Expression profiling identifies genes that predict recurrence of breast cancer after adjuvant CMF-based chemotherapy

Katja Specht, Nadia Harbeck, Jan Smida, Katja Annecke, Ulrike Reich, Joerg Naehrig, Rupert Langer, Joerg Mages, Raymonde Busch, Elisabeth Kruse, et al.

Breast Cancer Res Treat (2009) 118:45–56
DOI 10.1007/s10549-008-0207-y

PRECLINICAL STUDY

Expression profiling identifies genes that predict recurrence of breast cancer after adjuvant CMF-based chemotherapy

Katja Specht · Nadia Harbeck · Jan Smida · Katja Annecke · Ulrike Reich · Joerg Naehrig · Rupert Langer · Joerg Mages · Raymonde Busch · Elisabeth Kruse · Ludger Klein-Hitpass · Manfred Schmitt · Marion Kiechle · Heinz Hoefler

Received: 28 December 2007 / Accepted: 19 September 2008 / Published online: 17 October 2008

Abstract Cyclophosphamide, methotrexate and 5-fluorouracile (CMF)-based chemotherapy for adjuvant treatment of breast cancer reduces the risk of relapse. In this exploratory study, we tested the feasibility of identifying molecular markers of recurrence in CMF-treated patients. Using Affymetrix U133A GeneChips, RNA samples from 19 patients with primary breast cancer who had been uniformly treated with adjuvant CMF chemotherapy were analyzed. Two supervised class prediction approaches were used to identify gene markers that can best discriminate between patients who would experience relapse and patients who would remain disease-free. An additional independent validation set of 51 patients and 21 genes were analyzed by quantitative RT-PCR. Applying different algorithms to evaluate our microarray data, we identified two gene expression signatures of 21 and 12 genes containing eight overlapping genes, that predict recurrence in 19 cases with high accuracy (94%). Quantitative RT-PCR demonstrated that six genes from the combined signatures (*CXCL9*, *ITSN2*, *GNAI2*, *H2AFX*, *INDO*, and *MGC10986*) were significantly differentially expressed in the recurrence versus the non-recurrence group of the 19 cases and the independent breast cancer patient cohort ($n = 51$) treated with CMF. High expression levels of *CXCL9*, *ITSN2*, and *GNAI2* were associated with prolonged disease-free survival (DFS) ($P = 0.029$, 0.018 and 0.032, respectively). When patients were stratified by combined *CXCL9/ITSN2* or *CXCL9/FLJ22028* tumor levels, they exhibited significantly different disease-free survival curves ($P = 0.0073$ and $P = 0.005$, respectively). Finally, the *CXCL9/ITSN2* and *CXCL9/FLJ22028* ratio was an independent prognostic factor ($P = 0.034$ and $P = 0.003$, respectively) for DFS by multivariate Cox analysis in the 70-patient cohort. Our data highlight the feasibility of a prognostic assay that is applicable to therapeutic decision-making for breast cancer. Whether the biomarker profile is chemotherapy-specific or whether it is a more general indicator of bad prognosis of breast cancer patients remains to be explored.

Keywords Adjuvant CMF Chemotherapy · Breast Cancer · Microarray · Prediction · Quantitative RT-PCR · Recurrence

Electronic supplementary material The online version of this article (doi:10.1007/s10549-008-0207-y) contains supplementary material, which is available to authorized users.

K. Specht (✉) · J. Naehrig · R. Langer · H. Hoefler
Institute of Pathology, Technical University Munich, Ismaningerstrasse 22, 81675 Munich, Germany
e-mail: specht@lrz.tu-muenchen.de

N. Harbeck · K. Annecke · M. Schmitt · M. Kiechle
Department of Obstetrics and Gynecology, Klinikum Rechts der Isar, Technical University Munich, 81675 Munich, Germany

J. Smida · U. Reich · H. Hoefler
Institute of Pathology, GSF – National Research Centre for Environment and Health, 85764 Neuherberg, Germany

J. Mages
Institute of Medical Microbiology, Immunology and Hygiene, Technical University Munich, 81675 Munich, Germany

R. Busch
Institute of Medical Statistics and Epidemiology, Klinikum Rechts der Isar, Technical University Munich, 81675 Munich, Germany

E. Kruse
Institute of Medical Statistics, Biometry and Epidemiology, University Hospital Essen, 45122 Essen, Germany

L. Klein-Hitpass
Institute of Cell Biology, IFZ, University Hospital Essen, 45122 Essen, Germany

Introduction

Cyclophosphamide, methotrexate and 5-fluorouracil (CMF)-based chemotherapy is a chemotherapeutic regimen to treat breast cancer patients [1]. The agents cyclophosphamide and 5-fluorouracil are also part of another widely used polychemotherapy regimen containing anthracyclines such as epirubicin, FEC (5-fluorouracil, epirubicin, cyclophosphamide). Large metaanalyses have demonstrated to provide disease-free survival (DFS) as well as overall survival (OS) benefits in breast cancer patients treated with adjuvant chemotherapy [2, 3].

CMF as a regimen of six cycles of polychemotherapy has earlier been one of the most frequently used therapies in primary breast cancer patients. CMF has been shown to improve overall survival resulting in a 34% reduction in the relative risk of relapse and a 26% reduction in the relative risk of death [4]. In the late 1990s, however this regimen was shown to be less effective than a more intensive anthracycline-based regimen [5]. CMF was then considered to be reserved for patients at minor risk of disease recurrence, determined by factors such as the absence of axillary lymph node metastasis. According to the St Gallen consensus meeting 2005, patients with node-positive disease should be treated with more intensive regimens containing anthracyclines and taxanes [6]. However, recent data on predictive factors for anthracycline response seem to indicate that there is a substantial proportion of patients for whom CMF and six cycles of a 3-drug anthracycline combination yield comparable outcome results [7]. Moreover, CMF remains a recommended treatment option for patients in whom anthracyclines are contraindicated as reported in the most recent guidelines of the Breast Commission of the Gynecological Oncology Working Group (AGO) (www.ago-online.com).

Since CMF is still a valid therapy option in primary breast cancer patients, understanding of the development of resistance or failure to this therapy is an important item. Current clinical and pathological markers such as tumor size, nuclear grade, estrogen status or single molecular markers (i.e., p53 mutations) poorly predict the clinical course of these patients, especially patients that are at a high risk of tumor recurrence [8, 9]. In those patients, the use of alternative chemotherapeutic regimens might improve the clinical outcome. Therefore, the ability to predict the CMF treatment outcome should facilitate treatment planning. Moreover, in view of recently reported findings on other therapies, regimen-characteristic gene expression signatures are comparable in order to identify common or therapy-specific drug resistance pathways [10].

To investigate markers that have the potential to discriminate chemotherapy-sensitive, good outcome patients and chemotherapy-resistant, poor outcome breast cancer patients, we performed an exploratory microarray analysis using Affymetrix U133A oligonucleotide arrays from tumor samples of 19 patients ("training set") who after surgery had been uniformly treated by adjuvant CMF chemotherapy. QRT-PCR was subsequently used to apply the identified predictive gene set to 51 tumors from breast cancer patients ("validation group") treated by adjuvant CMF.

