Basal-like molecular subtype and HER4 up-regulation and response to neoadjuvant chemotherapy in breast cancer

ELMAR STICKELER[1], DIETMAR PILS[2], MAXIMILIAN KLAR[1], MARZENNA ORLOWSK-VOLK[3], AXEL ZUR HAUSEN[3], MARKUS JÄGER[1], DIRK WATERMANN[1], GERALD GITSCH[1], ROBERT ZEILLINGER[2] and CLEMENS B. TEMPFER[4]

[1]Department of Obstetrics and Gynecology, University of Freiburg, Freiburg, Germany; [2]Department of Obstetrics and Gynecology, Medical University of Vienna, Vienna, Austria; [3]Department of Pathology, University of Freiburg, Freiburg, Germany; [4]Department of Obstetrics and Gynecology, Marienhospital Herne-Ruhr University Bochum, Herne, Germany

Received May 26, 2011; Accepted June 25, 2011

DOI: 10.3892/or.2011.1392

Abstract. Alteration of gene expression profiles during chemotherapy may predict response to neoadjuvant chemotherapy (NAC) in breast cancer patients. In a prospective cohort study of 32 women with primary invasive breast cancer, we obtained tumor specimens before and after 4 cycles of NAC with epirubicine 90 mg/m^2 and cyclophosphamide 600 mg/m^2, followed by 4 cycles of docetaxel 100 mg/m^2. Total-RNA was extracted from tumor specimens and the whole transcriptome was analyzed with Agilent's 44K single color microarray. Data analysis was performed by GeneSpring v.11 and IBM SPSS v.18. Ten tumors were classified as basal-like and 22 tumors were classified as non-basal-like. Gene expression-based molecular subtype (basal-like vs. non-basal-like) (P=0.003), but not tumor grade (P=0.07), estrogen receptor (P=0.1), progesterone receptor (P=0.6) and HER2 status (P=0.4) predicted pathological complete response to NAC. Specifically, 7/10 basal-like tumors responded to NAC, whereas 19/22 non-basal-like tumors did not respond. Comparing gene expression signatures before and after 4 cycles of NAC, we found that all patients with an initial non-basal-like tumor retained this tumor type, whereas 5/7 basal-like tumors, including all responders, lost this molecular subtype. Complete prediction of response to NAC was achieved with a 21 gene list (P=0.000008). Of note, both the expression and up-regulation of a single gene, i.e. HER4, predicted the response to NAC in 26/32 (81%; P=0.002) and in 23/25 (92%; P<0.001) patients, respectively. These preliminary data indicate that therapy-induced HER4 gene up-regulation may be associated with response to NAC with epirubicine, cyclophosphamide and docetaxel.

Correspondence to: Dr Clemens Tempfer, Department of Obstetrics and Gynecology, Marienhospital Herne-Ruhr University Bochum, Hölkeskampring 40, D-44625 Herne, Germany
E-mail: clemens.tempfer@marienhospital-herne.de

Key words: gene expression, HER4, breast cancer, microarray, chemotherapy

Introduction

Chemotherapy, in the form of preoperative neo-adjuvant chemotherapy (NAC) or postoperative adjuvant chemotherapy, constitutes the standard of care for women with invasive breast cancer. The most effective combinations of cytotoxic drugs include anthracyclines, such as epirubicine and doxorubicine, and taxanes, such as docetaxel and paclitaxel. A well established and widely used regimen uses 4 cycles of epirubicine 90 mg/m^2 and cyclophosphamide 600 mg/m^2 every 3 weeks, followed by 4 cycles of docetaxel 100 mg/m^2. Depending on different probabilities of relapse, however, between 70 and 98% of patients undergoing chemotherapy will not benefit from this therapeutic intervention (1,2). Chemotherapy is applied empirically, since there are no markers sensitive and specific enough to predict the response and thus assign or withhold chemotherapy regimens to or from individual patients. Clinical parameters such as tumor size, regional lymph node status, tumor cell differentiation and expression markers on the protein level such as p53, bcl-2, and Ki-67, show no strong association with response to chemotherapy (3).

Response to NAC can be objectively measured by shrinkage of visible breast lesions as well as histological evaluation of the surgery specimen with respect to pathologic complete remission. Both items have demonstrated strong associations with disease-free and overall survival and are thus suitable surrogate markers (4). There is a need for sensitive and specific markers of response to chemotherapy to reduce the significant morbidity and costs associated with breast cancer treatment. Microarray technology provides the possibility to simultaneously assess thousands of genes by high volume quantification of gene expression. From a clinical perspective, this technology may be used before NAC in order to predict the response to chemotherapy in women with primary breast cancer.

Previous studies demonstrated the potential of gene expression profiles as predictive as well as prognostic markers of primary breast cancer (5). Various groups have investigated a variety of genes and chemotherapy regimens. There is, however, no agreement as to what set of genes should be recommended

for predicting the response to chemotherapy in general and the response to specific chemotherapy regimens in particular. For example, Chang *et al* studied core biopsies of 24 patients before NAC with 4 cycles of docetaxel 100 mg/m^2 every 3 weeks (6). Using a HgU95-Av2 GeneChip, they identified a set of 92 genes involved in cell cycle control, protein transport, cell adhesion, protein modification, cellular stress, and apoptosis, differentiating between response to therapy and lack thereof. Response was measured by the degree of tumor shrinkage with an arbitrarily chosen cut-off of 25% or less residual disease. They reported a sensitivity of 85% (11/13 resistant tumors identified) and a specificity of 90% (10/11 sensitive tumors identified). In another study, Ayers *et al* (7) used a cDNA array containing 30,721 human sequence clones to assess the gene expression in breast cancer specimens of 42 women before NAC with sequential weekly paclitaxel and fluorouracil/doxorubicin/cyclophosphamide. A set of 74 genes was shown to best predict the response to NAC defined as pathological complete response (pCR) with a sensitivity of 43% (3/7 responders identified) and a specificity of 100% (11/11 non-responders identified). It is of note, that only some, but not the majority of genes, listed in the works of Chang *et al* (6) and Ayers *et al* (7) are identical. It is unknown, whether this difference is due to differences in gene array methodology, chemotherapy regimens, patient population characteristics, or a combination of these factors.

