



## **Ep-CAM RNA expression predicts metastasis-free survival in three cohorts of untreated node-negative breast cancer**

Marcus Schmidt, Ilka Brigitte Petry, Daniel Böhm, Antje Lebrecht, Christian Törne, Susanne Gebhard, Aslihan Gerhold-Ay, Cristina Cotarelo, Marco Battista, Wiebke Schormann, et al.

### **► To cite this version:**

Marcus Schmidt, Ilka Brigitte Petry, Daniel Böhm, Antje Lebrecht, Christian Törne, et al.. Ep-CAM RNA expression predicts metastasis-free survival in three cohorts of untreated node-negative breast cancer. *Breast Cancer Research and Treatment*, 2010, 125 (3), pp.637-646. 10.1007/s10549-010-0856-5 . hal-00561313

**HAL Id: hal-00561313**

**<https://hal.science/hal-00561313>**

Submitted on 1 Feb 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## **Ep-CAM RNA expression predicts metastasis free survival in three cohorts of untreated node-negative breast cancer**

<sup>1</sup>Marcus Schmidt, <sup>1</sup>Ilka Brigitte Petry, <sup>1</sup>Daniel Böhm, <sup>1</sup>Antje Lebrecht, <sup>2</sup>Christian von Törne, <sup>1</sup>Susanne Gebhard, <sup>3</sup>Aslihan Gerhold-Ay, <sup>4</sup>Cristina Cotarelo, <sup>1</sup>Marco Battista, <sup>5</sup>Wiebke Schormann, <sup>5</sup>Evgenia Freis, <sup>6</sup>Silvia Selinski, <sup>6</sup>Katja Ickstadt, <sup>6</sup>Jörg Rahnenführer, <sup>7</sup>Martin Sebastian, <sup>8</sup>Martin Schuler, <sup>1</sup>Heinz Koelbl, <sup>2</sup>Mathias Gehrmann, <sup>5</sup>Jan G Hengstler

<sup>1</sup>Department of Obstetrics and Gynecology, University Medical Center of the Johannes Gutenberg-University, Mainz, Germany, <sup>2</sup>Siemens Medical Solutions, Diagnostics GmbH, Cologne, Germany, <sup>3</sup>Department of Medical Biometrics, Epidemiology and Informatics, Johannes Gutenberg University, Mainz, Germany  
<sup>4</sup>Department of Pathology, Johannes Gutenberg University, Mainz, Germany, <sup>5</sup>IfADo-Leibniz Research Centre for Working Environment and Human Factors (IfADo), TU Dortmund University, Dortmund, Germany, <sup>6</sup>Collaborative Research Center 475, Faculty of Statistics, TU Dortmund University, Dortmund, Germany, <sup>7</sup>Department of Medicine III, Johannes Gutenberg University, Mainz, Germany, <sup>8</sup>Department of Medicine (Cancer Research), West German Cancer Center, University Hospital Essen, Germany

Corresponding authors: Marcus Schmidt, M.D.

Mailing address: Department of Obstetrics and Gynecology, University Medical Centre of the Johannes Gutenberg-University, Medical School, Langenbeckstr. 1, 55131 Mainz, Germany

Telephone: 0049-6131-172683

Facsimile: 0049-6131-175673

e-mail: [marcus.schmidt@frauen.klinik.uni-mainz.de](mailto:marcus.schmidt@frauen.klinik.uni-mainz.de)

## Introduction

The Epithelial cell adhesion molecule (Ep-CAM) is a transmembrane glycoprotein expressed on most human epithelial cells, functioning as a homophilic cell-cell adhesion molecule [1]. Ep-CAM is a potential target for antibody based therapies [2-5]. Recently, we could show that a fully human monoclonal antibody against Ep-CAM has anti-tumor activity in heavily pretreated Ep-CAM positive metastatic breast cancer [6]. Overexpression of Ep-CAM has been reported to correlate with poor disease-free (DFS) and overall survival (OS) in breast cancer [7]. However, when the prognostic impact of Ep-CAM was later examined in a larger database of 1705 patients the authors postulated that the adverse prognostic effect of Ep-CAM overexpression was confined to node-positive patients [8]. Since node-positive breast cancer patients usually receive some form of adjuvant systemic therapy, this study does not explain whether the impact of Ep-CAM overexpression is purely prognostic in nature or confounded by predictive effects. The differentiation between prognostic and predictive factors is relevant, since therapeutic targets with prognostic relevance appear more promising [9]. The only possibility to rule out a potentially confounding predictive effect is to examine cohorts of patients without adjuvant systemic treatment. Recently, we have analyzed a cohort of 402 untreated node-negative breast cancer patients for immunohistochemically determined Ep-CAM expression [10]. Indeed, Ep-CAM overexpression carried prognostic significance and was associated with DFS independent from established clinical parameters [10]. Since monoclonal antibodies against Ep-CAM have shown clinical activity in metastatic breast cancer [6] the matter of definition of the subgroup of patients who profit most from this therapy becomes of increasing relevance. It is plausible that a subgroup of patients in whom Ep-CAM overexpression is associated with worse prognosis is most appropriate for an Ep-CAM antagonizing therapy. Therefore, the aim of the present