Materials and methods

Patients

We selected from our tumor bank of the Institute of Pathology, Technical University Munich fresh-frozen tissue specimens from patients with primary breast cancer who were treated during 1989–1998 at the Department of Obstetrics and Gynecology, Technical University Munich, Germany. After primary surgery (modified radical mastectomy or breast conserving therapy), patients received adjuvant chemotherapy, consisting of six cycles of CMF. CMF was administered intravenously, on days 1 and 8, and recycled on day 28, at the following doses: cyclophosphamide 500 mg/m^2, methotrexate 40 mg/m^2, and fluorouracil 600 mg/m^2. Treatment decision was based on consensus recommendations at the time and no additional endocrine treatment was given. Fresh-frozen tumor tissue was available from 79 consecutive patients that were receiving CMF adjuvant chemotherapy between 1989 and 1998. Seventy nine samples were processed and 9 samples had to be excluded on the basis of poor RNA quality; thus, 70 samples were eligible for further analysis. Patient characteristics are summarized in Table 1. Written informed consent for research use of tissue was obtained from all patients. Tumor samples were split into 19 samples where large quantities of tumor material was available in the tumor bank. However, this did not necessarily reflect the size of the original primary tumor, but depended on the tissue size available in the tumor bank. These samples were used for the predictive marker discovery ("training set"). The remaining tumor samples from 51 patients were available for an independent validation set ("validation set").

Follow-up data were obtained at regular intervals. Median follow-up time of all of the patients still alive at the time of analysis was 96.5 months (range, 32–174 months). Within the follow-up period, 31 patients (44%) had disease recurrence and of these, 22 patients (31%) died. Median follow-up time of the patients that didn't have disease recurrence was 99.5 months (range, 39–174 months).

Tumors were classified according to the pTNM (pathological tumor-node-metastasis) system and grading was

Table 1 Clinical information and demographics of the patients included in the study ($N = 70$)

	Training cases		Validation cases		All cases	
	N	%	*N*	%	*N*	%
Sex (female)	19	100	51	100	70	100
Age (years)						
Median	48		49		48.5	
Range	33–63		33–65		33–65	
Histology						
Invasive ductal	13	68	39	76	52	74
Invasive lobular	3	16	4	8	7	10
Mixed ductal/lobular	2	11	5[a]	10	7	10
Medullary	1	5	3	6	4	6
pTNM stage						
pT1	8	42	14	27	22	31.4
pT2	10	53	28	55	38	54.3
pT3/T4	1	5	9	18	10	14.3
pN0	2	11	18	35	20	28.5
pN1	9	47	21	41	30	43
pN2/N3	8	42	12	24	20	28.5
Nuclear grade						
1	0	0	0	0	0	0
2	12	63	22	43	34	48.5
3	7	37	29	57	36	51.5
ER/PR—positive	16	84	34	67	50	71.5
ER/PR—negative	3	16	17	33	20	28.5
Follow-up recurrent patients ($n = 31$), months						
Median	62		44.5		54	
Range	23–107		17–123		17–123	
Follow-up non-recurrent patients ($n = 39$), months						
Median	107		99		99	
Range	40–150		39–174		39–174	
Follow-up all patients still alive ($n = 48$), months						
Median					96.5	
Range					32–174	

N number of patients; *pTNM* pathological tumor-node-metastasis staging; *ER* estrogen receptor; *PR* progesterone receptor

[a] One case tubulo-lobular

performed according to Elston and Ellis [11]. Steroid hormone receptor content was established by immunohistochemical means. A graphical depiction of the selection process of the cases for the study is shown in Fig. 1.

RNA isolation and amplification

RNA was extracted from frozen tumor samples. To determine the histological representativeness of the tumors, frozen sections were cut and H&E stained. Tumor cells were then manually macrodissected by scraping off the cells from the slide with the help of a small sterile needle and pooled from 20 sections of 10 μm thickness to ensure at least a 70% pure tumor cell population. Total RNA was isolated with Trizol Reagent (Invitrogen, life technologies, Karlruhe, Germany) and was subsequently passed over a Qiagen RNeasy column (Qiagen, Hilden, Germany) to remove small fragments. RNA quality was assessed on the Agilent 2100 Bioanalyzer RNA 6000 Nano LabChip (Agilent Technologies, Palo Alto, CA). Only samples yielding profiles of intact total RNA (retention of both ribosomal bands and the broad central peak of mRNA) were used for the microarray and quantitative RT-PCR gene expression analyses.

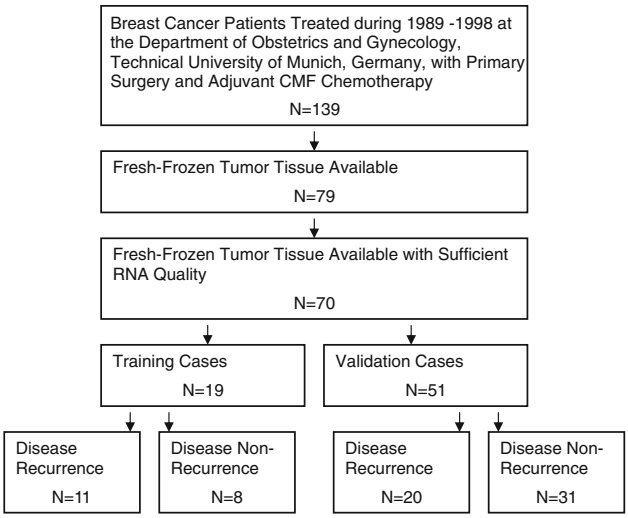

Fig. 1 Graphical depiction of the selection process of the patient's tumor RNA for the study

Preparation of in vitro transcription (IVT) products, oligonucleotide array hybridization and scanning was performed according to Affymetrix (Santa Clara, CA). In brief, 10–16 μg of total RNA from each breast tumor and T7-linked oligo-dT primers were used for first-strand cDNA synthesis. After IVT and biotin-labeling (RiboMax T7 kit, Promega, Madison, WI), fragmented biotinylated cRNA (15 μg) was hybridized onto the U133A GeneChip (Affymetrix) using recommended procedures for prehybridization, washing and staining with streptavidin–phycoerythrin (SAPE).

Microarray data processing and statistical analysis

Array images were analyzed using the Affymetrix Microarray Suite 5.0 software. Scaling across all probe sets of a given array to an average intensity of 1,000 was performed to compensate for variations in the amount and quality of the cRNA samples and other experimental variables of non-biological origin. Samples displaying a signal ratio for the β-actin 3′ and 5′ probesets of >3.0 were

considered as poor quality targets and excluded from the data set. The training set of the study contained 19 arrays (11 samples from patients with recurrence and eight samples from patients without recurrence). To reduce noise associated with low expressed transcripts, probesets receiving equal to or less than 30% present detection calls in both the recurrence and the non-recurrence group were eliminated. The signals of the remaining probesets (13,132) were used to perform prediction analyses employing two different algorithms. The first approach uses the k-nearest neighbors (knn) supervised learning method [12, 13]. Predictors consisting of 1–100 features (probesets) were determined using GeneCluster 2.0 software (http://www-genome.wi.mit.edu/cancer/software/genecluster2/gc2.html) and cross-validation of the resulting predictors was done by the leave-one-out procedure.

The second classification method uses an approach similar to the successful methods described previously [14, 15]. In short, expression values were transformed by taking logarithms and performing a z-transformation. We then calculated the correlation between the prognostic category (relapse vs. no relapse) and the expression ratio for each gene of the probesets (13,132) using the Bravais/Pearson correlation coefficient. The genes were ranked on the basis of the magnitude of the correlation coefficient. A tumor was classified as "relapse-free", if the correlation coefficient of the expression profile with the masterprofile was positive and if the resulting *P*-value for the correlation was ≤ 0.1. In a second step, the optimal set of reporter genes was determined starting with the three top genes using a leave-one-out cross validation procedure. The predictor was termed optimal when the number of misclassified tumors was minimal.