Sotiriou *et al* (8) described the differential expression of 37 genes based on a 7600 transcript cDNA array predicting clinical response to NAC with adriamycin 60 mg/m^2 and cyclophosphamide 600 mg/m^2. They also investigated the changes in gene expression after one cycle of chemotherapy. The number of genes that underwent expression changes was 10 times greater in the group of responders than in non-responders. Of note, the list of genes identified by Sotiriou *et al* (8) vastly differs from the list of genes identified by previous investigators (7,9).

Given the differences reported so far, it is a major challenge to identify molecular subtypes of breast cancer with specific response patterns to cytotoxic chemotherapy. Specific gene lists obtained by microarray expression profiling of invasive breast cancer specimens have led to the characterization of five molecular subtypes, ie luminal A, luminal B, basal-like, HER2 and normal breast-like (10). These molecular subtypes accurately predict prognosis, but have not been successfully used for prediction of the response to chemotherapy (11). One reason for the failure of incorporation of gene expression profiling and molecular subtypes into clinical practice is that individual specimens cannot be reliably assigned to the same molecular subtype. Specifically, molecular classification of individual samples is achieved by using single sample predictors (SSPs), based on statistical similarities between a given case and molecular subtype centroids (12-14). As recently demonstrated (11), assignment of individual cases to luminal A, luminal B, HER2, and normal breast-like was not reproducible and varied substantially depending on the SSP used. However, the proportion of cases classified as basal-like was consistent with all SSPs used. Whether or not these molecular subtypes respond differently to NAC is unknown.

We have designed a study using microarray technology to assess global gene expression at two time points during a common NAC regimen, epirubicine 90 mg/m^2 and cyclophosphamide 600 mg/m^2, followed by docetaxel 100 mg/m^2. We hypothesized that molecular tumor classification comparing basal-like and non basal-like subtypes as well as gene expression change during chemotherapy are associated with pathological complete remission after NAC.

Materials and methods

Patient selection and specimen sampling. We prospectively recruited 32 women with histologically diagnosed primary breast cancer and a sonographically measurable breast lesion undergoing NAC at the Department of Obstetrics and Gynecology, University of Freiburg, Freiburg, Germany. Patients were recruited over 20 months between March, 2008 and December, 2009. All women gave informed written consent before inclusion into the study. Approval by the Ethics Committee of the University of Freiburg was obtained.

Inclusion criteria were as follows: i) histological evidence of invasive breast cancer with a lesion measurable by breast ultrasound and magnetic resonance imaging (MRI) breast scan, ii) absence of contralateral breast cancer and distant metastases based on imaging studies, i.e. mammography, chest radiograph, liver sonography and bone scintigraphy and iii) absence of a family history of breast and/or ovarian cancer suggestive of hereditary breast cancer.

Exclusion criteria were as follows: i) personal history of exposure to cytotoxic chemotherapy, ii) blood counts, liver and renal function parameters below predefined standard cutoffs for cytotoxic chemotherapy, iii) severe co-morbidity, iv) age >70 years, v) concurrent use of cyclosporine or methrotrexate, vi) severe kidney disease with decreased renal function, vii) severe liver disease, viii) uncontrolled hypertension (>160/90 mmHg) and ix) asthma.

Before histological confirmation of a diagnosis of invasive breast cancer, all patients underwent an MRI scan of both breasts. Then, a two-pass, high speed core biopsy, evaluated the cancer cell content of >50% by fresh-frozen section analysis, was performed. After establishing the diagnosis of invasive breast cancer, all women underwent sequential NAC with 4 cycles of epirubicine 90 mg/m^2 and cyclophosphamide 600 mg/m^2 every 3 weeks, followed by 4 cycles of docetaxel 100 mg/m^2 every 3 weeks. At the time of diagnosis and at 4 time points during NAC and after completion of therapy, lesions were ultrasonographically measured to assess response. After completion of 4 cycles of epirubicine/cyclophosphamide, a second two-pass high speed core biopsy of the primary tumor was performed and tissue was stored for analysis. All patients underwent surgery after completion of chemotherapy. Patients without any residual invasive cancer in the breast and axillary lymph nodes were considered to have pCR. Patients with residual *in situ* cancer only were also considered to have pCR (15). Gene expression patterns were correlated to the response to NAC.

Microarray hybridization. Total-RNA was isolated using the RNA 6000 Nano LabChip kit (Agilent Technologies, Palo Alto, CA). Total-RNA (200 ng) was labelled and hybridized to the Agilent Whole Human Genome Microarray 4x44K (comprised of >41,000 unique human genes) according to the manufacturer's instructions. The following description, based on the guideline document 'Minimum Information About a Microarray Experiment-MIAME', developed by the

Table I. Clinicopathological characteristics of 32 breast cancer patients.