study was to analyze whether also Ep-CAM RNA expression can be used to identify a subgroup of patients with increased risk. Besides comparing the prognostic relevance of Ep-CAM RNA and protein, we additionally validated the prognostic relevance of Ep-CAM RNA expression in two additional cohorts of node-negative breast cancer [11-13]. Finally, we studied whether the prognostic influence of Ep-CAM RNA expression is independent from the recently described proliferation metagene [14].

## **Patients and Methods**

The Mainz node-negative cohort consisted of 409 consecutive node-negative breast cancer patients with tumor size pT1<sub>a</sub> – pT3 and adequate follow-up information who were treated at the Department of Obstetrics and Gynecology of the Johannes Gutenberg University Mainz between the years 1986 and 2000. Of these 409 patients paraffin blocks with tumor tissue were available for 402 individuals. In 194 patients fresh-frozen tissue for microarray-based gene-expression analysis was available [15]. All patients have been treated by surgical tumor resection, but had not received any systemic therapy in the adjuvant setting. Ep-CAM status was collected from a previously published database [10] IHC against Ep-CAM was available in 194 patients. Ep-CAM was assessed using an immunoreactive score defined by the product of a proportion score (0, none; 1, <10%; 2, 10-50%; 3, 50-80%; 4, >80% positive cells) and an intensity score (0, no staining; 1, weak; 2, moderate; 3, strong). Besides using the score, Ep-CAM expression was additionally assessed by the Ep-CAM status. Only tumors with strong specific surface membrane staining in more than 10% of the tumor cells were considered as Ep-CAM 3+ overexpressors and were thus assigned a positive Ep-CAM status. Patients had been treated either with modified radical mastectomy (44%) or breast conserving surgery followed by

irradiation (56%). Patients' characteristics are given in Table 1. The study was approved by the ethical review board of the medical association of Rhineland-Palatinate. We documented disease-free survival (DFS), metastasis-free survival (MFS) and breast cancer-specific overall survival (OS). The mean follow-up time was 10 years  $\pm$  5 years (standard deviation). 58 patients (30%) relapsed, 45 (23%) had distant metastases and 31 (16%) died of breast cancer (Table 1). The manuscript was prepared in agreement with the reporting recommendations for tumor marker reporting studies [15].

### **Gene expression profiling**

The Affymetrix (Santa Clara, CA) HG-U133A array and GeneChip System™ was used to quantify the relative transcript abundance in the breast cancer tissues as previously described [14]. To analyze Ep-CAM expression 201839\_s\_at was used. Raw .cel file, MAS 5.0 .txt file and patient data have already been deposited with accession number GSE11121 under <http://www.ncbi.nlm.nih.gov/geo>.

### **Previously published microarray datasets**

Two breast cancer Affymetrix HG-U133A microarray datasets including patient outcome information were downloaded from the NCBI GEO data repository (<http://www.ncbi.nlm.nih.gov/geo/>). 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 (n=286), [11]). The .txt file data were recalculated to a TGT of 500. The second data set (from GSE6532 and GSE7390) consists of 302 samples from breast cancer patients that remained untreated in the adjuvant setting after surgery (Transbig cohort, [12, 13]). Raw .cel

file data were processed by MAS 5.0 using a TGT of 500.

### **Analysis of microarray data**

For our Mainz dataset, 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 dataset as previously described (14). Genes passing the quality control filter of having a “present” call in at least 10 samples, median signal intensity above 75 and a coefficient of variation above 60% within our dataset were considered to be informative and used for subsequent analysis. 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, USA). Gene groups were defined after manual selection of nodes of the gene dendrogram as suggested by the occurrence of cluster regions within the heatmap. 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, Basel, Switzerland). 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 as previously described [14]. For further analysis, we used the previously described proliferation metagene [14].

