Gene expression profilers and conventional clinical markers to predict distant recurrences for premenopausal breast cancer patients after adjuvant chemotherapy

Emma Niméus-Malmström[a,d], Cecilia Ritz[b,d], Patrik Edén[b], Anders Johnsson[a],
Mattias Ohlsson[b], Carina Strand[a], Görel Östberg[c], Mårten Fernö[a,*], Carsten Peterson[b]

[a]Department of Oncology, Institute of Medical Sciences, University Hospital, Lund, Sweden
[b]Department of Theoretical Physics, Lund, Sweden
[c]Department of Pathology, Hospital, Halmstad, Sweden

ARTICLE INFO

ABSTRACT

A large proportion of breast cancer patients are treated with adjuvant chemotherapy after the primary operation, but some will recur in spite of this treatment. In order to achieve an improved and more individualised therapy, our knowledge in mechanisms for drug resistance needs to be increased. We have investigated to what extent cDNA microarray measurements could distinguish the likelihood of recurrences after adjuvant CMF (cyclophosphamide, methotrexate and 5-fluorouracil) treatment of premenopausal, lymph node positive breast cancer patients, and have also compared this with the corresponding performance when using conventional clinical variables.

We tried several gene selection strategies, and built classifiers using the resulting gene lists. The best performing classifier with odds ratio (OR) = 6.5 (95% confidence interval (CI) = 1.4–62) did not outperform corresponding classifiers based on clinical variables. For the clinical variables, calibrated on the samples, either using all the clinical parameters or the Nottingham Prognostic Index (NPI) parameters, the areas under the receiver operating characteristics (ROC) curve were 0.78 and 0.79, respectively. The ORs at 90% sensitivity were 15 (95% CI = 3.1–140) and 10 (95% CI = 2.1–97), respectively. Our data have provided evidence for a comparable prediction of clinical outcome in CMF-treated breast cancer patients using conventional clinical variables and gene expression based markers.

© 2006 Elsevier Ltd. All rights reserved.

1. Introduction

Breast cancer is a heterogeneous disease with a large variability in clinical outcome. Adjuvant polychemotherapy (e.g. with cyclophosphamide, methotrexate and 5-fluorouracil (CMF)) or anthracycline-containing regimes, produce substantial reduction in recurrence and mortality. In the metaanalysis, performed by the Early Breast Cancer Trialist's Collabration Group, the absolute improvement in 15-year breast cancer survival after adjuvant polychemotherapy was 10% (from 58% to 68%) for patients under the age of 50.[1] Besides an improvement in clinical outcome, these figures indicate that a large proportion of the patients will never recur after the primary operation and do not, consequently, need any further

treatment with unnecessary side effects. Also, a considerable proportion of the patients will recur despite treatment with adjuvant polychemotherapy. Substantial efforts have been made to identify the group that does not need adjuvant systemic therapy, and to explain mechanisms why some patients recur in spite of chemotherapy. Possible mechanisms for recurrence after treatment are low initial drug sensitivity or an acquired drug resistance. In order to achieve a more effective and individualised chemotherapeutic treatment of breast cancer patients in the future, it is essential to increase our knowledge in mechanisms responsible for drug resistance, and to define reliable indicators for response to therapy. Commonly accepted prognostic factors are lymph node status, tumour size, histological grade, and patient age. Predictors for the effect of endocrine treatment, currently used in clinical routine, are oestrogen (ER) and progesterone receptor (PgR) status, and for the effect of monoclonal antibodies (trastuzumab) c-erbB-2 is used. Useful markers for resistance and/or sensitivity of chemotherapy (CMF and/or antracyclin based regimes) have not, so far, been identified. Some markers have shown promising results in a limited number of studies, e.g. thymidylate synthase and thymidine kinase,[2–4] c-erbB-2,[5–7] p53,[8–11] topoisomerase IIα, and multidrug resistance-associated protein.[8,12–14]