Patient	ypT	ypN	Grade	ER (%)	PR (%)	HER2neu	Initial tumor (cm)	Tumor post CHT (cm)	pCR
1	0	0	3	0	0	1	1.0	-	Y
2	0	0	3	0	0	1	2.5	-	Y
3	1b	0	2	50	10	1	1.7	0.9	N
4	0	0	3	0	0	1	3.3	-	Y
5	1	1a	3	80	60	1	4.5	1.4	N
6	2	0	3	90	25	0	3.0	2.5	N
7	1b	0	2	15	30	3	2.1	0.6	N
8	1c	0	2	60	40	0	3.0	1.7	N
9	1b	0	2	90	3	1	1.8	0.7	N
10	1a	0	3	90	95	1	4.0	0.5	N
11	1mic	0	3	0	0	3	2.5	0.1	Y
12	1c	1a	2	50	50	0	3.3	1.2	N
13	1c	0	2	75	50	2	2.2	1.1	Y
14	is	0	3	0	0	0	2.2	-	Y
15	1a	0	2	70	10	0	1.3	0.4	N
16	0	0	2	0	0	3	2.4	-	Y
17	1c	x	2	80	95	1	2.5	1.4	N
18	is	0	3	0	0	3	3.0	0.3	N
19	0	0	3	60	90	1	2.0	-	Y
20	1c	1a	2	60	10	0	1.7	1.7	N
21	is	0	3	40	20	3	1.9	0.6	N
22	1a	0	1	0	0	3	3.3	0.4	N
23	1c	3	3	50	90	1	3.0	1.3	N
24	1c	2a	3	0	0	3	2.5	1.8	N
25	0	0	3	0	0	0	2.5	-	Y
26	1c	0	3	0	0	0	2.2	1.6	N
27	1b	2a	2	10	80	1	4.0	0.7	N
28	2	1a	2	50	60	1	3.5	3.1	N
29	1b	0	2	0	0	3	2.3	0.6	N
30	2	2	2	0	0	0	6.1	3.8	N
31	1a	0	2	80	80	0	2.2	0.6	N
32	1a	1	3	0	0	1	4.7	0.3	Y

ypT, pathological tumor stage; ypN, pathological lymph node status; ER, estrogen receptor; PR, progesterone receptor; CHT, chemotherapy; pCR, pathological complete response; Y, yes; N, no.

Microarray Gene Expression Data society (http://www.mged.org/miame) (16), provides additional information on the microarray experiments not described in detail. The protocol and conditions used during hybridization, blocking, and washing strictly followed the standard protocols recommended by the manufacturer (Agilent).

Microarray and statistical analysis. Raw microarray data were quantile-normalized, $\log_2$-transformed, and A(bsent)-, M(arginal)-, and P(resent)-flags set according to the GeneSpring software (Agilent) default settings. For statistical analysis, genes were pre-filtered under the following conditions: Flag=P in 75% of samples in each group, either pCR or no pCR (27,942 genes) or before chemotherapy, 4 cycles of epirubicine and cyclophosphamide, and after chemotherapy (28,612 genes). For analysis of the difference of expression before and after chemotherapy, the $\log_2$ expression values before chemotherapy were subtracted from the expression values after chemotherapy, corresponding to the $\log_2$-fold changes. For statistical analysis, the Excel Add-in Significance Analysis of Microarrays (SAM) was used (17). pCR (pCR vs. no pCR) was used as a quantitative outcome and a false discovery rate (FDR) of 10% was set as a cut-off.

The centroid correlations to the molecular subtypes (basal-like, luminal A, luminal B, HER2, and normal breast-like) using 3 different SSPs were made following the instructions from the supplementary web appendix to Weigelt *et al* (11). Gene symbols were used as identifiers, because they always had the highest number of matches. The correlation of molecular subtype clas-

Table II. Classification of patients according to molecular subtype using three single sample predictors.

Patient no.	SSP matched symbols						Summary				
	Sorlie 398/500		Hu 289/306		Parker 49/50						
	P	S	P	S	P	S	B	LA	LB	H	Classification
1	B	B	B	B	B	B	6	0	0	0	Basal
2	B	B	B	B	B	B	6	0	0	0	Basal
3	LB	LB	H	B	B	B	3	0	2	1	Non basal
4	B	B	B	B	B	B	6	0	0	0	Basal
5	LB	LB	LB	LB	LB	LB	0	0	6	0	Non basal
6	LB	LB	LB	LB	LB	LB	0	0	6	0	Non basal
7	LB	LB	H	(H)	LA	LA	0	2	2	2	Non basal
8	(LB)	LB	LB	LB	LA	LB	0	1	5	0	Non basal
9	LA	LB	LB	LB	LB	LB	0	1	5	0	Non basal
10	(LA)	(LA)	(LA)	(LA)	LA	LA	0	6	0	0	Non basal
11	B	B	B	B	B	B	6	0	0	0	Basal
12	LA	LA	LA	LB	LA	LB	0	4	2	0	Non basal
13	LA	LA	LA	LB	LA	LA	0	5	1	0	Non basal
14	B	B	B	B	B	B	6	0	0	0	Basal
15	LA	LA	(LA)	(LB)	LB	LB	0	3	3	0	Non basal
16	H	LB	H	(B)	H	H	1	0	1	4	Non basal
17	LA	LB	LA	(LB)	LA	LA	0	4	2	0	Non basal
18	LB	LB	H	H	H	H	0	0	2	4	Non basal
19	LB	LB	(LB)	(LB)	LA	LA	0	2	4	0	Non basal
20	LA	LA	LB	LB	LA	LB	0	3	3	0	Non basal
21	LB	LB	H	H	H	H	0	0	2	4	Non basal
22	(LA)	(LA)	H	(LB)	LA	LA	0	4	1	1	Non basal
23	(LB)	LB	H	(LB)	LA	LA	0	2	3	1	Non basal
24	B	B	B	B	B	B	6	0	0	0	Basal
25	B	B	B	B	B	B	6	0	0	0	Basal
26	B	B	B	B	B	B	6	0	0	0	Basal
27	(LB)	LB	LA	(LB)	LA	LA	0	3	3	0	Non basal
28	LB	LB	LA	(LA)	LA	LA	0	4	2	0	Non basal
29	LB	LB	H	LB	H	LB	0	0	4	2	Non basal
30	B	B	B	B	B	B	6	0	0	0	Basal
31	LB	LB	LB	LB	LA	LB	0	1	5	0	Non basal
32	B	B	B	B	B	B	6	0	0	0	Basal