### **Statistical Analysis**

Survival rates were calculated according to the Kaplan-Meier method. Metastasis-free survival (MFS) was computed from the date of diagnosis to the date of distant

metastasis or death from cancer if there was no earlier recurrence for the patients with gene-expression profiling available to allow for comparison with the pooled validation dataset. Survival time was compared with the Log-rank test. Multivariate Cox survival analyses were performed with inclusion [16]. Dichotomization was performed as described by Schmidt et al. [10]. Differences in Ep-CAM RNA expression between patients with different immunohistochemically determined Ep-CAM status or Ep-CAM scores, histological grade and pT stage were analyzed using the Mann-Whitney and the Kruskal-Wallis-test. Correlation between the proliferation metagene and Ep-CAM RNA expression was analyzed by the Spearman test. We used the gene list from Smid et al. [17] to cluster our patients (Mainz cohort) into intrinsic subtypes. Differences of Ep-CAM RNA levels between the individual intrinsic subtypes were analyzed by the Mann-Whitney-test. All p-values are two-sided. As no correction for multiple testing was performed they are descriptive measures. All analyses were performed using SPSS17.0.

## **Results**

### **Correlation of Ep-CAM RNA and protein expression**

The aim of this study was to compare the prognostic value of Ep-CAM RNA and immunohistochemically determined Ep-CAM protein expression in node negative breast cancer. Therefore, we first analyzed whether both parameters correlate with each other. Ep-CAM RNA and protein expression were determined in tumor tissue of 194 patients with node negative breast cancer (Mainz cohort). Ep-CAM RNA expression showed a unimodal distribution with a mean value and standard deviation of 11.6 and 0.84 (relative units, Fig. 1A). Immunohistochemically determined Ep-CAM was evaluated using the Ep-CAM score and the Ep-CAM status (Fig. 1B). Ep-CAM status also showed a statistically significant association with Ep-CAM RNA

expression (Fig. 1C,  $P=0.001$ , Mann-Whitney test). Also the Ep-CAM score was significantly associated with levels of Ep-CAM RNA (Fig. 1D,  $P=0.001$ , Kruskal-Wallis-test). However, the association between Ep-CAM RNA and protein was not linear. Ep-CAM scores of 0 showed the lowest and Ep-CAM scores of 12 the highest RNA expression levels, whereas Ep-CAM scores of 1-9 were intermediate (Fig. 1D). In addition, the difference in RNA expression levels between immunohistochemically positive and negative tumors was relatively small (Fig. 1C, D). Therefore, we could not simply replace immunostaining by the technically easier and better quantifiable RNA analysis but had to validate whether Ep-CAM RNA is adequate as a prognostic factor.

### **Ep-CAM RNA expression is an independent prognostic factor**

To study if Ep-CAM RNA expression is associated with prognosis we analyzed the Mainz cohort of 194 consecutive node negative breast cancer patients for whom gene array data as well as results from Ep-CAM immunostaining were available (patients characteristics: Table 1). Ep-CAM RNA expression was significantly associated with disease free survival in the univariate Cox analysis (Table 2A,  $P=0.001$ , HR=2.4). Also in the multivariate regression analysis adjusted for age, pTstage, grading, estrogen as well as progesterone receptor and Her-2 status, Ep-CAM RNA expression appeared to be prognostic, independent from the established clinical factors (Table 2B,  $P=0.017$ , HR=2.0). A similar result was obtained for analysis of metastasis free survival, where Ep-CAM RNA was significant in the univariate ( $P=0.003$ , HR=2.5) and in the multivariate ( $P=0.049$ , HR=1.9) Cox analysis (Table 2). Also for analysis of overall survival Ep-CAM RNA expression was significant in the univariate ( $P=0.002$ , HR=3.1) and in the multivariate ( $P=0.042$ , HR=2.3) Cox analysis (Supplemental Table 1). The influence of Ep-CAM RNA



expression on disease free (Fig. 2A), metastasis free (Fig. 2B), and overall (Fig. 2C) survival was visualized by Kaplan Meier analysis. Similar results were obtained if Ep-CAM RNA expression was dichotomized at the median (Fig. 2) or at the 75% percentile (Supplemental Fig. 1). Compared to Ep-CAM RNA the association of the immunohistochemically determined Ep-CAM status with prognosis was weaker. Trends towards shorter survival times were observed by Kaplan Meier analysis for disease free, metastasis free and overall survival time in Ep-CAM status positive patients (Supplemental Fig. 2). However, this amounted to statistical significance only for overall survival time ( $P=0.031$ , Supplemental Fig. 2C). In contrast to Ep-CAM RNA levels the immunological Ep-CAM status was not significantly associated with disease free survival (Supplemental Table 2), metastasis free interval (Supplemental Table 3) and overall survival (Supplemental Table 4) in the multivariate Cox regression analysis. In conclusion, high Ep-CAM RNA levels are associated with poor prognosis in the Mainz cohort and seem to perform even better compared to the immunohistochemically determined Ep-CAM status.