Real-time RT-PCR, data processing and prediction

Real-time RT-PCR was performed using TaqMan 384-well microfluidic card sets on an ABI PRISM 7900HT Sequence Detection system (Applied Biosystems). Sample RNA was extracted and quantified as for the microarray analysis and cDNA synthesis was carried out with Superscript Reverse Transcriptase (Invitrogen, life technologies, Karlruhe, Germany). 5.2 ng cDNA was used for 48 wells of the 384-well card. Gene-specific primers and probes for each gene were obtained from Assays-on-Demand Gene Expression Products (Applied Biosystems, Warrington, UK). Transcript levels of 21 genes of interest (Table 2) were determined in duplicates by the deltaCt method with 18S ribosomal RNA transcript as an internal reference. For prediction the PCR data set was split into a training set, which contained the very same 19 samples previously analyzed by microarray analysis and a validation set, which contained the remaining samples. Prediction analysis using predictors with 1–21 features derived from the training set were used for cross-validation using the leave-one-out procedure and used to classify the tumors of the test set as described for the microarray data analysis.

Heat maps

For heat map representation, expression measures were normalized to the mean expression value and log transformed (log10). The brightest red corresponds to 4-fold overexpression, the brightest green to 4-fold underexpression and black color represents the mean expression. In some cases, genes and samples were subjected to UPGMA clustering (Spotfire DecisionSite for Functional Genomics) using log transformed mean-normalized expression values and correlation as a similarity measure.

Statistical analysis

Time to first relapse was estimated and graphically presented according to the method of Kaplan and Meier. Differences between different subgroups of patients were assessed by a log-rank test for censured survival data. Relapse was defined as the first reappearance of breast cancer at any local or distant site. Significance level was set to 0.05 two-sided. Multivariate analysis was performed using the Cox proportional hazards regression model, with the clinical variables tumor size, grade, lymph node status, ER and PR content, and the *CXCL9*/*ITSN2* and *CXCL9*/*FLJ22028* ratios.

Results

Identification of differentially expressed genes and development of a gene predictor set

To identify differentially expressed genes between the recurrence and the non-recurrence group, we first analyzed primary tumor samples from 19 breast cancer patients who had been uniformly treated with adjuvant CMF chemotherapy after surgery. Within this cohort, 11 women developed disease recurrence with a median time to recurrence of 54 months and eight women remained disease-free with median follow-up of 107 months. While it would be optimal to study more samples for marker discovery, several studies have shown that successful class prediction is possible using a limited number of samples [16–18]. Oligonucleotide microarray-based gene expression profiling was carried out using Affymetrix U133A arrays containing 22,283 probe sets representing more than 14,500 well-substantiated human genes. The resulting

Table 2 List of Top 25 Predictor Genes

Rank[a]				Gene symbol	Annotation	Affymetrix probe set ID	Applied Biosystems[b]	Gene ontology
1	5	5	2	*ANGPTL2*	Angiopoietin-like 2//(NM_012098)	219514_at	Hs00171912_m1	Receptor binding//extracellular space
2	4	1	5	*CXCL9*	Chemokine (C-X-C motif) ligand 9//(NM_002416)	203915_at	Hs00171065_m1	Signal transduction//inflammatory response//chemotaxis
3	6	3	10	*A4GALT*	Alpha 1,4-galactosyltransferase//(NM_017436)	219488_at	Hs00213726_m1	Lipid biosynthesis//galactosyltransferase activity
4	2	2	7	*PGDS*	Prostaglandin D2 synthase, hematopoietic//(NM_014485)	206726_at	Hs00183950_m1	Signal transduction//prostaglandin biosynthesis
5	3	–	26	*INDO*	Indoleamine-pyrrole 2,3 dioxygenase//(NM_002164)	210029_at	Hs00158027_m1	Oxidoreductase activity//electron transporter activity
6	9	–	80	*ITSN*	Intersectin 2 (ITSN2), transcript variant 1//(NM_006277)	209898_x_at	Hs00219661_m1	Endocytosis//SH3/SH2 adaptor protein activity//calcium ion binding
7	7	–	28	*MGC10986*	Hypothetical protein MGC10986//(NM_030576)	218600_at	Hs00228984_m1	Electron transporter activity
–	1	9	8	*PDE4A*	Phosphodiesterase 4A, cAMP-specific (phosphodiesterase E2 dunce homolog, Drosophila)//(NM_006202)	211447_s_at	Hs00183479_m1	Signal transduction//3′,5′-cyclic-nucleotide phosphodiesterase
–	8	4	1	*APOM*	Apolipoprotein M//(NM_019101)	205682_x_at	Hs00219533_m1	Membrane lipid metabolism//lipid transporter activity
–	10	6	9	*PRND*	Prion protein 2 (dublet)//(NM_012409)	222106_at	NA	Prion/doppel alpha-helical domain
–	11	–	105	*IDI1*	Isopentenyl-diphosphate delta isomerase//(NM_004508)	204615_x_at	NA	Cholesterol biosynthesis//magnesium ion binding
–	12	–	24	*PMF1*	Polyamine-modulated factor 1//(NM_007221)	206956_at	NA	Transcription coactivator activity
–	13	–	15	*CEBPA*	CCAAT/enhancer binding protein (C/EBP), alpha//(NM_004364)	204039_at	NA	Regulation of transcription, DNA-dependent//RNA polymerase II transcription factor activity
–	14	–	190	*GPR126*	G protein-coupled receptor 126//(NM_020455)	213094_at	Hs00607475_m1	Neuropeptide signaling pathway
–	15	–	16	*FLJ22028*	Hypothetical protein FLJ22028//(NM_024854)	219802_at	Hs00430724_m1	Disulfide oxidoreductase activity
–	16	–	21	*PEX1*	Peroxisome biogenesis factor 1//(NM_000466)	204873_at	Hs00166599_m1	Peroxisome-assembly ATPase activity
–	17	11	–	*MGC16824*	Esophageal cancer associated protein//(NM_020314)	203173_s_at	Hs00220422_m1	Unknown
–	18	–	29	*TIMM44*	Translocase of inner mitochondrial membrane 44 homolog//(NM_006351)	203092_at	Hs00197152_m1	Mitochondrial translocation//ATP binding
–	19	–	40	*TOX*	Thymus high mobility group box protein//(NM_014729)	204529_s_at	Hs00207075_m1	Regulation of transcription, DNA-dependent
–	20	–	18	*H2AFX*	H2A histone family, member X//(NM_002105)	205436_s_at	Hs00266783_s1	Chromosome organization and biogenesis//nucleosome assembly
–	–	8	22	*FLJ20257*	Hypothetical protein FLJ20257//(NM_019606)	219798_s_at	Hs00219704_m1	Unknown; probably methyltransferase activity
–	–	12	–	*PCDH16*	Protocadherin 16 dachsous-like (Drosophila)//(NM_003737)	218892_at	Hs00371436_m1	Calcium-dependent cell adhesion
–	–	10	31	*BGN*	Biglycan//(NM_001711)	201262_s_at	Hs00156076_m1	Extracellular matrix structural constituent
–	–	7	–	*GNAI2*	Guanine nucleotide binding protein (G protein), alpha inhibiting activity negative regulation polypeptide 2//(NM_002070)	201040_at	Hs00179998_m1	Signal transduction//negative regulation of adenylate cyclase activity
–	21	–	50	*TEM8*	Tumor endothelial marker 8//(NM_032208)	220092_s_at	Hs00216777_m1	Unknown//probably receptor activity

[a] Genes were rank-ordered according to the results of the 7-feature knn algorithm (1st column), 21-feature knn algorithm (2nd column), an algorithm similar to the already published approach [14], [15] (3rd column) and the Mann–Whitney test (4th column)

[b] Gene-specific primers and probes for each gene are from Assays-on-Demand Gene Expression Products (Applied Biosystems, Warrington, UK)

expression dataset was first filtered to reduce noise associated with low expressed transcripts. Probesets receiving equal to or <30% present detection calls in both the recurrence and the non-recurrence group were eliminated, leaving 13,132 probesets for the analysis.