SSP, single sample predictor. Tumors were classified as B, basal-like; LA, luminal A; LB, luminal B; and H, HER2, based on the P, Pearson correlation and S, Spearman correlation analyses. Sybtypes in parentheses indicate those unclassified using a correlation cut-off <0.1. Free-marginal κ, 0.57 (chance-adjusted measure of agreement).

sifications derived form different SSPs was evaluated using the free-marginal multi-rater κ (18). Gene sets for the prediction of responder status (pCR vs. no pCR) were selected using the uncorrelated shrunken centroids (USC)-approach implemented in the MultiExperiment Viewer (MeV) using default parameters (19). Finally, linear regression models were built from gene-sets derived from the USC-approach and analyzed using receiver operating characteristics (ROC). The first 4 eigenvectors of a principal component analysis (PCA) normalizing all columns to zero mean and unit standard deviation were calculated and the first two of them are shown in a 2D graph. Gene lists were functionally annotated with a database for annotation, visualization, and integrated discovery (DAVID) Tools and were corrected for multiple testing according to Benjamini and Hochberg. A network analysis was performed using the Network Analysis, Visualization, and Graphing Toronto (NAViGaTOR) tool (20) and the Interologous Interaction Database v.1.72 (http://ophid.utoronto.ca/i2d). Random networks were built with eight

Table III. Clinicopathological characteristics, molecular subtype, and response to neoadjuvant chemotherapy comparing responders (pCR) to non-responders (no pCR).

	Total	No pCR	pCR	P-value
Total	32	22	10	
Initial tumor size, mean ± SD (range)	2.76±1.06 (1.0-6.1)	2.86±1.11 (1.3-6.1)	2.53±0.95 (1.0-4.7)	0.424[a]
Grade, n				
1	1	1	0	0.070[b]
2	15	13	2	
3	16	8	8	
ER, n				
0.0-33.3%	16	8	8	0.130[b]
33.3-66.6%	8	7	1	
66.6-100.0%	8	7	1	
PR, n				
0.0-33.3%	21	13	8	0.630[b]
33.3-66.6%	5	4	1	
66.6-100.0%	6	5	1	
HER2neu, n				
0	10	8	2	0.374[b]
1	13	8	5	
2	1	0	1	
3	8	6	2	
Subtype, n				
Basal	10	3	7	0.003[b]
Non-Basal	22	19	3	

[a]t-test. [b]Fisher's exact test. ER, estrogen receptor; PR, progesterone receptor.

seeding proteins from the whole annotated proteome based on Agilents' Whole Human Genome Microarray 4x44K. For statistical analysis, a one-sample t-test was used. Statistically significant differences in clinicopathological parameters and molecular subtypes were calculated using t-test and Fisher's exact test as appropriate. Predictive capabilities were assessed using the area under the ROC curves (AUC). The 95%-confidence intervals and P-values are shown where appropriate. P-values below 0.05 were considered statistically significant. All statistical analyses were performed with R (v.2.10.0) (21), BioConductor (v.2.5) (22), GeneSpring GX (v.11.0.1), DAVID Tools (v.6.7) (23) and IBM SPSS v.18 (IBM, Chicago, IL).

Results

Classification of molecular subtypes and prediction of responder-status. In this prospective series, 32 patients completed sequential NAC with 4 cycles of epirubicine 90 mg/m^2 and cyclophosphamide 600 mg/m^2 every 3 weeks, followed by 4 cycles of docetaxel 100 mg/m^2 every 3 weeks. No patient dropped out of the study. Median age of the patients was 50.0 (range, 30-68) years. Tumor characteristics of the investigated patients are shown in Table I. Using 3 SSPs, which are based on correlations between single samples and molecular subtype centroids (12-14), tumors were classified as luminal A, luminal B, basal-like, HER2, and normal breast-like (Table II). Classification agreement was low, especially for the non basal-like subtypes, i.e. the free-marginal κ, a chance-adjusted measure of agreement, was 0.57. Therefore, and following the classification suggested by Weigelt *et al* (11), tumors were further characterized as basal-like (n=10) and non basal-like (n=22) (Table II).

Comparing molecular subtypes and response to NAC [pCR (responders) vs. no pCR (non-responders)], we found that molecular subtype (basal-like vs. non basal-like) (P=0.003), but not tumor grade (P=0.070), estrogen receptor (P=0.130), progesterone receptor (P=0.630) and HER2 status (P=0.374) were significantly different (Table III). Specifically, 7/10 tumors with a basal-like molecular subtype responded to NAC, whereas 19/22 tumors with non basal-like gene molecular subtype did not respond. The AUC for molecular subtype as a predictor of response to NAC was 0.782 (P=0.012) (Table IV). Comparing gene expression signatures before and after 4 cycles of epirubicine and cyclophosphamide in 25 patients (in the remaining seven patients the percentage of tumor cells in the biopsy was too low for analysis), we found that all patients with an initial non basal-like molecular subtype retained this subtype, i.e. non basal-like. Of the 5 responder patients with an initial basal-like molecular subtype, all lost the subtype and turned to a non basal-like molecular subtype.

Table IV. Area under the curve (AUC) from receiver operating characteristics (ROC) for parameters predicting responder (pCR vs. no pCR)-status.