### **Validation in two independent cohorts**

In a next step we studied whether the prognostic relevance of Ep-CAM RNA observed in the Mainz cohort can be validated in two independent cohorts. For this purpose the availability of the established clinical parameters (age, pTstage, grading, hormone and Her-2 receptor status) is important to allow multivariate Cox analysis similar as shown for the Mainz cohort in Table 2. For the published Transbig cohort data on age, pTstage and grading are available. However, the Her-2 status determined by immunostaining and FISH as well as the estrogen and progesterone receptor status assessed by an immunoreactive score (as available for the patients from Mainz) are not available for the Transbig cohort. In order to achieve a

multivariate analysis we reverted to the published gene array data. Estrogen receptor, Her-2 and progesterone receptor RNA expression showed a bimodal distribution in tumor tissue (Supplemental Fig. 3). The cutpoints differentiating between high and low expression were identified in the combined cohort of patients from Mainz, Rotterdam and the Transbig group (Supplemental Fig. 3). This seems to be adequate, because similar bimodal distributions and cutpoints were observed in each individual subcohort (Transbig, Mainz and Rotterdam subcohorts, Supplemental Fig. 4A-C). Therefore, the cutpoints of 10.0 (estrogen receptor), 12.6 (Her-2) and 4.9 (progesterone receptor) could be used for all cohorts. In the Mainz cohort results of positive HER2, ER and PR status derived from the RNA expression data and from the immunostaining results showed an excellent correlation in the Chi-square test with  $P < 0.001$ , respectively (Supplemental Table 5). Using these cutpoints, patients in the Transbig cohort were dichotomized into Her-2 as well as hormone receptor high *versus* low (patients characteristics: Supplemental Table 6). Both, univariate ( $P = 0.006$ ) and multivariate ( $P = 0.035$ ) analysis of the Transbig cohort demonstrated the prognostic relevance of Ep-CAM RNA for metastasis free survival (Table 3). Since this type of analysis differs from that shown for the Mainz cohort in Table 2 concerning the classification of Her-2 and hormone receptors, we applied the technique of gene array based classification also for the Mainz cohort to obtain comparable conditions to the Transbig cohort (Supplemental Table 8). Dichotomizing Her-2 and the hormone receptors by gene array data in the Mainz cohort resulted in a similarly significant association of Ep-CAM RNA with metastasis free survival (Supplemental Table 8) as for the conventional Her-2 and hormone receptor status based on immunostaining and FISH (Table 2). In conclusion, similar to the results obtained from the Mainz patients, Ep-CAM RNA is also an independent prognostic factor in the Transbig cohort.

In a next step the cohorts were combined and the influence of Ep-CAM RNA expression was visualized by Kaplan Meier analysis. In the Transbig cohort alone (n=302) Ep-CAM showed a significant association with metastasis free survival ( $P=0.042$ ; Fig. 3A), whereas in the Rotterdam cohort only a trend was obtained ( $P=0.089$ ; Fig. 3B). If the Transbig and Rotterdam cohorts were combined increasing the case number increased to n=588 a p-value of  $P=0.002$  was obtained (Fig. 3C). Not all clinical parameters of the Rotterdam cohort are publicly available. Therefore, this group of patients could not be analyzed by multivariate Cox regression. Next, we included also the Mainz patients into a large combined cohort with 782 patients, resulting in a p-value of  $P<0.001$  (Fig. 3D). Therefore, the prognostic relevance of Ep-CAM initially analyzed in the Mainz cohort (Fig. 2, Table 2) could be confirmed in two further cohorts of node negative breast cancer.

### **Ep-CAM correlates with the proliferation metagene and is associated with the intrinsic subtype**

Recently, progress has been achieved in predicting breast cancer prognosis by gene expression metagenes [14, 18]. Therefore, we analyzed, whether Ep-CAM correlates with these metagenes. A relatively good correlation was observed with the proliferation metagene ( $P<0.001$ ,  $R=0.425$ , Supplemental Fig. 5). Therefore, one might expect that the metagene that comprises the information of a multitude of genes performs superior over an individual gene in multivariate analysis. Surprisingly, this was not the case. When all three cohorts were combined (n=782) Ep-CAM was significant in multivariate analysis ( $P=0.004$ ), in contrast to the proliferation metagene ( $P=0.114$ ) (Table 4A). Also if the individual subcohorts were analyzed, including the clinical factors, Ep-CAM and not the proliferation metagene was significant in multivariate analysis (Table 4B).