An initial comparison of the gene expression profiles of recurrent versus non-recurrent breast tumors using a non-parametric Mann–Whitney-Test identified 464 probe sets that showed expression levels changes of at least 50% and were differentially expressed at a significance of $P < 0.05$. This represents 3.53% of the probesets available for the analysis after filtering. 254 probesets (54.7%) had significantly increased signal intensity (1.5–10-fold increases) in recurrent breast tumors and 210 probesets (45.3%) had reduced signal intensity (1.5–6.25-fold decreases). The 464 genes classified as most significantly differentially expressed at $P < 0.05$ are listed in the Supplementary Table 1.

To achieve a higher level of stringency for identification of differentially expressed genes, two algorithms for class prediction were used. First, predictors consisting of 1–100 features (probesets) were determined using the knn supervised learning method [12, 13] using GeneCluster 2.0 software. Cross-validation of the resulting predictors by the leave-one-out procedure showed that predictors consisting of 2–21 features classified 17–18 out of the 19 samples correctly when no confidence threshold was applied.

Class prediction models were built using 1, 7, 10, 21, 50 or 100 features (genes), and the training error for each model was calculated using leave-one-out cross validation (Supplementary Table 2). Although the accuracy of the models was comparable, the 7- and the 21-feature knn model were chosen for further study because they predicted most accurately the class distinctions of the training set (18 of 19 correct calls; 94% accuracy).

In addition, a second analytical approach similar to the successful methods described previously [14, 15] has been taken. After filtering and mathematical transformation of the data (see Materials and methods), a selection of discriminatory genes by their correlation with the clinical outcome (relapse vs. non-relapse) was performed and the optimal set of reporter genes was determined by the leave-one-out cross validation procedure. Using this approach, a set of 12 genes was found that predicted correctly 18 of 19 cases (94% accuracy) (Table 2). Four of these genes formed part of the 7-knn-feature model and eight of the 12-gene classifier were part of the 21-knn feature classifier.

A cluster map generated by using the 21 knn feature (Fig. 2a) and the 12-gene prognosis set (Fig. 2b) demonstrates the clear differences between the two clinical groups i.e., between patients who later relapsed and those who did not.

Confirmation of gene expression measurements by real-time RT-PCR

Gene expression values derived from Affymetrix microarray hybridization were further analyzed by real-time RT-PCR using the microfluidic card platform. The cards consisted of 24 primer and probe sets preloaded on the assay cards in duplicates. Assays for 21 genes included the 7 genes from the 7-feature knn, 10 additional genes derived from the 21-feature knn (four of the 21 genes were not available as predesigned primer sets from the company) and 11 out of 12 genes from the second statistical algorithm used in this study. The 18S ribosomal RNA was used as an internal reference. cDNA samples generated from the 19 primary tumor RNA samples previously analyzed by microarrays were subjected to real-time PCR and the resulting expression values were compared with the Affymetrix data. Pearson and Spearman rank correlations were positive for 20 genes and significantly positive for 16 and 15 of 21 genes, respectively (Table 3). A cluster analysis of the training set ($n = 19$ tumors) with RT-PCR derived values for 17 genes out of the 21-knn classifier (Fig. 2c) and 11 genes of the 12-gene prognosis set (Fig. 2d) confirms that 18/19 tumor were classified correctly.

Validation of predictor set on independent cases by real-time RT-PCR

We next analyzed the 21 genes by RT-PCR in an independent patient cohort ($n = 51$ patients) that had received the same treatment as the training set patients. Of these 51 patients, 20 had suffered a relapse with a median time to recurrence of 32 months and 31 patients had remained disease-free with a median follow-up of 99 months. We used the 17 genes from the 21-feature knn set and the 11 genes from the 12 gene prognosis set and attempted to predict relapse in the validation set. While the two gene sets predicted relapse in the training set with high accuracy both by microarray derived gene expression values as well as QRT-PCR data, they failed to predict the clinical outcome in the validation set (clinical outcome was misclassified in 60% of the cases).

Expression levels of candidate genes in breast cancer

The mRNA expression levels of the 21 genes measured by QRT-PCR showed a marked variation in the 70 breast cancer cases, which included the training set ($n = 19$ cases) and the validation set ($n = 51$ cases) of tumors. The median expression levels were 8.9 (range, 2.2–53) for *A4GALT*, 27 (range, 5.9–265) for *ANGPTL2*, 2 (0.5–14) for *APOM*, 125 (19–501) for *BGN*, 50 (3.9–1498) for *CXCL9*, 25 (8.1–112) for *FLJ20257*, 6.4 (1.8–38) for *FLJ22028*, 56

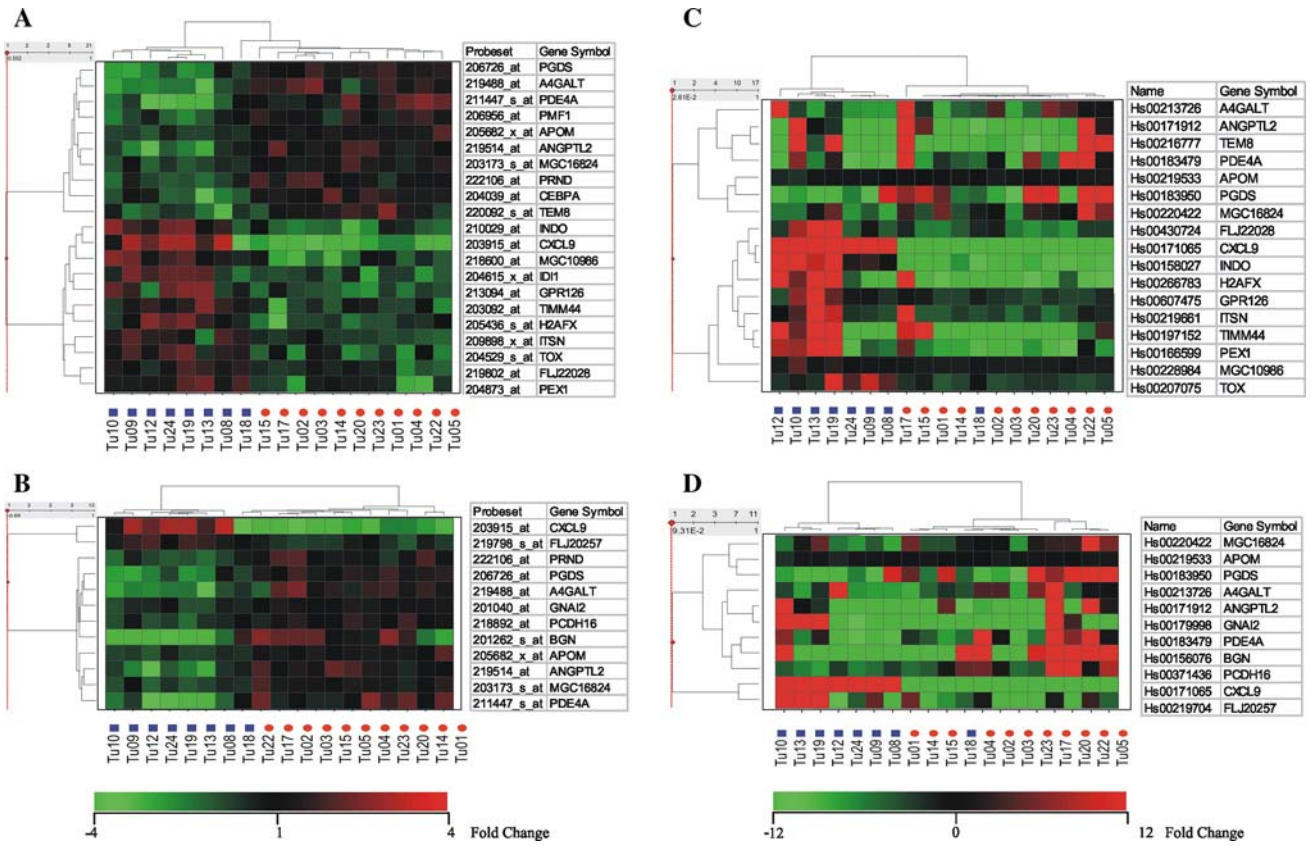

Fig. 2 Clustermap of differentially expressed genes in CMF-treated breast cancer patients who relapsed and those who did not. **a** and **b** A 21-feature (**a**) and a 12 gene-prognosis set (**b**) accurately classify 18/19 breast cancer cases using Affymetrix oligonucleotide-derived values. **c** and **d** Clustermap of differentially expressed genes generated with QRT-PCR-derived values using 17 genes from the 21-feature classifier (**c**) and 11 genes from the 12-gene prognosis set (**d**) correctly classify 18/19 tumors into the two categories. The red and blue dots below each dendrogramm indicate recurrences and non-recurrences, respectively. Bottom, color scale bar

