metagene on survival time in tumors with high expression of proliferation genes was visualized by Kaplan-Meier analysis in all three cohorts using a cutoff value determined in the Mainz cohort. For highly proliferating tumors, a similar influence was seen in both the discovery and validation cohort studies (Fig. 4). This association was not solely restricted to *ESR1*-negative or *ERBB2*-positive tumors, which were

categorized to almost 90% into the fast-proliferating subgroup, but was also observed in the remaining *ESR1*-positive/*ERBB2*-negative fast-proliferating tumors (Supplementary Figs. S5 and S6).

## Discussion

In an unsupervised discovery approach, we combined two-dimensional hierarchical clustering with PCA (21, 22). We used PC1 to PC3, which retain the largest possible variation that can be displayed in three dimensions, and projected gene expression and clinical information into the distribution of tumor samples. Samples were separated on PC1 predominantly according to the expression of the ER metagene, reiterating the pivotal influence of ER for the molecular profile of breast cancer. The proliferation metagene formed another axis. Notably, almost all ER-negative breast cancer samples were characterized by high proliferation. Tumors with intermediate ER expression showed the highest variation in proliferative activity. High expression of proliferation-associated genes in this subtype seemed to be linked with an equally poor prognosis as for ER-negative tumors. When systematically using different metagenes for an explanation for the noticeably low number of early metastases in the region with concurrent low ER and high proliferation, we detected a third axis. This axis was almost perpendicular to the proliferation axis, indicating mutually independent information. It was formed by the B-cell metagene, containing B-cell–associated genes such as immunoglobulins, and, to a lesser extent, by the T-cell metagene, containing T-cell–related genes such as the T-cell receptor (*TCR*). These two metagenes are largely overlapping. In a region in which these metagenes were highly expressed, metastases occurred rarely despite high proliferation and low ER expression. Because gene expression profiling is a quantitative technique, it was possible to analyze B-cell and T-cell metagenes in relation to metastasis-free survival and to establish a hypothesis to obtain the highest predictive value in each breast cancer subtype. As suggested by the PCA results, we performed survival analysis separately for patients with low and high expression of proliferation-associated genes. The B-cell metagene was associated with metastasis-free survival in the subtype of highly proliferating tumors. Multivariate analysis suggested that the B-cell metagene is a prognostic factor for highly proliferating tumors, independent of the proliferation, ER, and T-cell metagenes. To validate our

hypothesis in separate gene expression data sets, we used two previously published cohort studies of untreated, node-negative breast cancer patients. The influence of the B-cell metagene was unequivocally confirmed in the Rotterdam and TRANSBIG cohorts by Cox regression analysis. However, in contrast to Mainz, the B-cell metagene was associated with outcome in the whole validation cohorts as well. This difference might be explained by cohort bias (i.e., differences between patients in the respective cohorts). The Rotterdam cohort is clearly enriched for high-grade tumors compared with Mainz, which is reflected by a shift in the distribution of the proliferation metagene within the Rotterdam cohort (Supplementary Fig. S7). When we used the Mainz-defined cutoff value to distinguish between high and low proliferation, it became evident that a larger proportion of Rotterdam samples is categorized as high-proliferating (184 high-proliferating versus 102 low-proliferating tumors). The discrepancy between our Mainz cohort and the TRANSBIG cohort might be explained by a striking difference in patient age. Whereas the TRANSBIG cohort hardly contains elderly patients, the Mainz cohort is depleted for ER-negative tumors (Supplementary Fig. S7), which are more often found in younger women. Interestingly, a dependence of the prognostic effect of lymphoid infiltration on young patient age has been described in a large series of 1,919 patients (23). Our observations indicate that cohort bias has to be taken into account in breast cancer analysis, even in homogeneously treated patients lacking two layers of complexity (i.e., mixed node status and systemic treatment).