Recently, Sieuwerts and co-workers have analyzed Ep-CAM expression in breast cancer cell lines with different intrinsic subtype characteristics [19]. They reported that cell lines of the "normal-like" subgroup express particularly low levels of Ep-CAM. Therefore, we clustered our patients into intrinsic subtypes and compared levels of Ep-CAM RNA (Supplemental Fig. 6). We confirmed the observation of Sieuwerts and co-workers [19]. Expression of Ep-CAM was significantly lower in normal-like carcinomas compared to basal-like, ERBB2-like or luminal B subtypes ( $P < 0.001$ ; Supplemental Fig. 6).

Gene array data presented in this study were log2 transformed and dichotomized cohortwise at the 75% percentile. We used this type of transformation, because in previous studies of adverse prognostic factors we have observed that particularly the 25% of patients with highest expression show an increased hazard ratio [20, 16]. However, the prognostic influence of Ep-CAM RNA does not depend on a particular mode of data transformation, since also untransformed Ep-CAM RNA expression data and log2 transformed data dichotomized at the median instead of the 75% percentile also resulted in significant associations.

## Discussion

Even though Ep-CAM was initially described as a cellular adhesion molecule [1], recent insights revealed a more complex role for Ep-CAM, not merely limited to cell adhesion but also facilitating processes such as cell signalling, cell migration, proliferation, and differentiation [21]. Silencing Ep-CAM gene expression with Ep-CAM short interfering RNA resulted in a substantial decrease in the rate of cell proliferation, migration and invasion in breast cancer cell lines [22] which supports the notion that the role of Ep-CAM in carcinomas is pleiotropic [21]. Ep-CAM overexpression has been reported to be associated with adverse prognosis in node-positive [8] as well as node-negative breast cancer [10], whereby Ep-CAM has been evaluated by immunostaining. Compared with immunostaining, RNA analysis has the advantage of less inter- and intra-observer variability. Furthermore, the results can be validated in previously published cohorts of patients characterized by gene array analyses. To validate Ep-CAM RNA expression as a possible prognostic factor in breast cancer, we first studied the correlation between Ep-CAM RNA and protein expression. Affymetrix HG-U133A and immunostaining data was available for 194 node-negative breast carcinomas from the Mainz cohort. Ep-CAM status positive tumors had significantly higher levels of Ep-CAM RNA expression. However, there was no linear relationship between immunostaining and RNA expression data. This became obvious when we compared the Ep-CAM score, a product of the fraction of positive cells and their staining intensity, with RNA expression levels (Fig. 1). Although the highest and lowest scores were associated with high and low RNA expression no correlation was seen for the intermediate scores. Therefore, we next compared the prognostic relevance of RNA and protein in the Mainz cohort. Interestingly, Ep-CAM expression was associated with disease free survival and also with metastasis free interval, independently from the clinically established factors

age, pTstage, grading, hormone receptor and HER-2 status, whereas Ep-CAM immunostaining data showed only a trend. The latter result does not contradict previously published studies where Ep-CAM was reported to be associated with prognosis, because the discrepancy can be explained by differences in case numbers. Both, RNA and immunostaining data were available in only 194 patients (present study), whereas the previously published associations between Ep-CAM immunostaining and prognosis were found in cohorts of 1,705 [8] and 402 [10] patients. In a next step, we validated the prognostic relevance of Ep-CAM RNA expression in independent cohorts. Very similar results as for the Mainz cohort were found in the previously published Transbig cohort [12, 13] of 302 node negative breast cancer patients, where Ep-CAM RNA was also associated with metastasis free interval independent from age, pTstage, grading, hormone receptor and HER-2 status. Unfortunately, a third cohort, the Rotterdam patients of node-negative breast cancer [11] could not be evaluated by multivariate analysis, because the clinical data have not been completely published. However, in Kaplan-Meier analysis the single cohorts as well as combined cohorts were associated with survival time. The relevance of Ep-CAM is also illustrated by the decreasing p-value of the log-rank-test with increasing case numbers with  $P=0.042$  for the Transbig cohort only ( $n=302$ ),  $P=0.002$  for the combined Transbig plus Rotterdam cohorts ( $n=588$ ) and  $P<0.001$  for all three cohorts with  $n=782$  patients.