The development of techniques for gene expression analyses enables an extensive characterisation of malignant tumours. Studies using these techniques in breast cancer have shown distinct differences in gene expression profile between hereditary and sporadic breast cancer,[15] and between ER positive and ER negative cancer.[16,17] Promising results have also been obtained for predicting clinical outcome,[17–21] both in patients not treated with adjuvant therapy[19–21] and in patients treated with adjuvant therapy, endocrine, chemotherapy, or both.[17–19] Furthermore, gene expression analysis have identified genes involved in mediating the response to cytotoxic drugs, e.g. 5-fluoruracil in breast and colorectal cancer cell-lines, and oesophageal cancer,[11,22] cisplatin in oesophageal cancer,[22] anthracyclines in breast cancer cell-lines,[23] and neoadjuvant taxane treatment in breast cancer.[24] However, before the above mentioned results can be applied in clinical routine, the data needs to be confirmed, using other array platforms and other patient materials. Furthermore, one important issue concerns whether the gene expression analysis provides information about clinical outcome and treatment sensitivity, in addition to the information obtained by conventional clinical factors, already in routine use. A recent publication from our group[25] has stressed this issue, by showing that clinical markers have similar power in predicting breast cancer prognosis as cDNA microarray gene expression profilers, using publicly available data.[20]

In this study, we have used cDNA microarray analysis to predict recurrences after adjuvant treatment of CMF in a well-defined cohort of patients (premenopausal and lymph node positive). The ability to predict recurrences after CMF was also evaluated using clinical markers, publicly available cDNA expression data used for predicting clinical outcome,[20] and a gene expression profile associated with response to chemotherapy, based on prior knowledge, obtained after literature search.

2. Patients and methods

2.1. Patient selection

According to treatment guidelines in the regional care program for breast cancer in southern Sweden issued 1991, premenopausal lymph node positive (N+) breast cancer patients were recommended postoperative radiation and adjuvant chemotherapy. Radiotherapy was delivered to ipsilateral axillary and supraclavicular lymph nodes and the remaining breast parenchyma after breast conservation surgery or thoracic wall after mastectomy. The absorbed target dose was 50 Gy in 25 fractions in one series during 5 weeks. The standard chemotherapy at that time period was nine cycles of CMF. Patients for the present study were stringently selected in a stepwise manner to fulfil the following criteria: premenopausal women with primary breast carcinoma, stage T1-3N1-2M0, diagnosed 1992–97, frozen primary tumour samples were still available, referred to the Department of Oncology in Lund or Malmö for adjuvant radiotherapy, treatment with nine cycles of CMF, either distant recurrence within 40 months after completion of CMF or remained free from distant recurrence for 40 months or longer, good quality of extracted RNA, and successful hybridisation. After this selection process (Fig. 1) we ended up with 29 recurrences and 56 recurrence-free patients that were included in the analysis (Table 1). The study was approved by the ethics committee at Lund University.
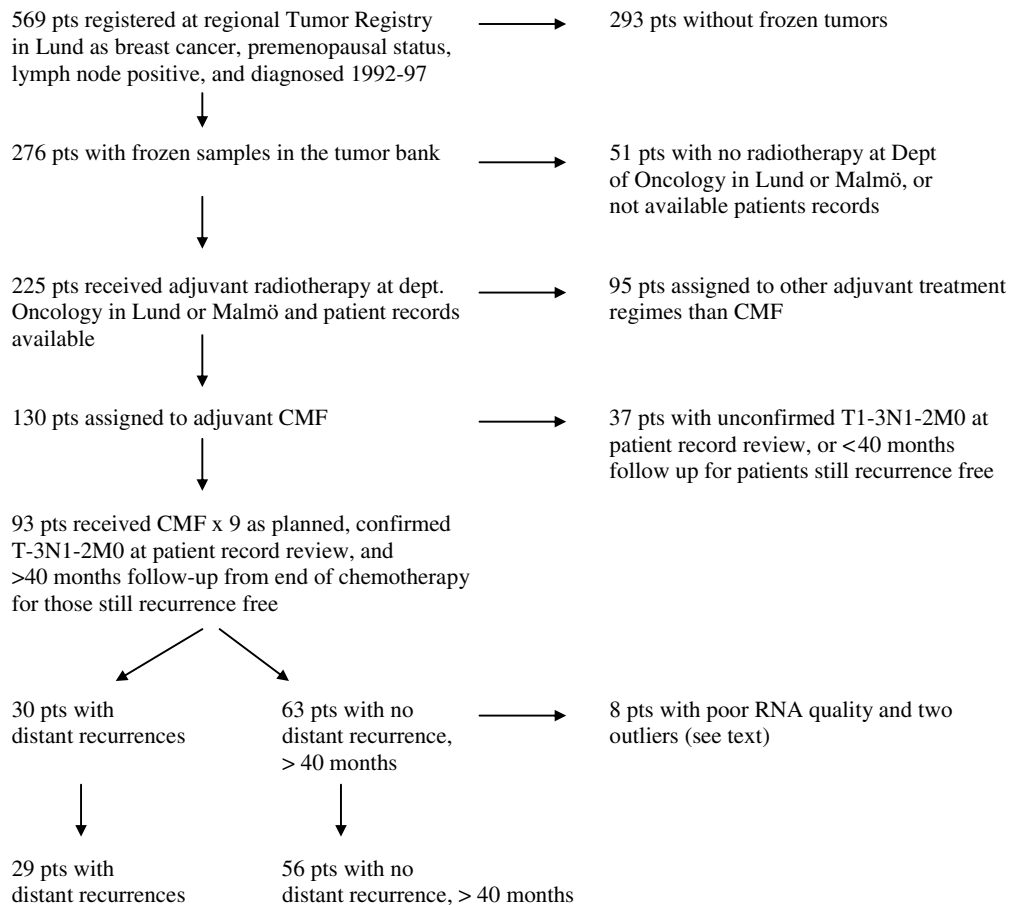
2.2. Chemotherapy