(14–377) for *GNAI2*, 2.5 (0.4–28) for *GPR126*, 13 (1.6–124) for *H2AFX*, 5.3 (0.2–365) for *INDO*, 5.2 (2.5–36) for *ITSN2*, 0.7 (0.07–6) for *MGC10986*, 16 (4.2–48) for *MGC16824*, 8.6 (1.4–36) for *PCDH16*, 12 (4.6–45) for *PDE4A*, 11 (4.1–68) for *PEX1*, 12 (0.5–59) for *PGDS*, 58 (8.1–288) for *TEM8*, 42 (20–225) for *TIMM44*, and 3.2 (0.9–21) for *TOX*. To identify genes of predictive value, we evaluated the prognostic utility of each of these genes by itself. mRNA levels were categorized according to quartile distribution for gene expression values and either the lower (0.25) or the upper quartiles (0.75) were tested against the rest of the quartiles. Six genes (*ITSN2*, *CXCL9*, *GNAI2*, *H2AFX*, *INDO*, and *MGC10986*) of the 21 genes tested were significantly differentially expressed in the recurrence vs. the non-recurrence group ($n = 70$ patients; $P = 0.01$, 0.01. 0.01. 0.04, 0.04, and 0.04, respectively).

Expression levels in relation to DFS

The expression levels of the 21 genes measured by QRT-PCR were further analyzed in relation to the length of DFS. For this purpose, mRNA levels were categorized according to quartile distribution for gene expression values and either the lower (0.25) or the upper quartiles (0.75) were tested against the rest of the quartiles. Kaplan–Meier curves revealed that high expression levels of *CXCL9*, *ITSN2*, and *GNAI2* were significantly associated with a favorable clinical course in terms of length of DFS (log-rank = 4.74, $P = 0.0295$; log-rank = 4.61, $P = 0.0318$; log-rank = 5.58, $P = 0.018$, respectively; Table 4 and Fig. 3a, b). The expression levels of the other genes were not significantly related with the length of DFS (Table 4). We then tested the prognostic utility of each of the 21 genes in combination with all the others. As cut-off points, we used again quartiles (one quartile versus the other three quartiles) to categorize tumors as high versus low for the respective factors. When patients had both high *CXCL9* and *ITSN2* levels in their tumors, they had a longer disease-free survival time than those exhibiting low levels (log rank = 7.19, $P = 0.0073$, Fig. 3c). When the patients were stratified by *CXCL9* levels combined with either low or high *FLJ22028* levels, they showed significantly different disease-free survival curves: patients whose tumors exhibited low *CXCL9* gene expression levels were found to

Table 3 Correlation of Affymetrix expression data with real-time RT-PCR generated expression values

	Pearson correlation		Spearman rank	
	r	*P*	*r*	*P*
A4GALT	0.339	0.156	0.363	0.126
ANGPTL2	0.16	0.514	0.389	0.099
APOM	0.471[a]	0.042	0.465[a]	0.045
BGN	0.645[b]	0.003	0.574[a]	0.01
CXCL9	0.683[b]	0.001	0.795[b]	0.001
FLJ20257	0.611[b]	0.005	0.542[a]	0.016
FLJ22028	0.77[b]	≤0.0001	0.754[b]	≤0.0001
GNAI2	0.115	0.638	−0.137	0.576
GPR126	0.89[b]	≤0.0001	0.805[b]	≤0.0001
H2AFX	0.762[b]	≤0.0001	0.604[b]	≤0.006
INDO	0.746[b]	≤0.0001	0.839[b]	≤0.0001
ITSN2	−0.085	0.729	0.13	0.596
MGC10986	0.711[b]	0.001	0.609[b]	0.006
MGC16824	0.587[b]	0.008	0.46[a]	0.048
PCDH16	0.774[b]	≤0.0001	0.746[b]	≤0.0001
PDE4A	0.549[a]	0.015	0.512[a]	0.025
PEX1	0.597[b]	0.007	0.242	0.318
PGDS	0.574[a]	0.01	0.6[b]	0.007
TEM8	0.343	0.151	0.551[a]	0.015
TIMM44	0.591[b]	0.008	0.17	0.486
TOX	0.839[b]	≤0.0001	0.714[b]	≤0.001

Correlations are significantly positive for 16 of 21 genes by the Pearson correlation and significantly positive for 15 of 21 genes by Spearman rank correlation

[a] Correlation is significant at the 0.05 level (2-tailed)

[b] Correlation is significant at the 0.01 level (2-tailed)

Table 4 Probablity of survival after 120 months (%)

	High expression level[a] ($N = 17$)	Low expression levels[b] ($N = 53$)	*P*-value (log rank)
A4GALT	56.3	51.3	0.789
ANGPTL2	60.3	50.5	0.363
APOM	45.8	55.2	0.961
BGN	44.5	54.8	0.828
CXCL9	80.9	44.6	0.029
FLJ20257	59.6	50.4	0.346
FLJ22028	43.7	55.7	0.368
GNAI2	80.7	45.0	0.032
GPR126	62.8	49.6	0.484
H2AFX	64.7	49.0	0.098
INDO	74.5	46.3	0.088
ITSN2	79.6	44.4	0.018
MGC10986	74.1	46.8	0.107
MGC16824	58.6	52.8	0.638
PCDH16	47.0	54.1	0.734
PDE4A	62.7	49.7	0.243
PEX1	52.4	53.1	0.659
PGDS	52.3	53.5	0.881
TEM8	67.9	48.1	0.180
TIMM44	50.3	52.9	0.878
TOX	32.9	58.7	0.450

[a, b] mRNA levels of the 21 genes measured by QRT-PCR were categorized according to quartile distribution for gene expression values and either the lower or the upper quartiles were tested against the rest of the quartiles. Expression levels were then analyzed in relation to the length of disease-free survival (DFS) of the 70 patients

Expression levels of three genes (*CXCL9*, *ITSN2* and *GNAI2*) were significantly correlated with a favorable clinical course in terms of length of DFS

be significantly associated with a shortened disease-free survival when compared to the patients with high *CXCL9* tumor levels. When, in addition to low *CXCL9* levels, patients had high levels of *FLJ22028*, they had a worse clinical course than those with low *FLJ22028* levels (log-rank = 15.3, $P = 0.002$, Fig. 3d). Finally, a multivariate analysis of the risk of developing disease recurrence shows that the combined *CXCL9*/*ITSN2* or CXCL9/*FLJ22028* ratios were the only independent predictive factor of DFS ($P = 0.034$ and $P = 0.003$, respectively) ($n = 70$ tumors), when compared with traditional histological or clinical factors such as tumor size, nodal status, grading or hormone receptor status. A significant correlation between the level of mRNA expression and the length of overall survival was not found for any of the investigated genes (data not shown).