Recently, a genome wide overview of the natural history of node-negative breast cancer has been published, in which three major biological axes, (i) the transition from slow to fast proliferation, (ii) immune cell infiltration and (iii) steroid hormone receptor expression allow deeper understanding of biological processes relevant for breast cancer progression [14, 18]. Transition from slow to fast proliferation, leads to

the most dramatic aggravation of prognosis. When we studied the relationship between Ep-CAM and the above mentioned “coordinates in the universe of node-negative breast cancer” [18], we observed a strong correlation between Ep-CAM RNA expression and the proliferation metagene. A metagene roughly represents a distinct biological motif as indicated by the genes constituting the respective cluster. Proliferation was the most significant independent prognostic motif in earlier studies [14, 23-26]. Since we initially expected that the prognostic information of Ep-CAM expression may be already covered by the powerful proliferation metagene, we included both, Ep-CAM and the proliferation metagene, into the multivariate Cox model. Bearing the well documented strong correlation of proliferation and outcome in mind, it is astonishing that a single gene, Ep-CAM, is superior to the proliferation metagene in multivariate analysis (Table 3). On the other hand, considering the numerous potential functions of Ep-CAM aside from proliferation like cell adhesion [1], cell migration [22], cell cycle regulation [27], apoptosis resistance [28], tumor immune evasion [29] and metastasis [30], it is conceivable that Ep-CAM indeed possesses a significant association with outcome of breast cancer patients independent of proliferation. In conclusion, our large-scale validation of the prognostic significance of Ep-CAM RNA expression for the natural history of node-negative breast cancer makes it a valuable prognostic factor independent from established clinical factors and from the proliferation metagene which might serve as a suitable objective for targeted therapies.

Acknowledgment: This study was supported by the BMBF (NGFN project Oncoprofile). We thank Ms. Susanne Lindemann for competent bibliographic assistance.

## References

1. Litvinov SV, Velders MP, Bakker HA et al (1994) Ep-CAM: a human epithelial antigen is a homophilic cell-cell adhesion molecule. *J Cell Biol* 125:437–46
2. Naundorf S, Preithner S, Mayer P et al (2002) In vitro and in vivo activity of MT201, a fully human monoclonal antibody for pancarcinoma treatment. *Int J Cancer* 100:101–10
3. Prang N, Preithner S, Brischwein K et al (2005) Cellular and complement-dependent cytotoxicity of Ep-CAM-specific monoclonal antibody MT201 against breast cancer cell lines. *Brit J Cancer* 92:342–9
4. Amann M, Brischwein K, Lutterbuese P et al (2008) Therapeutic window of MuS110, a single-chain antibody construct bispecific for murine EpCAM and murine CD3. *Cancer Res* 68:143–51
5. Sebastian M, Kiewe P, Schuette W et al (2009) Treatment of malignant pleural effusion with the trifunctional antibody catumaxomab (Removab) (anti-EpCAM x Anti-CD3): results of a phase 1/2 study. *J Immunother* 32:195–202
6. Schmidt M, Scheulen ME, Ditttrich C et al (2010) An open-label, randomized phase II study of adecatumumab, a fully human anti-EpCAM antibody, as monotherapy in patients with metastatic breast cancer. *Ann Oncol* 21: 275-82
7. Gastl G, Spizzo G, Obrist P et al (2000) Ep-CAM overexpression in breast cancer as a predictor of survival. *Lancet* 356:1981–2
8. Spizzo G, Went P, Dirnhofer S et al (2004) High Ep-CAM expression is associated with poor prognosis in node-positive breast cancer. *Breast Cancer Res Treatment* 86:207–13
9. Hengstler JG, Bockamp EO, Hermes M et al (2006) Oncogene-blocking therapies: new insights from conditional mouse tumor models. *Curr Cancer Drug Targets* 6:603–12
10. Schmidt M, Hasenclever D, Schaeffer M et al (2008) Prognostic effect of epithelial cell adhesion molecule overexpression in untreated node-negative breast cancer. *Clin Cancer Res* 14:5849–55
11. Wang Y, Klijn JGM, Zhang Y et al (2005) Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* 365:671–9
12. Desmedt C, Piette F, Loi S et al (2007) 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* 13:3207–14
13. Loi S, Haibe-Kains B, Desmedt C et al (2007) Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. *J Clin Oncol* 25:1239–46
14. Schmidt M, Böhm D, von Törne C et al (2008) The humoral immune system has a key prognostic impact in node-negative breast cancer. *Cancer Res* 68:5405–13
15. McShane LM, Altman DG, Sauerbrei W et al (2005) Reporting recommendations for tumor marker prognostic studies. *J Clin Oncol* 23:9067-72