A relationship between host defense mechanisms and prognosis of breast cancer has been discussed for decades (24). In a pioneering study, Aaltomaa and colleagues (19) showed that lymphocytic infiltrates were related to a good outcome in breast cancer, especially in rapidly proliferating tumors. Recently, a close association between marked diffuse inflammation and good outcome was described especially for grade 3 breast carcinomas (25). Furthermore, medullary breast cancer has been identified to be closely related to the basal-like tumor type (26), which suggests that the described poor outcome of the basal subtype (4) could be improved by the influence of the immune system. All these results are in good agreement with our observations made by gene expression profiling. A recent meta-analysis of 186 ER-negative breast tumors revealed five major subtypes, two of them characterized by prominent up-regulation of immune
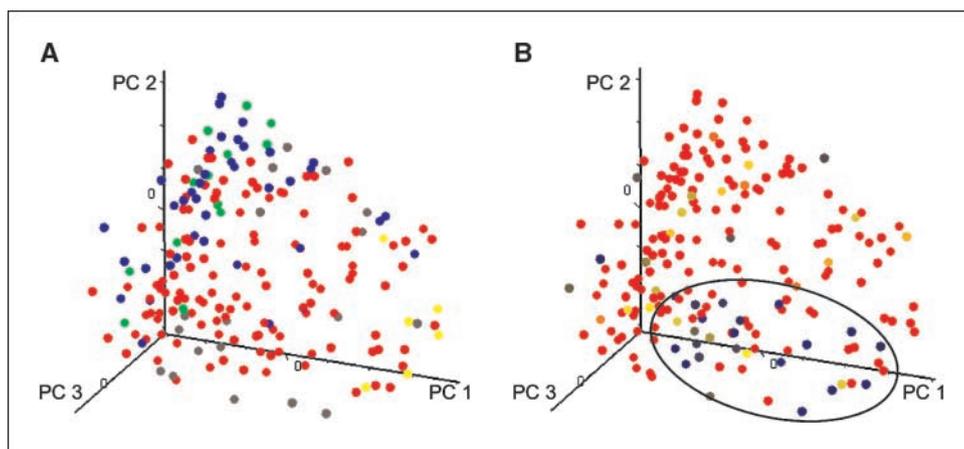


**Figure 3.** Histologic tumor types and time to metastasis visualized in the Mainz cohort study by PCA. *A*, histologic tumor types. *Blue*, lobular; *green*, tubular; *red*, ductal; *yellow*, medullary; *gray*, others. *B*, all samples of patients who developed a distant metastasis are colored on a continuous scale from early (*blue*) to late metastasis (*dark orange*). *Red*, samples from patients that remained disease-free until end of follow-up, regardless of time. *Region marked by a circle*, an obvious accumulation of early metastases.

**Table 2.** Cox regression analyses of metagene expression in relation to time to metastasis in the Mainz (*A*), Rotterdam (*B*), and TRANSBIG (*C*) cohorts
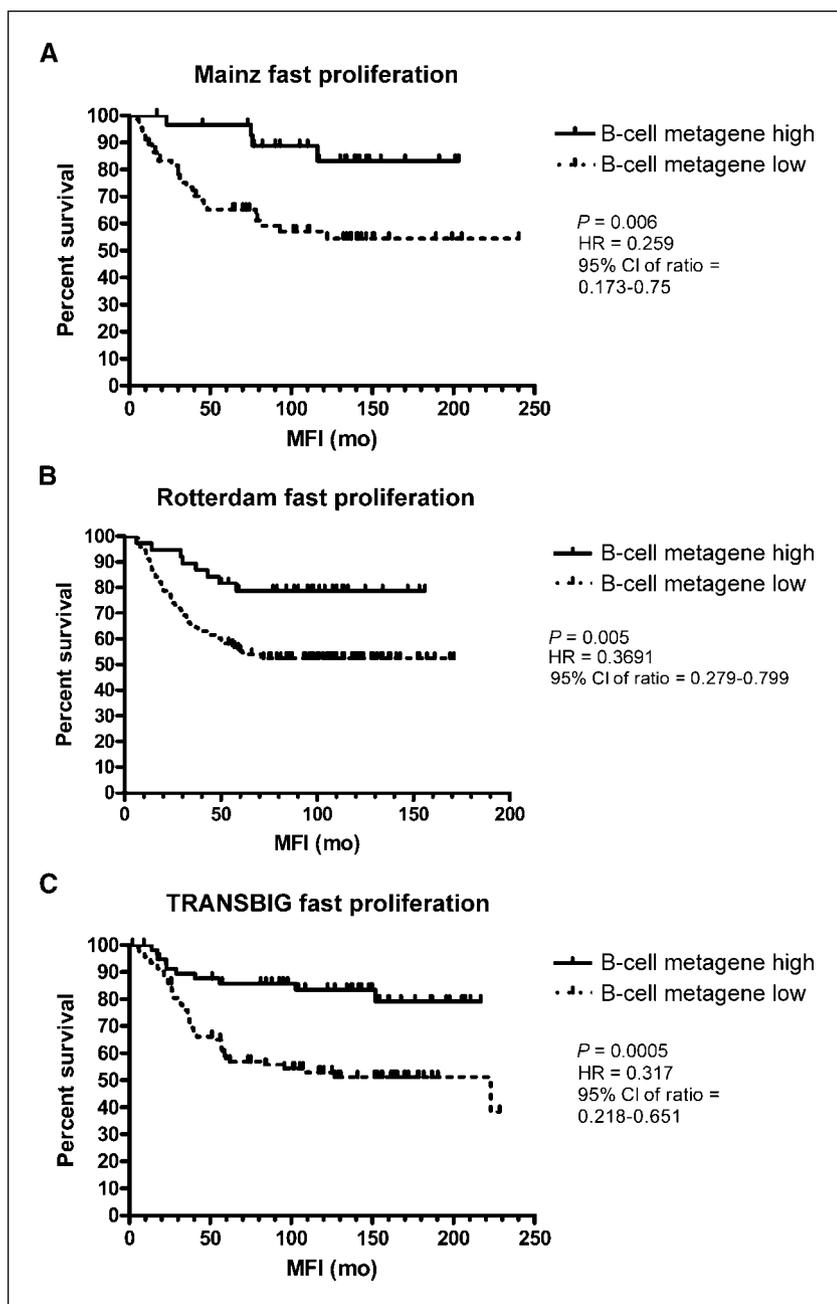
**(A) Mainz**