Patients were treated with an intravenous CMF schedule; cyclophosphamide 600 mg/m^2, methotrexate 40 mg/m^2 and 5-fluorouracil 600 mg/m^2, on day 1, every 3 weeks, for nine cycles. According to the regional guidelines, chemotherapy should be started within 1 month after surgery. Radiotherapy was started within 1 month after initiation of CMF. During the 5 weeks of radiotherapy, cyclophosphamide was given at a dose of 850 mg/m^2 every 3 weeks, while methotrexate and 5-fluorouracil were omitted. The delivered chemotherapy doses were calculated and could be retrieved in 83 of the 85 patients records. The actual dose intensities mg/m^2/week were calculated and showed to be almost identical in the two groups; 93% of the planned doses for recurrence-free patients compared to 92% of the patients with recurrences. The main toxicity of CMF treatment was leucopenia. Dose reduction due to leucopenia (white blood cells <3.0 × 10^9/L) was performed in 65% of the recurrence-free patients and in 60% of the patients that later developed distant recurrence ($p = 0.63$, chi-square-test).

2.3. Methods

2.3.1. Conventional prognostic and treatment predictive factors
Histological grade was re-evaluated for all the samples by the same observer according to Elston and Ellis.[26] The grading procedure consisted of judgment of tubule formation, nuclear plemorphism, and mitotic count. Each of these morphological features was given a score of 1–3 points. The overall histolog-

```
569 pts registered at regional Tumor Registry     ────▶   293 pts without frozen tumors
in Lund as breast cancer, premenopausal status,
lymph node positive, and diagnosed 1992-97

          │
          ▼

276 pts with frozen samples in the tumor bank     ────▶   51 pts with no radiotherapy at Dept
                                                          of Oncology in Lund or Malmö, or
          │                                               not available patients records
          ▼

225 pts received adjuvant radiotherapy at dept.   ────▶   95 pts assigned to other adjuvant treatment
Oncology in Lund or Malmö and patient records             regimes than CMF
available
          │
          ▼

130 pts assigned to adjuvant CMF                  ────▶   37 pts with unconfirmed T1-3N1-2M0 at
                                                          patient record review, or <40 months
          │                                               follow up for patients still recurrence free
          ▼

93 pts received CMF x 9 as planned, confirmed
T-3N1-2M0 at patient record review, and
>40 months follow-up from end of chemotherapy
for those still recurrence free

        ╱        ╲
       ▼          ▼

30 pts with        63 pts with no              ────▶   8 pts with poor RNA quality and two
distant recurrences   distant recurrence,              outliers (see text)
                   > 40 months

       │              │
       ▼              ▼

29 pts with        56 pts with no
distant recurrences   distant recurrence, > 40 months
```

Fig. 1 – The selection of patients included in the study.

ical grade was obtained by adding these points, and was categorized as follows: grade 1, 3–5 points, grade 2, 6–7 points, and grade 3, 8–9 points. The Nottingham Prognostic Index (NPI) is a linear combination of lymph node status, tumour size, and histological grade, according to the formula:[27] $NPI = 0.2 \times$ tumour size (in cm) + lymph node status + histological grade, where lymph node status is 1 for node negative, 2 for 1–3 tumour-involved nodes and 3 when 4 or more nodes are tumour-involved.

ER and PgR were analysed routinely, at the time of the primary operation, with enzyme immunoassay according to kit instructions (Abbott Laboratories, Diagnostic Division, Chicago, IL, USA), and expressed as fmol per mg cytosol protein. Receptor values above or equal to 25 fmol/mg protein were considered positive.