16. Hengstler JG, Lange J, Kett A et al (1999) Contribution of c-erbB-2 and topoisomerase IIalpha to chemoresistance in ovarian cancer. *Cancer Res* 59:3206–14
17. Smid M, Wang Y, Zhang Y et al (2008) Subtypes of breast cancer show preferential site of relapse. *Cancer Res* 68: 3108-14
18. Schmidt M, Hengstler JG, von Törne C et al (2009) Coordinates in the universe of node-negative breast cancer revisited. *Cancer Res* 69:2695-8
19. Sieuwerts AM, Kraan J, Bolt J et al (2009) Anti-epithelial cell adhesion molecule antibodies and the detection of circulating normal-like breast tumor cells. *J Natl Cancer Inst* 101: 61-6
20. Tanner B, Hasenclever D, Stern K et al (2006) ErbB-3 predicts survival in ovarian cancer. *J Clin Oncol* 24:4317–23
21. Trzpis M, McLaughlin PMJ, Leij LMFH de et al (2007) Epithelial cell adhesion molecule: more than a carcinoma marker and adhesion molecule. *Am J Pathol* 171:386–95
22. Osta WA, Chen Y, Mikhitarian K et al (2004) EpCAM is overexpressed in breast cancer and is a potential target for breast cancer gene therapy. *Cancer Res* 64:5818–24
23. Oh DS, Troester MA, Usary J et al (2006) Estrogen-regulated genes predict survival in hormone receptor-positive breast cancers. *J Clin Oncol* 24:1656–64
24. Dai H, van't Veer L, Lamb J et al (2005) A cell proliferation signature is a marker of extremely poor outcome in a subpopulation of breast cancer patients. *Cancer Res* 65:4059–66
25. Sotiriou C, Wirapati P, Loi S et al (2006) Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J Natl Cancer Inst* 98:262–72
26. Fan C, Oh D, Wessels L et al (2006) Concordance among gene-expression-based predictors for breast cancer. *N Engl J Med* 355:560–9
27. Münz M, Zeidler R, Gires O (2005) The tumour-associated antigen EpCAM upregulates the fatty acid binding protein E-FABP. *Cancer Lett* 225:151–7
28. Nübel T, Preobraschenski J, Tuncay H et al (2009) Claudin-7 regulates EpCAM-mediated functions in tumor progression. *Mol Cancer Res* 7:285–99
29. Ziegler A, Heidenreich R, Braumüller H et al (2009) EpCAM, a human tumor-associated antigen promotes Th2 development and tumor immune evasion. *Blood* 113:3494–3502
30. Schmidt D-S, Klingbeil P, Schnölzer M et al (2004) CD44 variant isoforms associate with tetraspanins and EpCAM. *Exp Cell Res* 297:329–47

## Figure legends

**Fig. 1:** A. Frequency distribution of Ep-CAM RNA expression. B. Definition of the Ep-CAM status. C. Relationship between immunohistochemically determined Ep-CAM status and Ep-CAM RNA expression. D. Immunohistochemically determined Ep-CAM score is associated with levels of Ep-CAM RNA expression. Log2 transformed gene array data of 194 patients (Mainz cohort) were analyzed in A, C and D.

**Fig. 2:** High Ep-CAM RNA levels is associated with shorter disease free survival time (A), shorter metastasis free survival time (B) and shorter overall survival time in the Mainz cohort of node negative breast cancer patients. Levels of Ep-CAM RNA were dichotomized at the median. The respective data after dichotomization at the 75% percentile resulted in similar associations between Ep-CAM RNA and prognosis (Supplemental Fig. 1). The numbers below the diagrams represent patients at risk at the time point indicated on the x-axis.

**Fig. 3:** Validation of the association between Ep-CAM RNA levels and metastasis free survival time in additional, independent cohorts. The association was confirmed in the individual and particularly in the combined cohorts. The numbers below the diagrams represent patients at risk at the time point indicated on the x-axis.

**Table 1:** Clinicopathological characteristics of node negative breast cancer patients from the Mainz cohort (n=194)

Characteristics	n	%
<b>Age at diagnosis</b>		
<50	45	23.2
≥50	149	76.8

<b>pT stage</b>		
pT <sub>1</sub>	106	54.6
pT <sub>2</sub>	85	43.8
pT <sub>3</sub>	3	1.5
<b>Histological grade</b>		
G I	41	21.1
G II	105	54.1
G III	48	24.7
<b>Estrogen receptor status</b>		
Negative	45	23.2
Positive	149	76.8
<b>Progesterone receptor status</b>		
Negative	81	41.8
Positive	113	58.2
<b>Hormone receptor status<sup>1</sup></b>		
Negative	39	20.1
Positive	155	79.9
<b>HER-2 status</b>		
Negative	168	86.6
Positive	26	13.4
<b>Death</b>		
Of cancer	31	16.0
Unrelated to cancer	25	12.9
Surviving	138	71.1
<b>Relapse</b>		
Regional	58	29.9
Metastasis	21	10.8
Contralateral	45	23.2
No relapse	5	2.6
	136	70.1

<sup>1</sup>The hormone receptor status is positive as soon as one of both, the estrogen or the progesterone receptor status, is positive.

**Table 2:** Association of Ep-CAM RNA expression with breast cancer specific disease-free survival (DFS; A, B) and metastasis free survival (MFS; C, D) in the Mainz cohort of node negative breast cancer patients (n=194).