| Metagene | All data (*n* = 200; 47 events) | | High proliferation (*n* = 95; 31 events) | | Low proliferation (*n* = 105; 16 events) | |
|---|---|---|---|---|---|---|
| | HR (95% CI) | *P* | HR (95% CI) | *P* | HR (95% CI) | *P* |
| Univariate | | | | | | |
| T-Cell | 0.80 (0.53–1.19) | 0.263 | 0.61 (0.38–0.98) | 0.041 | 0.99 (0.38–2.6) | 0.986 |
| B-Cell | 0.84 (0.68–1.03) | 0.095 | 0.70 (0.53–0.94) | 0.018 | 1.16 (0.83–1.62) | 0.395 |
| Proliferation | 2.30 (1.64–3.24) | <0.001 | 2.2 (1.36–3.57) | 0.001 | 3.6 (0.3–43.46) | 0.313 |
| ER | 0.47 (0.24–0.92) | 0.028 | 0.8 (0.33–1.89) | 0.605 | 0.62 (0.12–3.20) | 0.571 |
| Multivariate | | | | | | |
| T-Cell | 0.83 (0.50–1.39) | 0.48 | 0.93 (0.53–1.61) | 0.783 | 0.47 (0.12–1.80) | 0.269 |
| B-Cell | 0.79 (0.61–1.04) | 0.093 | 0.66 (0.46–0.97) | 0.034 | 1.41 (0.90–2.20) | 0.136 |
| Proliferation | 2.20 (1.40–3.46) | <0.001 | 2.02 (1.14–3.59) | 0.016 | 7.06 (0.49–101) | 0.151 |
| ER | 0.63 (0.26–1.56) | 0.318 | 0.61 (0.21–1.79) | 0.365 | 0.54 (0.08–3.58) | 0.519 |