The analysis of S-phase fraction (SPF) was also performed as part of clinical routine in an Ortho Cytoron Absolute flow cytometer (Ortho Diagnostic Systems, Raritan, NJ, USA). Samples with an SPF $\geqslant$ 12% were classified as high SPF, and those samples with values below these levels as low SPF.[28]

2.3.2. RNA isolation and cDNA microarray
Total RNA was isolated from fresh frozen tumours using Trizol (Invitrogen, Carlsbad, CA) and purified with the RNeasy® Midikit (Qiagen Inc, Valencia, CA). RNA quality was assessed with an Agilent 2100 Bioanalyzer RNA 6000 Lab.Chip kit (Agilent Technologies, Palo Alto, CA) and six samples were excluded due to poor RNA quality. The protocol for cDNA microarray has been reported previously.[29] Briefly, the arrays were spotted with 27,648 sequence-verified cDNA clones (Unigene). Labeled cDNA was produced using 25 μg of tumour RNA and 10 μg Stratagene Reference RNA (Stratagene, La Jolla, CA) by anchored primed reverse transcriptase using CyscriptRT from the Cyscribe post labelling kit and Cy5-dUTP or Cy3-dUTP (Amersham Biosciences, Piscataway, NJ). Agilent software (Agilent technologies, Palo Alto, CA) was used for fluorescence scanning at 5 μm resolution and Gene pix Pro software (Axon Instruments, Inc., Union City, CA) for image analysis.

2.3.3. Data mining methods
2.3.3.1. Gene expression analysis. Gene expression analysis proceeded in three steps: (i) preprocessing, (ii) selection of significant genes, and (iii) construction of classifier.

(i) *Preprocessing*. The data was stored in BASE [30] (BioArray Software Environment) after the initial image processing step. Pearson correlations of log reference intensities were calculated for all pairs of assays. The mean Pearson correlation for an assay ranged from 0.88–0.93, except for two assays, which had average Pearson correlation 0.73 and 0.13, respectively. These two assays were excluded from the following analysis. In BASE, a LOWESS normalisation was applied to the log ratios.[31] Replicate measurements x_i of the same reporter on an assay were merged as in [32] and represented by a weighted mean $m = \Sigma_i w_i x_i / \Sigma_i w_i$, where the weight w_i is

Table 1 – Clinical and biological characteristics of 85 premenopausal patients, with lymph node positive breast cancer, treated with adjuvant CMF, subdivided with respect to if they have developed distant recurrences or not		
Clinical parameter	Distant rec	No rec
Age at diagnosis		
< 40 years	6	9
40–50 years	16	44
> 50 years	5	5
Tumour size		
T1, ⩽ 20 mm	3	24
T2, > 20–50 mm	21	33
T3, > 50 mm	0	1
missing value	3	0
Lymph nodes		
1–3 pos lymph nodes	16	45
⩾ 4 pos lymph nodes	11	13
Histological grade		
1	1	12
2	2	15
3	23	27
Missing value	1	4
ER		
< 25 fmol/mg protein	19	20
⩾ 25 fmol/mg protein	8	38
PgR		
< 25 fmol/mg protein	18	22
⩾ 25 fmol/mg protein	9	36
SPF		
< 12%	7	29
⩾ 12%	16	24
Missing value	4	5

$\exp(-3u_i^{1/2}/|x_i - m|)$, the estimated uncertainty of a spot u is $SNR_1^{-2} + SNR_2^{-2}$, and SNR_i is the signal to background noise ratio for channel i. The set of equations for m was solved numerically by simple iteration. The error of the merged value was defined as $U = 1/\Sigma_i(1/u_i) + \Sigma_i w_i^2(x_i - m)^2/(\Sigma_i w_i)^2$. We then modified expression values according to an error model [29] where expression values x_i, now representing the value merged on reporter, with large uncertainties u_i were moved towards the weighted mean m across assays for that reporter. The modified expression value was given by $x_i' = w_i(x_i - m)$. After reducing the importance of low-quality measurements in this way, the quality weights were not used in the following analysis. Reporters were excluded if missing in more than 10% of the samples or if the standard deviation of the modified log ratios was less or equal to 0.3. After these steps, 4484 reporters remained for further processing.

(ii) *Selection of significant genes*. Reporters were ranked according to the Pearson correlation between (modified) gene expression log ratios and the clinical outcome M (M = 1 for recurrence and M = 0 for no recurrence). The false-discovery rate (FDR), defined as the fraction of reporters having a Pearson correlation higher than a chosen cut-off value by chance,[33] was estimated from the Pearson correlation density of 1000 sample label permutations.