A. Univariate Cox analysis

<b>Prognostic factor</b>	<b>p</b>	<b>HR</b>	<b>95% - CI</b>
Ep-CAM RNA expression	0.001	2.371	1.394-4.034

#### B. Multivariate Cox analysis

<b>Prognostic factors</b>	<b>p</b>	<b>HR</b>	<b>95% - CI</b>
Age (<50 vs ≥50 years)	0.514	0.825	0.463-1.470
pT stage (≤2cm vs >2cm)	0.590	0.859	0.495-1.492
Histological grade (Grade 3 vs grade 1 and 2)	<0.001	4.802	2.729-8.449
HR <sup>1</sup> (ER or PR) (negative vs. positive)	0.460	1.268	0.676-2.379
HER-2 status (positive vs negative)	0.733	1.128	0.563-2.260
Ep-CAM RNA Expression	0.017	1.994	1.129-3.523

### C. Univariate Cox analysis

<b>Prognostic factor</b>	<b>p</b>	<b>HR</b>	<b>95% - CI</b>
Ep-CAM RNA expression	0.003	2.476	1.362-4.499

### D. Multivariate Cox analysis

<b>Prognostic factors</b>	<b>p</b>	<b>HR</b>	<b>95% - CI</b>
Age (<50 vs ≥50 years)	0.700	0.876	0.446-1.718
pT stage (≤2cm vs >2cm)	0.886	1.046	0.565-1.939
Histological grade (Grade 3 vs grade 1 and 2)	<0.001	4.689	2.477-8.875
HR <sup>1</sup> (ER or PR) (negative vs. positive)	0.293	1.471	0.716-3.022
HER-2 status (positive vs negative)	0.306	1.475	0.701-3.103
Ep-CAM RNA Expression	0.049	1.936	1.003-3.735

<sup>1</sup>The hormone receptor status (HR) is positive as soon as one of both, the estrogen (ER) or the progesterone receptor status (PR), is positive.

**Table 3:** Association of Ep-CAM RNA expression with **metastasis free survival (MFS)** in the Transbig cohort of node negative breast cancer patients (n=302).

A. Univariate Cox analysis

Prognostic factor	p	HR	95% - CI
Ep-CAM RNA expression	0.006	1.908	1.203-3.026

B. Multivariate Cox analysis

Prognostic factors	p	HR	95% - CI
Age (<50 vs ≥50 years)	0.176	1.398	0.861-2.270
pT stage (≤2cm vs >2cm)	<0.001	3.230	1.696-6.154
Histological grade (Grade 3 vs grade 1 and 2)	0.396	0.789	0.457-1.362
HR <sup>1</sup> (ER or PR) (RNA high vs. low)	0.525	0.834	0.477-1.459
HER-2 (RNA high vs. low)	0.754	1.102	0.601-2.021
Ep-CAM RNA Expression	0.035	1.734	1.038-2.895

<sup>1</sup>The hormone receptor status (HR) is high as soon as RNA expression of one of both, the estrogen (ER) or the progesterone receptor status (PR), is high.

**Table 4:** EP-CAM RNA expression is independent from the proliferation metagene. Association of Ep-CAM RNA expression with **metastasis free survival (MFS)** was analyzed in the combined cohorts (Transbig, Rotterdam and Mainz) and in the Transbig cohort of node negative breast cancer patients.

A. Multivariate Cox analysis of the combined cohort (n=782)

<b>Prognostic factor</b>	<b>p</b>	<b>HR</b>	<b>95% - CI</b>
Ep-CAM RNA expression	0.004	1.519	1.141-2.021
Proliferation metagene	0.114	1.269	0.945-1.704

B. Multivariate Cox analysis of the Transbig cohort (n=302)

<b>Prognostic factors</b>	<b>p</b>	<b>HR</b>	<b>95% - CI</b>
Age (<50 vs ≥50 years)	0.179	1.395	0.859-2.266
pT stage (≤2cm vs >2cm)	< 0.001	3.163	1.655-6.045
Histological grade (Grade 3 vs grade 1 and 2)	0.770	0.916	0.508-1.650
HR <sup>1</sup> (ER or PR) (RNA high vs. low)	0.325	0.743	0.411-1.342
HER-2 (RNA high vs. low)	0.827	1.070	0.583-1.964
Proliferation metagene	0.237	0.669	0.344-1.302
Ep-CAM RNA Expression	0.030	1.770	1.059-2.960

<sup>1</sup>The hormone receptor status (HR) is high as soon as RNA expression of one of both, the estrogen (ER) or the progesterone receptor status (PR), is high.