**(B) Rotterdam**

| Metagene | All data (*n* = 286; 106 events) | | High proliferation (*n* = 184; 77 events) | | Low proliferation (*n* = 102; 29 events) | |
|---|---|---|---|---|---|---|
| | HR (95% CI) | *P* | HR (95% CI) | *P* | HR (95% CI) | *P* |
| Univariate | | | | | | |
| T-Cell | 0.77 (0.60–0.98) | 0.036 | 0.66 (0.48–0.89) | 0.006 | 1.00 (0.65–1.53) | 0.999 |
| B-Cell | 0.78 (0.65–0.94) | 0.009 | 0.71 (0.57–0.89) | 0.003 | 0.95 (0.73–1.25) | 0.701 |
| Proliferation | 1.53 (1.17–2.01) | 0.002 | 1.30 (0.79–1.92) | 0.183 | 21.7 (0.29–61.7) | 0.016 |
| ER | 0.61 (0.38–0.98) | 0.042 | 0.80 (0.45–1.44) | 0.46 | 0.63 (0.27–3.35) | 0.403 |
| Multivariate | | | | | | |
| T-Cell | 0.80 (0.5–1.15) | 0.22 | 0.74 (0.48–1.13) | 0.159 | 0.88 (0.45–1.73) | 0.719 |
| B-Cell | 0.78 (0.62–0.98) | 0.034 | 0.74 (0.56–0.97) | 0.03 | 0.91 (0.64–1.29) | 0.6 |
| Proliferation | 1.27 (0.88–1.82) | 0.197 | 1.01 (0.62–1.63) | 0.976 | 23.2 (1.77–305) | 0.017 |
| ER | 0.50 (0.25–1.00) | 0.048 | 0.46 (0.21–1.01) | 0.053 | 0.58 (0.15–2.24) | 0.428 |

**(C) TRANSBIG**

| Metagene | All data (*n* = 302; 77 events) | | High proliferation (*n* = 151; 54 events) | | Low proliferation (*n* = 151; 23 events) | |
|---|---|---|---|---|---|---|
| | HR (95% CI) | *P* | HR (95% CI) | *P* | HR (95% CI) | *P* |
| Univariate | | | | | | |
| T-Cell | 0.79 (0.59–1.05) | 0.101 | 0.62 (0.44–0.86) | 0.004 | 0.99 (0.55–1.77) | 0.96 |
| B-Cell | 0.85 (0.75–0.97) | 0.019 | 0.78 (0.66–0.91) | 0.001 | 0.98 (0.80–1.20) | 0.86 |
| Proliferation | 1.35 (1.00–1.82) | 0.049 | 0.53 (0.30–0.95) | 0.034 | 4.47 (0.47–42.6) | 0.192 |
| ER | 0.64 (0.38–1.07) | 0.089 | 2.02 (1.06–3.84) | 0.032 | 0.15 (0.03–0.62) | 0.009 |
| Multivariate | | | | | | |
| T-Cell | 0.86 (0.6–1.23) | 0.418 | 0.83 (0.55–1.26) | 0.384 | 0.63 (0.25–1.54) | 0.306 |
| B-Cell | 0.83 (0.7–0.97) | 0.021 | 0.82 (0.68–0.99) | 0.035 | 1.01 (0.78–1.30) | 0.942 |
| Proliferation | 1.17 (0.77–1.79) | 0.463 | 0.68 (0.30–1.18) | 0.135 | 3.68 (0.42–32.1) | 0.24 |
| ER | 0.49 (0.22–1.09) | 0.08 | 0.69 (0.30–1.82) | 0.510 | 0.10 (0.02–0.48) | 0.004 |

response genes (16). A subtype called "SR+," characterized by overexpression of steroid response genes and correlation with the HER2+ intrinsic subtype, showed a poorer outcome than a cluster of tumors called "IR+," which is characterized by overexpression of immune response–related genes. In addition, Alexe and colleagues (18) showed that high expression of lymphocyte-associated genes in node-negative HER2+ breast cancer correlates with good outcome. The discrepancy with

HER2+ tumors might be related to the fact that the first study analyzed tumors of mixed node status and heterogeneously treated patients, whereas the second study included tumors from node-negative, untreated patients only. Anyhow, we confirm the prognostic relevance of immune cell infiltration in both *ESR1*-negative and *ERBB2*-positive node-negative breast cancers (Supplementary Figs. S5 and S6). Furthermore, we show that almost 90% of both subtypes belong to the "high proliferation"

type as defined in our discovery cohort. Finally, we show that the prognostic influence of B-cell transcripts is also seen in "high proliferation" type tumors classified as *ESR1* positive and *ERBB2* negative (Supplementary Figs. S5 and S6). Functional interrelations between immune and tumor cells and their consequence for patient outcome are still unclear. Because *IGHM* transcripts are found in low-proliferating and high-proliferating tumors whereas *IGHG* transcripts are primarily found in fast-proliferating tumors, it is tempting to speculate that the humoral immune response matures with tumor progression. Several reports have focused on the oligoclonal expansion of B cells in both medullary and ductal breast carcinomas (27–29). Hansen and colleagues (30) described an oligoclonal B-cell response targeting actin that was exposed on the tumor cell surface as an early