Discussion

The aim of this exploratory study was to identify genes that are useful in prediction of relapse of breast cancer patients that were uniformly treated with surgery and CMF, a standard combination chemotherapy. Using microarray and quantitative RT-PCR technologies, we identified six genes (*CXCL9*, *ITSN2*, *GNAI2*, *H2AFX*, *INDO*, and *MGC10986*) that were significantly differentially expressed in the recurrence versus the non-recurrence group of 70 breast cancer patients. We have found that high expression levels of *CXCL9*, *ITSN2*, and *GNAI2* were associated with prolonged disease-free survival (DFS) ($P = 0.029$, 0.018, and 0.032, respectively) and that patients stratified by combined *CXCL9*/*ITSN2* or *CXCL9*/*FLJ22028* tumor levels exhibited significantly different disease-free survival curves ($P = 0.0073$ and $P = 0.005$, respectively). Finally, the *CXCL9*/*ITSN2* and *CXCL9*/*FLJ22028* ratio was an independent prognostic factor ($P = 0.034$ and $P = 0.003$, respectively) for DFS by multivariate Cox analysis in the 70-patient cohort.

Our study group was carefully selected: patients had received no additional endocrine treatment, reflecting the

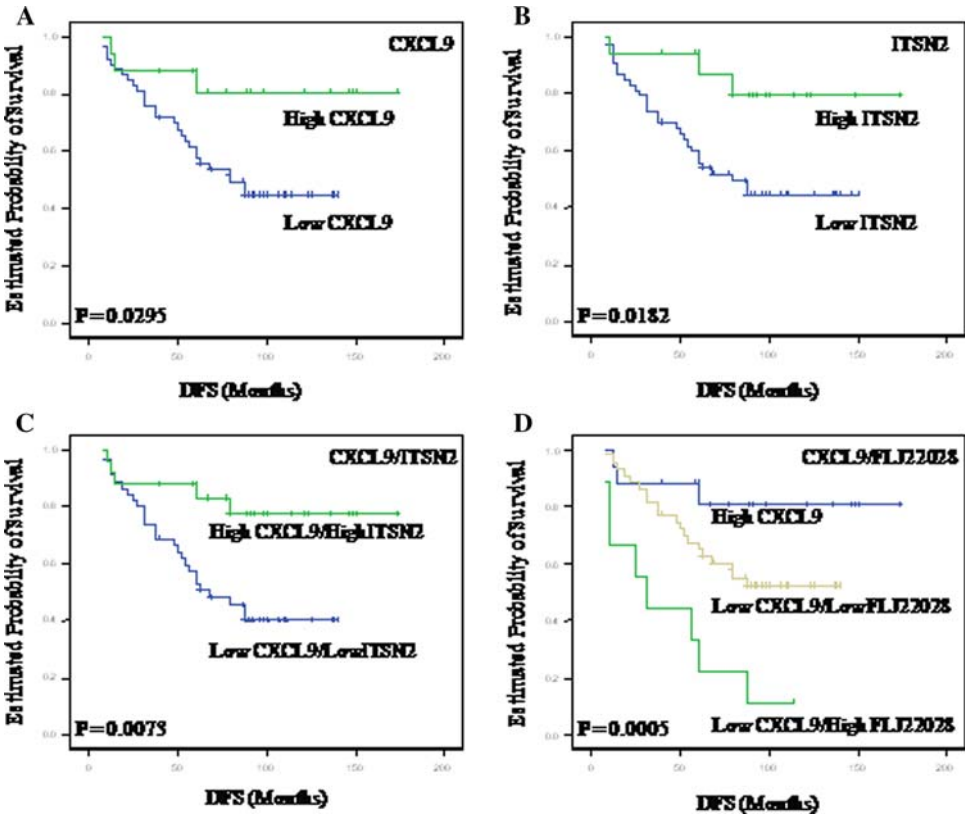

Fig. 3 DFS according to gene expression values categorized into quartiles in CMF-treated breast cancer patients. **a** *CXCL9* (log-rank = 4.74, $P = 0.0295$) in 70 patients. **b** *ITSN2* (log-rank = 4.74, $P = 0.0295$) in 70 patients. **c** *CXCL9/ITSN2* (log-rank = 7.19, $P = 0.0073$) in 70 patients. **d** *CXCL9/FLJ22028* (log-rank = 15.3 $P = 0.0005$) in 70 patients

therapeutic consensus recommendation of the time, so the assessment of relapse was not subject to potentially confounding contributions of additional systemic treatment. Moreover, a long follow-up (median: 96.5 months) of the patients had been obtained. Tumor RNA was of high quality and gained by manual macrodissection to ensure at least a 70% pure tumor cell population.

We started our exploratory study by generating gene expression profiles from patient specimens using Affymetrix HG-U133A microarrays. Among various published multivariate statistical methods and clustering algorithms to analyze microarray expression data [19, 20], we decided to use approaches similar to successful studies on tumors from breast cancer patients described previously [12–15]. Gene expression was analyzed using standard statistical methods (Mann–Whitney-Test and fold change) and, in addition, data were analyzed by clustering and iterative processes to identify gene sets predictive of relapse. A training set consisting of 19 specimens was used to obtain two gene sets of 12 and 21 genes (together 25 distinct genes) that correctly clustered 18 of 19 cases into recurrent or non-recurrent categories. Importantly, reproducible detection of the differential expression levels for discriminating genes was demonstrated by an independent method, quantitative RT-PCR. Quantitative RT-PCR validated gene expression changes for 21 of these 25 genes, 16 of which were significantly correlated with the microarray-derived values, and thus confirmed the predictive power of the gene set in the training set cases. A number of recent studies have shown that gene expression patterns can be used for prognostication and chemotherapy response prediction of breast cancer [14, 15, 18, 21, 22]. Van't Veer et al. [14] reported a 70-gene classifier that outperformed standard clinical or histological criteria in disease outcome prediction of 78 node negative breast cancer patients. The same investigators validated this 70-gene classifier in a cohort of 217 node-positive and node-negative patients. However, treatment of the patients group was not homogenous: either no additional treatment was given after surgery or patients were treated nonuniformly with hormonal or chemotherapeutic agents. Using this published gene classifier in our uniformly CMF-treated patient cohort for prediction of relapse, we did not observe a significant correlation with clinical outcome, however, only 36 probe sets covering 36 genes of the 70 genes were present on the oligonucleotide arrays used in our study (data not shown).While it would be optimal to study as many samples as possible for the initial marker discovery, several studies have shown that successful class prediction is possible using a limited number of samples [16–18]. Due to the limited number of tumor samples with appropriate characteristics, uniform treatment and clinical follow-up available, we started our exploratory microarray analysis with 19 samples and used the 51 samples of our study as validation set. The two gene sets of

12 and 21 genes (together 25 distinct genes) that we identified, correctly clustered 18 of 19 cases into recurrent or non-recurrent categories. However, when 21 of the 25 genes where applied to an independent test group consisting of 51 cases to test their accuracy to predict clinical outcome, the gene set failed to correctly classify the categories. There are a number of possible explanations for this: we used QRT-PCR technique to measure transcript levels and while we demonstrated that the RT-PCR-derived values significantly correlated with the microarray-derived values in the test set ($n = 19$ cases) in 16 of 21 genes, the different techniques employed for quantification of gene expression still might account for the failure of accurate outcome prediction in the validation set. Secondly, primer and probe set for QRT-PCR were only available for 21 of the 25 genes and it could be conceived that the missing genes would have given additional information. Our data stress the need to study large patient cohorts with microarrays when aiming at the identification of biomarkers for clinical use in solid tumors, representing a genetically far more heterogeneous group than for example hematological malignancies.