Parameter	AUC	CI_{95}	P-value
Initial tumor size	0.432	0.221-0.643	0.542
Grade (pathology)	0.723	0.536-0.910	0.046
GGI score	0.718	0.509-0.927	0.051
GGI-grade 3 vs. grade 1	0.559	0.346-0.772	0.597
GGI-grade change[a]	0.728	0.510-0.946	0.071
ER (%)	0.755[b]	0.572-0.937	0.023
PR (%)	0.736[b]	0.538-0.935	0.035
HER2neu	0.555	0.347-0.762	0.626
Molecular subtype	0.782	0.592-0.972	0.012
Initial HER4 expression	0.855[b]	0.725-0.984	0.002
Change of HER4 expression[a]	0.941	na[c]	<0.001

[a]Change of GGI-grade or HER4 expression before and after NAC. [b]Variables were inverted to get AUC values above 0.5. [c]Not applicable (asymptotic 95% confidence interval (CI_{95}) not calculated); ER, estrogen receptor; PR, progesterone receptor.

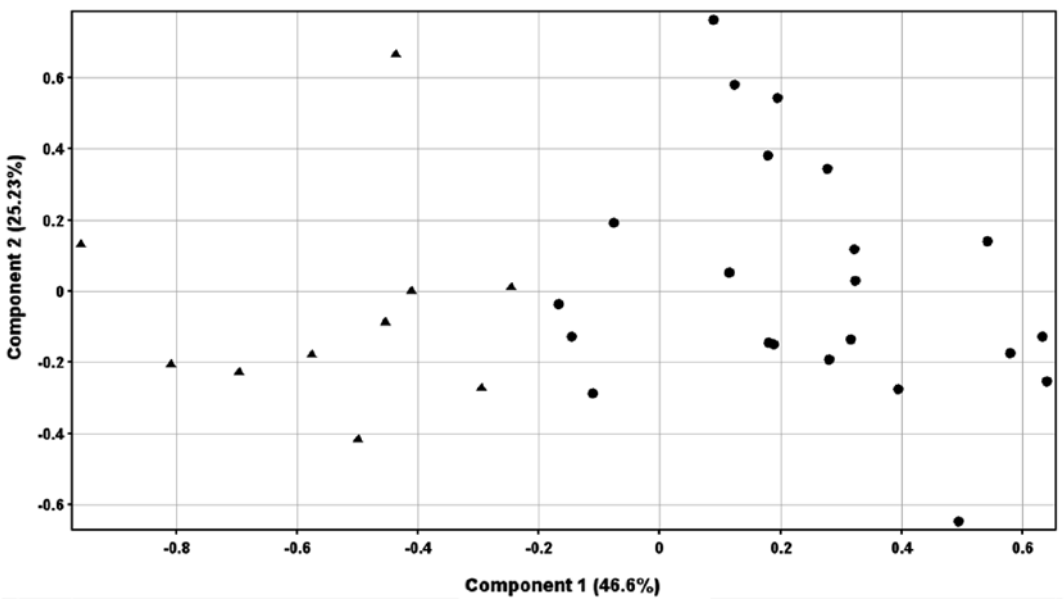

Figure 1. Principal component analysis (PCA) of the 21 USC Gene Set. The first 2 of the 4 calculated eigenvalues are shown, capturing approximately 72% of the variation in the data. A complete separation at -0.2 on the X-axis (component 1) of all responders from all non-responders is shown.

Both tumors of non-responder patients with an initial basal-like molecular subtype retained their basal-like expression pattern (Table V). Therefore, the change of a basal-like subtype to a non basal-like subtype after 4 cycles of NAC was a predictor of responder status in all investigated patients.

Differentially expressed genes in responders compared to non-responders. SAM using pCR as a quantitative outcome and an FDR cut-off of 10% revealed 284 genes as significantly higher and 30 genes as significantly lower expressed in initial tissues of responders compared to non-responders (data not shown). A DAVID analysis with the 219 DAVID-annotated genes from this 314-gene-list found a number of functional clusters (data not shown) with the first six functional clusters all associated with immune system regulation (lymphocytes, leukocytes and T-cells) and/or the response to chemokines.

Differentially regulated genes before and after NAC in responders compared to non-responders. Twenty-five tumor biopsies after 4 cycles of NAC were available. In the remaining 7 biopsies, percentage of tumor cells was below 50% and these samples were therefore excluded from further analysis. To find differentially regulated genes after NAC in responders compared to non-responders, the difference after chemotherapy minus before chemotherapy in $\log_2$ expression-space was calculated (corresponding to the $\log_2$-fold-change) and

Table V. Classification of tumors according to molecular subtype before and after neoadjuvant chemotherapy in 25 patients.