apoptotic event in medullary breast cancer. The observed IgG antibody response showed all criteria of an antigen-driven, high-affinity response. In addition, ganglioside D3 has been identified as another target for an oligoclonal B-cell response in medullary breast cancer (31). These authors interpreted their findings as proof of principle about the active role of tumor-infiltrating B lymphocytes. Despite tempting implications about the prognostic effect of these findings, none of these studies actually analyzed the significance of the described B-cell response for patient survival. In fact, conflicting results have led to a dispute about the actual role of tumor-associated leucocytes (32). Based on animal experiments, the notion is being promoted that B cells inhibit the immune response against tumors (33, 34), and clinical trials using rituximab for B-cell depletion have been proposed.



**Figure 4.** Survival analyses according to B-cell metagene expression. Metastasis-free survival of fast-proliferating tumors stratified according to high or low expression of the B-cell metagenes of the Mainz cohort study [*A*; hazard ratio (HR), 0.238 (95% CI, 0.171–0.786); *P* = 0.01], Rotterdam cohort study [*B*; HR, 0.416 (95% CI, 0.284–0.851); *P* = 0.011], and TRANSBIG cohort study [*C*; HR, 0.098 (95% CI, 0.088–0.653); *P* = 0.005]. *MFI,* metastasis-free interval.

However, our findings in human breast tumors show a correlation between B-cell infiltration and good prognosis, suggesting that further research should be done before cancer patients are depleted of B cells.

In conclusion, we have shown a strong association of the expression of B-cell related mRNA transcripts with metastasis-free survival of rapidly proliferating node-negative breast cancer. Extending the knowledge of the complex role of immune cells and their interaction in breast cancers should ultimately pave the way for the long-awaited successful development of therapeutics aiming at the third prognosis axis.

## Disclosure of Potential Conflicts of Interest

## References

1. Osborne CK, Yochmowitz MG, Knight WA III, McGuire WL. The value of estrogen and progesterone receptors in the treatment of breast cancer. Cancer 1980;46:2884–8.
2. Gentili C, Sanfilippo O, Silvestrini R. Cell proliferation and its relationship to clinical features and relapse in breast cancers. Cancer 1981;48:974–9.
3. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumors. Nature 2000;406: 747–52.
4. Sorlie T, Perou CM, Tibshirani R, 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.
5. Rouzier R, Perou CM, Symmans WF, et al. Breast cancer molecular subtypes respond differently to preoperative chemotherapy. Clin Cancer Res 2005;11: 5678–85.
6. Van't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002;415:530–6.
7. Wang Y, Klijn JG, Zhang Y, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet 2005;365:671–9.
8. Van de Vijver MJ, He YD, van't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 2002;347:1999–2009.
9. Foekens JA, Atkins D, Zhang Y, et al. Multicenter validation of a gene expression-based prognostic signature in lymph node-negative primary breast cancer. J Clin Oncol 2006;24:1665–71.
10. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 2004;351:2817–26.
11. Sotiriou C, Wirapati P, Harris A, 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.
12. Oh DS, Troester MA, Usary J, et al. Estrogen-regulated genes predict survival in hormone receptor-positive breast cancers. J Clin Oncol 2006;24:1656–64.

13. Fan C, Oh DS, Wessels L, et al. Concordance among gene-expression-based predictors for breast cancer. N Engl J Med 2006;355:560–9.
14. Desmedt C, Piette F, Loi S, et al. Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. Clin Cancer Res 2007;13:3207–14.
15. Loi S, Haibe-Kains B, Desmedt 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.
16. Teschendorff AE, Naderi A, Barbosa-Morais NL, et al. A consensus prognostic gene expression classifier for ER positive breast cancer. Genome Biol 2006;7:R101.
17. Teschendorff A, Miremadi A, Pinder SE, Ellis IO, Caldas C. An immune response gene expression module identifies a good prognosis subtype in estrogen receptor negative breast cancer. Genome Biol 2007;8:R157.
18. Alexe G, Dalgin GS, Scanfeld D, et al. High expression of lymphocyte-associated genes in node-negative HER2[+] breast cancer correlates with lower recurrence rates. Cancer Res 2007;67:10669–76.
19. Aaltomaa S, Lipponen P, Eskelinen M, et al. Lymphocyte infiltrates as a prognostic variable in female breast cancer. Eur J Cancer 1992;28:859–64.
20. Ridolfi R, Rosen P, Port A, Kinne D, Mike V. Medullary carcinoma of the breast. A clinicopathologic study with 10-year follow up. Cancer 1977;40:1365–85.
21. Alter O, Brown PO, Botstein D. Singular value decomposition for genome-wide expression data processing and modeling. Proc Natl Acad Sci U S A 2000;97: 10101–6.
22. Selaru FM, Yin J, Olaru A, et al. An unsupervised approach to identify molecular phenotypic components influencing breast cancer features. Cancer Res 2001;64: 1584–8.
23. Menard S, Tomasic G, Casalini P, et al. Lymphoid infiltration as a prognostic variable for early-onset breast carcinomas. Clin Cancer Res 1997;3:817–9.
24. Di Paola M, Angelici L, Bertolotti A, Colizza S. Host