Among the 21 genes that were tested with QRT-PCR, six genes, *ITSN2*, *CXCL9*, *GNAI2*, *H2AFX*, *INDO*, and *MGC10986*, proved to be of significance when relapsed versus non-relapsed patients of an independent patient cohort ($n = 51$ patients) uniformly treated with CMF were tested. Interestingly, *CXCL9* ranked among the top predictor genes using the different algorithms. Furthermore, *ITSN2*, *CXCL9*, and *GNAI2* and combined *CXCL9*/*ITSN2* were significantly associated with a favorable clinical course ($P = 0.018$, 0.029, 0.032, and 0.0073, respectively) as measured by the time to disease recurrence. For some of these genes, an association with breast cancer had already been demonstrated. *CXCL9* is a chemotactic factor for activated T cells and NK cells and has been reported to be a suppressor of angiogenesis [23]. Furthermore, *CXCL9*, as a single agent, has been shown to exhibit tumor-inhibitory-activity in a model of breast cancer [24] and high *CXCL9* expression has been recently linked to a favorable clinical outcome of renal cell carcinoma [25]. Interestingly, in our patient cohort, high *CXCL9* expression is also associated with a prolonged time to disease progression. Whether CXCL9 is secreted by the tumor cells or is derived from tumor infiltrating lymphocytes is currently not known-although we manually macrodissected the tumor tissue we cannot rule out the possibility that some "contaminating" inflammatory cells might have contributed to high intratumoral CXCL9 levels. Gi alpha 2 (GNAI2) belongs to the family of Gi alpha proteins that includes 3 polypeptides: Gi alpha 1 (GNAI1), Gi alpha 2 (GNAI2), and Gi alpha 3 (GNAI3). It was recently reported that the alpha-subunit of G(i) can activate the Ras-ERK/MAPK mitogenic pathway by membrane recruitment of rap1GAPII and reduction of GTP-bound Rap1 [26]. Little is known about the relevance of intersectin 2 *(ITSN2)* in breast cancer. Intersectin 2 encodes an adapter protein which is part of the endocytotic machinery of the cell [27]. It binds N-WASp, a potent activator of actin assembly via the Arp2/3 complex and thus coordinates actin assembly and trafficking events [28]. When patients of our study were stratified by *CXCL9* levels and then combined with either low or high *FLJ22028* levels, they showed significantly different disease-free survival curves when considering the 70 patients ($P = 0.0005$) (Fig. 3d). Patients with high *CXCL9* levels in their tumors were found to be significantly associated with a prolonged disease-free survival; conversely, patients with low *CXCL9* tumor levels combined with high *FLJ22028* levels had the worst prognosis. A multivariate analysis of the risk of developing disease recurrence shows that the combined *CXCL9*/*FLJ22028* ratios was the only independent predictive factor of DFS ($P = 0.003$, respectively) ($n = 70$ tumors), when compared with traditional histological or clinical factors such as tumor size, nodal status, grading or hormone receptor status. *FLJ22028* codes for a hypothetical protein with predicted pyridine nucleotide-disulphide oxidoreductase activity. No association with breast cancer or chemotherapy response prediction has been associated so far with this gene.

There are only a few studies describing gene expression profiling of CMF-treated breast cancer patients. Nimeus et al. [29] reported on a comparison of gene expression profiling and conventional clinical markers to predict distant recurrences for premenopausal breast cancer patients after adjuvant CMF chemotherapy. Interestingly, the authors concluded that classifiers using cDNA microarray-based gene lists did not outperform corresponding classifiers based on clinical variables. Paik et al. [30] used a commercially available 21-gene recurrence score assay which includes genes involved in tumor cell proliferation and hormonal response to quantify the likelihood of breast cancer recurrence in women with node-negative, estrogen receptor-positive breast cancer and to predict the magnitude of CMF or methotrexate and fluorouracil chemotherapy benefit. CMF was the chemotherapeutic agent of choice at the time our patient cohort was treated for advanced breast cancer (between 1989 and 1996). Later on, anthracyclines such as doxorubicin and its analogue epirubicin have been gradually included in chemotherapeutic regimens [3]. Today, 5-fluorouracil and cyclophosphamide, two of the CMF components in the treatment of breast cancer are commonly used in combination with anthracyclines (FEC/FAC) [31, 32]. It would be therefore interesting to test whether the markers identified in the present study can predict the clinical outcome of patients who received FAC or FEC treatment. Moreover, as CMF is currently also being used in

a neoadjuvant setting before surgery, it would be interesting to see whether the two gene ratios have the potential to predict chemotherapy response in preoperatively treated patients. To our knowledge, gene expression analysis of tumors from patients undergoing primary neoadjuvant CMF chemotherapy has not been performed so far, however, there are several studies reporting on gene expression profiling for response prediction of breast cancer patients treated with neoadjuvant taxanes and anthracyclines. Chang et al. [33] identified a 92-gene set that predicted response to docetaxel in primary breast tumors of 24 patients, with positive and negative predictive values of 92 and 83%, respectively. Another study [34] analyzed gene expression in diagnostic core biopsy tissue samples from breast cancer patients treated with gemcitabine, epirubicine and docetaxel. A gene expression signature consisting of 512 genes was identified, that predicted pathologic tumor response with a sensitivity of 78%, a specificity of 90%, and an overall accuracy of 88%. Rody et al. [35, 36] performed gene expression analysis on pretherapeutic core biopsies from patients treated with docetaxel, adriamycin and cyclophosphamide neoadjuvant combination chemotherapy. Based on the Sorlie classification, they demonstrated that the erbB2-positive cluster of the intrinsic gene set predicted tumor response. Generally, while it is impossible with this kind of investigations to foresee whether the outcome associations are related to the actual sensitivity of the tumors to the agents or to the biology of the tumor itself, they might provide markers for the personalized choice of alternative treatment regimens.

In summary, our findings demonstrate the utility of a small set of biomarkers in identifying breast cancer patients still at high risk of relapse after CMF chemotherapy treatment. Whether the biomarker profile is regimen-specific or a more general indicator of chemotherapy resistance or whether it reflects the aggressiveness of the tumor merits further research. Moreover, a more thorough investigation of one of the key markers found in our study, CXCL9, may reveal important insights into the understanding of basic mechanisms underlying chemotherapy- and immune response.

Acknowledgments Supported by a grant to M. Kiechle and H. Hoefler from the BMBF (Federal Ministry of Education and Research), Germany, German National Genome Project (KR-S15T03).