Patient no.	R/NR	Classification before chemotherapy	SSP matched symbols after chemotherapy: Sorlie 398/500 P	Sorlie 398/500 S	Hu 289/306 P	Hu 289/306 S	Parker 49/50 P	Parker 49/50 S	Summary after chemotherapy: B	LA	LB	H	N	Classification
3	NR	Non basal	LB	LB	LB	(LB)[e]	LA	LB	0	1	5	0	0	Non basal
6	NR	Non basal	LB	LB	LB	LB	LB	LB	0	0	6	0	0	Non basal
9	NR	Non basal	LA	LA	LA	LB	LA	LA	0	5	1	0	0	Non basal
12	NR	Non basal	LA	LA	LA	LA	LA	LA	0	6	0	0	0	Non basal
15	NR	Non basal	(N)	(LA)	LA	(LA)	LA	LA	0	5	0	0	1	Non basal
23	NR	Non basal	N	(N)	N	N	N	N	0	0	0	0	6	Non basal
26	NR	Basal	B	B	B	B	B	B	6	0	0	0	0	Basal
27	NR	Non basal	(LA)	(LB)	LA	(LA)	N	LA	0	4	1	0	1	Non basal
28	NR	Non basal	(LA)	(LB)	LA	LA	N	LA	0	4	1	0	1	Non basal
30	NR	Basal	B	B	(B)	B	N	N	4	0	0	0	2	Basal/Normal
31	NR	Non basal	LB	LB	LB	LB	LA	LB	0	1	5	0	0	Non basal
5	NR	Non basal	(LB)	LB	LB	(LB)	LA	LB	0	1	5	0	0	Non basal
7	NR	Non basal	(LB)	LB	H	(B)	LA	LA	1	2	2	1	0	Non basal
8	NR	Non basal	(N)	(N)	LA	(LA)	N	LA	0	3	0	0	3	Non basal
10	NR	Non basal	(LB)	LB	(LA)	(LA)	LA	LA	0	4	2	0	0	Non basal
18	NR	Non basal	H	LB	H	(N)	LA	LA	0	2	1	2	1	Non basal
22	NR	Non basal	LA	(LA)	LA	LA	LA	LA	0	6	0	0	0	Non basal
1	R	Basal	(N)	(LB)	(H)	(N)	LA	LA	0	2	1	1	2	Non basal
4	R	Basal	(LB)	LB	(H)	(B)	N	LA	1	1	2	1	1	Non basal
11	R	Basal	(LB)	(LB)	(N)	(B)	N	LA	1	1	2	0	2	Non basal
13	R	Non basal	(LA)	LB	LA	LA	LA	LA	0	5	1	0	0	Non basal
2	R	Basal	(N)	(LB)	(N)	(N)	N	LA	0	1	1	0	4	Non basal
16	R	Non basal	LB	LB	(N)	(N)	LA	LA	0	2	2	0	2	Non basal
19	R	Non basal	LB	LB	(LA)	(LA)	LA	LA	0	4	2	0	0	Non basal
25	R	Basal	(N)	(B)	(H)	(B)	N	N	2	0	0	1	3	Non basal

SSP, single sample predictor; R, responder (pCR); NR, non-responder (no pCR). Tumors were classified as B, basal-like; LA, luminal A; LB, luminal B; and H, HER2, based on the P, Pearson correlation and S, Spearman correlation analyses. Sybtypes in parentheses indicate those unclassified using a correlation cut-off <0.1. Free-marginal κ, 0.40 (chance-adjusted measure of agreement).

used for SAM with the same conditions as above. Two hundred fifty-eight genes were differentially up-regulated after NAC whereas 109 genes were differentially down-regulated after NAC comparing responders to non-responders (data not shown). Only the following 5 genes of the differentially regulated genes after NAC in responders compared to non-responders (4 are expected by chance) overlapped with the 314-gene list of differentially expressed genes from above: A_32_P34941 (THC2714457), A_24_P125839 (C21orf91), A_23_P339240 (PLCH1), A_32_P24295, A_24_P918518 (P4HTM). In the DAVID analysis with the 266 DAVID-annotated genes from this 367-gene-list (data not shown) functional cluster one was associated with the DNA-protein complex, the nucleosome and chromatin organization, especially histone 2B (H2B). Cluster two was associated with the extracellular matrix. Fig. 3 shows gene expression values of HER4 before and after 4 cycles of NAC indicating HER4 up-regulation in responders.

Prediction of responder-status from microarray data. Using a linear regression model built from a predictor gene set found with an USC approach, complete prediction of response to NAC based on the initial tumor biopsy was achieved with a 21 gene set (AUC=1.0; P=0.000008), encompassing the following genes: A_32_P6015 (MNX1), A_24_P843020 (LOC729111), A_32_P20997 (BU561469), A_32_P227043 (THC2722767), A_23_P256425 (ADAMDEC1), A_23_P1322 (AKR1E2), A_24_P347065 (CPT1A), A_32_P148745 (VWDE), A_32_P183765 (ERBB4), A_23_P339240 (PLCH1), A_24_P923381 (EPR1), A_24_P123190 (PLD1), A_23_P66481 (RTN4RL1), A_32_P10886 (C6orf52), A_23_P9086 (ZDHHC2), A_32_P515920 (LOC400573), A_23_P362694 (C4orf7), A_24_P4171 (FGFR1), A_32_P163739 (KIAA1257), A_24_P156748 (SLC30A2), A_24_P193011 (CCND1). Fig. 1 shows a PCA of the 21 gene set demonstrating complete separation of responders and non-responders.

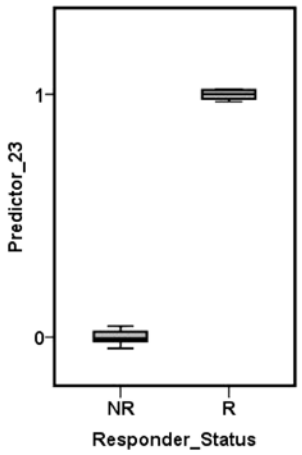

Figure 2. Complete prediction of response to neo-adjuvant chemotherapy using a regression model from a 23 USC-gene set comparing the initial tumor biopsy and the tumor specimens obtained after 4 cycles of neo-adjuvant chemotherapy.