resistance in relation to survival in breast cancer. Br Med J 1974;4:268–70.
25. Lee AH, Gillett CE, Ryder K, Fentiman IS, Miles DW, Millis RR. Different patterns of inflammation and prognosis in invasive carcinoma of the breast. Histopathology 2006;48:692–701.
26. Bertucci F, Finetti P, Cervera N, et al. Gene expression profiling shows medullary breast cancer is a subgroup of basal breast cancers. Cancer Res 2006;66:4636–44.
27. Hansen MH, Nielsen H, Ditzel HJ. The tumor-infiltrating B cell response in medullary breast cancer is oligoclonal and directed against the autoantigen actin exposed on the surface of apoptotic cancer cells. Proc Natl Acad Sci U S A 2001;98:12659–64.
28. Coronella JA, Telleman P, Kingsbury GA, et al. Antigen-driven oligiclonal expansion of tumor-infiltrating B cells in infiltrating ductal carcinoma of the breast. J Immunol 2002;169:1829–36.
29. Nzula S, Going JJ, Stott DI. Antigen-driven clonal proliferation, somatic hypermutation, and selection of B lymphocytes infiltrating human ductal breast carcinomas. Cancer Res 2003;63:3275–80.
30. Hansen MH, Nielsen HV, Ditzel HJ. Translocation of an intracellular antigen to the surface of medullary breast cancer cells early in apoptosis allows for an antigen-driven antibody response elicited by tumor-infiltrating B cells. J Immunol 2002;169:2701–11.
31. Kotlan B, Simsa P, Teillaud JL, et al. Novel ganglioside antigen identified by B cells in human medullary breast carcinomas: the proof of principal concerning the tumor-infiltrating B lymphocytes. J Immunol 2005;175: 2278–85.
32. O'Sullivan C, Lewis CE. Tumor-associated leucocytes: friends or foes in breast carcinoma. J Pathol 1994;172: 229–35.
33. Inoue S, Leitner WW, Golding B, Scott D. Inhibitory effects of B cells on antitumor immunity. Cancer Res 2006;66:7741–7.
34. Shah S, Divekar AA, Hilchey SP, et al. Increased rejection of primary tumors in mice lacking B cells: inhibition of anti-tumor CTL and Th1 cytokine responses by B cells. Int J Cancer 2005;117:574–86.

# The Humoral Immune System Has a Key Prognostic Impact in Node-Negative Breast Cancer

Marcus Schmidt, Daniel Böhm, Christian von Törne, et al.

| | |
|---|---|
| **Updated version** | Access the most recent version of this article at: http://cancerres.aacrjournals.org/content/68/13/5405 |
| **Supplementary Material** | Access the most recent supplemental material at: http://cancerres.aacrjournals.org/content/suppl/2008/06/19/68.13.5405.DC1 |

| | |
|---|---|
| **Cited articles** | This article cites 34 articles, 17 of which you can access for free at: http://cancerres.aacrjournals.org/content/68/13/5405.full#ref-list-1 |
| **Citing articles** | This article has been cited by 68 HighWire-hosted articles. Access the articles at: http://cancerres.aacrjournals.org/content/68/13/5405.full#related-urls |

| | |
|---|---|
| **E-mail alerts** | Sign up to receive free email-alerts related to this article or journal. |
| **Reprints and Subscriptions** | To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org. |
| **Permissions** | To request permission to re-use all or part of this article, use this link http://cancerres.aacrjournals.org/content/68/13/5405. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site. |