References

1. Eifel P, Axelson JA, Costa J, Crowley J, Curran WJ Jr, Deshler A et al (2001) National Institutes of Health Consensus Development Conference Statement: adjuvant therapy for breast cancer, November 1–3, 2000. J Natl Cancer Inst 93:979–989. doi:10.1093/jnci/93.13.979
2. Early Breast Cancer Trialists' Collaborative Group (1992) Systemic treatment of early breast cancer by hormonal, cytotoxic, or immune therapy. 133 randomised trials involving 31,000 recurrences and 24,000 deaths among 75,000 women. Lancet 339:1–15
3. Early Breast Cancer Trialists' Collaborative Group (1998) Polychemotherapy for early breast cancer: an overview of the randomised trials. Lancet 352:930–942. doi:10.1016/S0140-6736(98)03301-7
4. Bonadonna G, Valagussa P, Moliterni A, Zambetti M, Brambilla C (1995) Adjuvant cyclophosphamide, methotrexate, and fluorouracil in node-positive breast cancer: the results of 20 years of follow-up. N Engl J Med 332:901–906. doi:10.1056/NEJM199504063321401
5. Levine MN, Bramwell VH, Pritchard KI, Norris BD, Shepherd LE, Abu-Zahra H et al (1998) Randomized trial of intensive cyclophosphamide, epirubicin, and fluorouracil chemotherapy compared with cyclophosphamide, methotrexate, and fluorouracil in premenopausal women with node-positive breast cancer. National Cancer Institute of Canada Clinical Trials Group. J Clin Oncol 16:2651–2658
6. Goldhirsch A, Glick JH, Gelber RD, Coates AS, Thurlimann B, Senn HJ (2005) Meeting highlights: international expert consensus on the primary therapy of early breast cancer 2005. Ann Oncol 16:1569–1583. doi:10.1093/annonc/mdi326
7. Pritchard KI, Shepherd LE, O'Malley FP, Andrulis IL, Tu D, Bramwell VH et al (2006) HER2 and responsiveness of breast cancer to adjuvant chemotherapy. N Engl J Med 354:2103–2111. doi:10.1056/NEJMoa054504
8. Askmalm MS, Carstensen J, Nordenskjold B, Olsson B, Rutqvist LE, Skoog L et al (2004) Mutation and accumulation of p53 related to results of adjuvant therapy of postmenopausal breast cancer patients. Acta Oncol 43:235–244. doi:10.1080/02841860410029474
9. Andersson J, Larsson L, Klaar S, Holmberg L, Nilsson J, Inganas M et al (2005) Worse survival for TP53 (p53)-mutated breast cancer patients receiving adjuvant CMF. Ann Oncol 16:743–748. doi:10.1093/annonc/mdi150
10. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M et al (2004) A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 351:2817–2826. doi:10.1056/NEJMoa041588
11. Elston CW, Ellis IO (1991) Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. Histopathology 19:403–410. doi:10.1111/j.1365-2559.1991.tb00229.x
12. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP et al (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 286:531–537. doi:10.1126/science.286.5439.531
13. Tamayo P, Slonim D, Mesirov J, Zhu Q, Kitareewan S, Dmitrovsky E et al (1999) Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. Proc Natl Acad Sci USA 96:2907–2912. doi:10.1073/pnas.96.6.2907
14. van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M et al (2002) Gene expression profiling predicts clinical outcome of breast cancer. Nature 415:530–536. doi:10.1038/415530a
15. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW et al (2002) A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 347:1999–2009. doi:10.1056/NEJMoa021967
16. Hwang D, Schmitt WA, Stephanopoulos G (2002) Determination of minimum sample size and discriminatory expression patterns in microarray data. Bioinformatics 18:1184–119317

17. Mukherjee S, Tamayo P, Rogers S, Rifkin R, Engle A, Campbell C et al (2003) Estimating dataset size requirements for classifying DNA microarray data. J Comput Biol 10:119–142
18. Ayers M, Symmans WF, Stec J, Damokosh AI, Clark E, Hess K et al (2004) Gene expression profiles predict complete pathologic response to neoadjuvant paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide chemotherapy in breast cancer. J Clin Oncol 22:2284–2293. doi:10.1200/JCO.2004.05.166
19. Datta S, Datta S (2003) Comparisons and validation of statistical clustering techniques for microarray gene expression data. Bioinformatics 19:459–466. doi:10.1093/bioinformatics/btg025
20. Hilsenbeck SG, Friedrichs WE, Schiff R, O'Connell P, Hansen RK, Osborne CK et al (1999) Statistical analysis of array expression data as applied to the problem of tamoxifen resistance. J Natl Cancer Inst 91:453–459. doi:10.1093/jnci/91.5.453
21. Huang E, Cheng SH, Dressman H, Pittman J, Tsou MH, Horng CF et al (2003) Gene expression predictors of breast cancer outcomes. Lancet 361:1590–1596. doi:10.1016/S0140-6736(03)13308-9
22. Sotiriou C, Neo SY, McShane LM, Korn EL, Long PM, Jazaeri A et al (2003) Breast cancer classification and prognosis based on gene expression profiles from a population-based study. Proc Natl Acad Sci USA 100:10393–10398. doi:10.1073/pnas.1732912100
23. Sgadari C, Farber JM, Angiolillo AL, Liao F, Teruya-Feldstein J, Burd PR et al (1997) Mig, the monokine induced by interferon-gamma, promotes tumor necrosis in vivo. Blood 89:2635–2643
24. Dorsey R, Kundu N, Yang Q, Tannenbaum CS, Sun H, Hamilton TA et al (2002) Immunotherapy with interleukin-10 depends on the CXC chemokines inducible protein-10 and monokine induced by IFN-gamma. Cancer Res 62:2606–2610
25. Kondo T, Ito F, Nakazawa H, Horita S, Osaka Y, Toma H (2004) High expression of chemokine gene as a favorable prognostic factor in renal cell carcinoma. J Urol 171:2171–2175. doi:10.1097/01.ju.0000127726.25609.87
26. Mochizuki N, Ohba Y, Kiyokawa E, Kurata T, Murakami T, Ozaki T et al (1999) Activation of the ERK/MAPK pathway by an isoform of rap1GAP associated with G alpha(i). Nature 400:891–894. doi:10.1038/23738
27. Pucharcos C, Estivill X, de la Luna S (2000) Intersectin 2, a new multimodular protein involved in clathrin-mediated endocytosis. FEBS Lett 478:43–51. doi:10.1016/S0014-5793(00)01793-2
28. Schafer DA (2002) Coupling actin dynamics and membrane dynamics during endocytosis. Curr Opin Cell Biol 14:76–81. doi:10.1016/S0955-0674(01)00297-6
29. Niméus-Malmström E, Ritz C, Edén P, Johnsson A, Ohlsson M, Strand C et al (2006) Gene expression profiles and conventional clinical markers to predict distant recurrences for premenopausal breast cancer patients after adjuvant chemotherapy. Eur J Cancer 42:2729–2737. doi:10.1016/j.ejca.2006.06.031
30. Paik S, Tang G, Shak S, Kim C, Baker J, Kim W et al (2006) Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. J Clin Oncol 24:3726–3734. doi:10.1200/JCO.2005.04.7985
31. Martin M, Villar A, Sole-Calvo A, Gonzalez R, Massuti B, Lizon J et al (2003) Doxorubicin in combination with fluorouracil and cyclophosphamide (i.v. FAC regimen, day 1, 21) versus methotrexate in combination with fluorouracil and cyclophosphamide (i.v. CMF regimen, day 1, 21) as adjuvant chemotherapy for operable breast cancer: a study by the GEICAM group. Ann Oncol 14:833–842. doi:10.1093/annonc/mdg260
32. Bonadonna G, Zambetti M, Moliterni A, Gianni L, Valagussa P (2004) Clinical relevance of different sequencing of doxorubicin and cyclophosphamide, methotrexate, and fluorouracil in operable breast cancer. J Clin Oncol 22:1614–1620. doi:10.1200/JCO.2004.07.190
33. Chang JC, Wooten EC, Tsimelzon A, Hilsenbeck SG, Gutierrez MC, Elledge R et al (2003) Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer. Lancet 362:362–369. doi:10.1016/S0140-6736(03)14023-8
34. Thuerigen O, Schneeweiss A, Toedt G, Warnat P, Hahn M, Kramer H et al (2006) Gene expression signature predicting pathologic complete response with gemcitabine, epirubicin, and docetaxel in primary breast cancer. J Clin Oncol 24:1839–1845. doi:10.1200/JCO.2005.04.7019
35. Rody A, Karn T, Gätje R, Kourtis K, Minckwitz G, Loibl S et al (2006) Gene expression profiles of breast cancer obtained from core cut biopsies before neoadjuvant docetaxel, adriamycin, and cyclophoshamide chemotherapy correlate with routine prognostic markers and could be used to identify predictive signatures. Zentralbl Gynakol 128:76–81. doi:10.1055/s-2006-921508
36. Rody A, Karn T, Solbach C, Gaetje R, Munnes M, Kissler S et al (2007) The erbB2+ cluster of the intrinsic gene set predicts tumor response of breast cancer patients receiving neoadjuvant chemotherapy with docetaxel, doxorubicin and cyclophosphamide within the GEPARTRIO trial. Breast 16:235–240. doi:10.1016/j.breast.2007.02.006