Using a similar linear regression model from the 25 paired tumor tissues with the $\log_2$-fold change calculated from the expression values before and after 4 cycles of NAC, a complete prediction with a 23 gene list was possible (AUC=1.0; P=0.000007), encompassing the following genes: A_32_P30330 (AK130878), A_23_P6714 (SENP7), A_24_P771880 (T70285), A_24_P652174 (AK001007), A_23_P19333 (TREM1), A_32_P105825 (MPPED2), A_24_P567298 (CSAG2), A_32_P141948 (THC2564562), A_23_P203439 (KCNC1), A_32_P183765 (ERBB4), A_23_P203150 (TMPRSS13), A_23_P93348 (LTB), A_32_P152986 (THC2634713), A_32_P47166 (DQ655984), A_32_P112331 (PLD1), A_24_P649050 (AK023816), A_24_P14464 (WFDC2), A_23_P154986 (GGT1), A_23_P215459 (ELN), A_24_P195669 (MYO15B), A_24_P160466 (GPRIN1), A_32_P190303 (LONRF2), A_23_P71170 (TRPV6). Fig. 2 shows a box plot demonstrating complete prediction of response using the 23 gene list. Both the initial expression and the change of expression of the only gene present in both predictor gene sets, i.e. HER4, predicted response to NAC in 26/32 (81%) and 23/25 (92%) patients, respectively. Comparing the AUC values of ROC curves, both the initial expression and the change of expression of HER4 were better predictors than tumor grade, estrogen receptor, progesterone receptor, HER2 status and molecular subtype (basal-like vs. non basal-like) (Table IV).

Network analysis. Eight of 21 gene products of the predictive 21 gene set as described above were annotated in the I2D-database. A network analysis of protein-protein-interactions using these 8 gene products resulted in 3 networks, which is significantly less than expected from 1000 networks seeded by 8 arbitrary seeding-proteins (mean 7.1±1.1 networks, P<0.001): one network consisted of 192 proteins with HER4 as the central component, and two small two-component networks. Using these 192 proteins in a functional DAVID-analysis revealed the KEGG-pathway 'Pathways in cancer' as the most over-represented category with 52/192 proteins involved in this pathway. This indicates that HER4 plays a central role in connecting cancer pathways in breast cancer, ie apoptosis, proliferation, cytokine-signaling and MAPK-signaling. In the 21 gene set, these pathways are represented by BIRC5, CCND1, FGFR1 and PLD1. The validity of the microarray findings was confirmed by real-time polymerase chain reaction (PCR) for selected genes in a subset of patient samples. As expected, reverse transcriptase (RT)-PCR experiments paralleled the findings obtained by the cDNA arrays in all cases (data not shown).

Discussion

In this study, we found that gene expression-based molecular subtype, the basal-like subtype, was associated with pCR after NAC. Specifically, 7/10 basal-like tumors responded to NAC, whereas 19/22 non basal-like tumors did not respond. All patients with an initial non basal-like tumor retained this tumor subtype after 4 cycles of NAC, whereas all basal-like tumors responding to NAC lost this gene expression pattern. A 21 gene

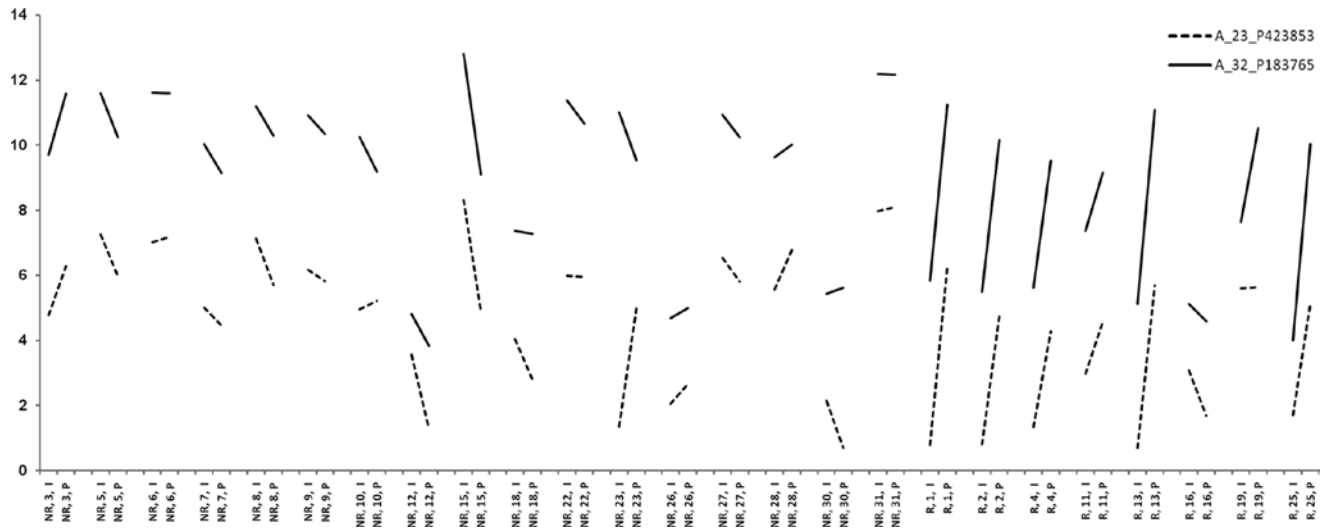

Figure 3. Log_2 gene expression values of both probes for HER4 (A_23_P423853 and A_32_P183765) before (I) and after 4 cycles of neo-adjuvant chemotherapy (P) for each patient. On the left side all non-responders (NR) are shown; on the right side all responders (R) are shown.

list derived from gene expression before NAC and a 23 gene list derived from gene expression changes during chemotherapy was significantly associated with pCR after NAC. Of note, both initial low expression and up-regulation of expression during chemotherapy of a single gene, i.e. HER4, was found in 26/32 (81%) and 23/25 (92%) responders to NAC, respectively. These data indicate that basal-like molecular subtype, low initial HER4 expression, and HER4 up-regulation during therapy are associated with response to NAC with epirubicine, cyclophosphamide and docetaxel.

Molecular subtype classification into basal-like vs. non basal-like subtypes has recently been proposed (11). Using this approach, we found that tumors expressing a basal-like molecular subtype, but not luminal A, luminal B, or HER2-subtypes, were associated with pCR after NAC, a finding previously reported by others.

Comparing differentially expressed genes in the initial tumor samples from responders and non-responders to NAC, genes involved in the regulation and signalling of the immune system were significantly overrepresented. Comparison of differentially expressed genes calculated from the expression levels after NAC (biopsy two) minus the expression levels before NAC (initial biopsy) revealed overrepresentation of genes from the DNA-protein complex. While the latter is not of surprise, given the DNA damaging potential of epirubicine and cyclophosphamide, the former is of more interest, indicating a substantial influence of the immune system on the response to NAC.

Our study has limitations. The sample size of this preliminary study is small and the low number of basal-like tumors may bias the results. Our data have to be confirmed in an independent validation set of patients. Also, complete prediction of response was achieved by a 21 gene list which is overlapping in 4 genes, but is not identical with all genes constituting the basal-like molecular subtype. This indicates that relevant genes involved in determining the response to NAC classify into various molecular subtypes, raising questions about the validity of gene expression profiling as a predictive marker. In this respect, it is of note, that we identified HER4 as the single most important factor in predicting response to the NAC regimen tested in this study.

A network analysis of protein-protein interactions using annotated genes of the 21 gene list identified in this study pointed to a protein network involved in cancer pathways with HER4 as the central component. This indicates that HER4 may play a central role in connecting cancer pathways in breast cancer, apoptosis, proliferation, cytokine-signaling, and MAPK-signaling. In the 21 gene set, these pathways were represented by BIRC5, CCND1, FGFR1 and PLD1.

In summary, we present gene expression data from tumor samples of patients with primary invasive breast cancer, identifying the basal-like molecular subtype and HER4 up-regulation during chemotherapy as potential predictors of response to NAC with epirubicine, cyclophosphamide and docetaxel.

Acknowledgements

This study was supported by an unrestricted grant from Sanofi-Aventis.

References

1. Carlson RW, Anderson BO and Bensinger W: NCCN practice guidelines for breast cancer. Oncology 14: 33-49, 2000.
2. National Institutes of Health Consensus Development Conference. [http://www.consensus.nih.gov; accession date: 05-04-2011]
3. Hortobagyi GN, Hayes D and Pusztai L: Integrating newer science into breast cancer prognosis and treatment: molecular predictors and profiles. American Society of Clinical Oncology, Annual Meeting Summaries, pp191-202, 2002.
4. Fisher B, Bryant J, Wolmark N, *et al*: Effect of preoperative chemotherapy on the outcome of women with operable breast cancer. J Clin Oncol 16: 2672-2685, 1998.
5. van de Vijver MJ, He YD, van't Veer LJ, *et al*: A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 347: 1999-2009, 2002.
6. Chang JC, Wooten EC, Tsimelzon A, *et al*: Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer. Lancet 362: 362-369, 2003.
7. Ayers M, Symmans WF, Stec J, *et al*: Gene expression profiles predict complete pathologic response to neoadjuvant paclitaxel and fluorouracil, doxorubicine, and cyclophosphamide chemotherapy in breast cancer. J Clin Oncol 22: 2284-2293, 2004.
8. Sotiriou C, Powles TJ, Dowsett M, *et al*: Gene expression profiles derived from fine needle aspiration correlate with response to systemic chemotherapy in breast cancer. Breast Cancer Res Treat 4: R3, 2002.
9. Kelly CM, Warner E, Tsoi DT, Verma S and Pritchard KI: Review of the clinical studies using the 21-gene assay. Oncologist 15: 447-456, 2010.
10. Perou CM, Sorlie T, Eisen MB, *et al*: Molecular portraits of human breast tumours. Nature 406: 747-752, 2000.
11. Weigelt B, Mackay A, A'hern R, *et al*: Breast cancer molecular profiling with single sample predictors: a retrospective analysis. Lancet Oncol 11: 339-349, 2010.
12. Sorlie T, Tibshirani R, Parker J, *et al*: Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci USA 100: 8418-8423, 2003.
13. Hu Z, Fan C, Oh DS, *et al*: The molecular portraits of breast tumors are conserved across microarray platforms. BMC Genomics 7: 96, 2006.
14. Parker JS, Mullins M, Cheang MC, *et al*: Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol 27: 1160-1167, 2009.
15. Jeruss JS, Mittendorf EA, Tucker SL, Gonzalez-Angulo AM, Buchholz TA, Sahin AA, Cormier JN, Buzdar AU, Hortobagyi GN and Hunt KK: Staging of breast cancer in the neoadjuvant setting. Cancer Res 68: 6477-6481, 2008.
16. Brazma A, Hingamp P, Quackenbush J, *et al*: Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nat Genet 29: 365-371, 2001.
17. Tusher VG, Tibshirani R and Chu G: Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci USA 98: 5116-5121, 2001.
18. Brennan RL and Prediger DJ: Coefficient Kappa. Some uses, misuses, and alternatives. Educ Psychol Meas 41: 687-699, 1981.
19. Saeed AI, Bhagabati NK, Braisted JC, *et al*: TM4 microarray software suite. Methods Enzymol 411: 134-193, 2006.
20. Brown KR, Otasek D, Ali M, *et al*: NAViGaTOR: network analysis, visualization and graphing Toronto. Bioinformatics 25: 3327-3329, 2009.
21. R Development Core Team: R. A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria, 2009. http://www.R-project.org; accession date 05-04-2011.
22. Talloen W, Clevert DA, Hochreiter S, *et al*: I/NI-calls for the exclusion of non-informative genes: a highly effective filtering tool for microarray data. Bioinformatics 23: 2897-2902, 2007.
23. Huang W, Sherman BT and Lempicki RA: Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4: 44-57, 2009.