(iii) *Construction of classifier*. This step was done following closely what was done earlier.[16,34,35] The top ten or top 100 genes with the highest Pearson correlation to clinical out-

come were subject to principal component analyses (PCA), and the principal components with largest eigenvalues were used for construction of a committee of artificial neural network (ANN) classifiers. The performance was tested by applying the committee of networks to blind tests. In Khan et al.[35] and Gruvberger et al.[16] single test sets were used. Our goal was to compare different classifier performances, and multiple test set divisions then provide more reliable estimates.[36] The need to retain sufficiently large training sets motivated small test sets. However, this leads to large variation between test set results,[36] and many random test sets must be considered. Already facing substantial computational costs when ranking genes and selecting ANN designs, we therefore adopted a slightly different approach, where the ANN output values for all test samples were compiled and finally used to produce a single test result, as an estimate of the average result. With this approach, the test set size no longer poses a major problem, and we adopted a leave-one-out procedure.

In the cross-testing scheme, every member of a predefined pool of different ANN designs (and a new set of genes) was considered for each new blind test selection. The pool contained all combinations of the following parameters: number of inputs = 2, 4, 6, 8, 10; number of hidden nodes = 0, 2; and weight decay parameter = 0, 0.01, 0.03, 0.1. Back propagation (with learning rate = 0.75 and momentum parameter = 0.1) was used to minimise the error function during 50 training epochs, and for each iteration the learning rate was decreased by a factor of 0.98.

The performance of the classifiers from the different gene sets was measured by the area under the receiver operating characteristics curve (ROC area).[37] We also calculated the odds ratios (ORs) after setting the thresholds corresponding to 10% misclassified in the distant recurrence group. The interpretation of ORs is known to be delicate [38] but they are included here for easier comparison to other studies.[20,25] All ORs in this paper are calculated at 90% sensitivity, making the comparison between them more straightforward than in the most general case. Compared to the ROC area, the OR is closer to a clinical reality, where a decision threshold must be implemented, but sensitive to noise in the studied data set. The ROC area represents a performance average over a wide range of thresholds, and is therefore less sensitive to noise and may better indicate which classifying approach that has the highest potential.

2.3.3.2. *Clinical variables analysis*. Five samples missing histological grade were excluded from the clinical variables analysis. Missing values for SPF were replaced with the mean over all samples. All tumors were annotated as T1 (⩽ 20 mm), T2 (> 20 mm–50 mm), or T3 (> 50 mm). In two cases, T stage was the only available information of tumour size. In order to get a numerical value of size for all samples, these two missing values were replaced by the mean size over the samples with the same T stage annotation (T1 and T3, respectively).

ER, PgR, SPF, and tumour size were in the statistical analysis used as continuous variables.

Two approaches were taken using conventional variables only. In the first one, the NPI [27] was computed for all patients without any learning steps. In the other approach, ANN mod-

els were constructed according to the cross-testing and cross validation scheme above, with the exception that the PCA step was not performed, since there were only seven variables (number of tumour-involved lymph nodes, tumour size, histological grade, age, ER, PgR, and SPF). The second approach was employed using all seven clinical variables, and also using only the three parameters included in the NPI (number of tumour-involved lymph nodes, tumour size, and histological grade).

Hybrid classifiers were constructed in the same way as for the gene expression data, but with the clinical variables added as input nodes. The pool of ANN designs was identical except for the number of inputs. The hybrid classifier with seven clinical variables had 9, 11, 13, 15, or 17 numbers of inputs, and the hybrid classifier with the NPI variables had 5, 7, 9, 11 or 13.

2.3.4. Search strategy and selection criteria for drug associated genes

Data for the list of known drug associated genes (drug-genes) was identified during December 2003 – February 2004 (see Supplement) in two ways. First, already available articles within this subject were selected. Secondly, published articles, since 1997, were obtained by two separate searches of PubMed (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi). The first one included the search terms 'drug resistance', 'cancer', 'cyclophosphamide', 'methotrexate', and '5-FU'. The second search included 'drug resistance' and 'cancer', in order to find genes involved in sensitivity or resistance to other regimes than CMF (see Supplement).

The selection of genes was performed prior to the data analyses. To find the reporters on our array corresponding to the genes on our pre-defined gene lists, we used the official gene symbol. The gene symbols for the van 't Veer gene list [20] were obtained through ACID [39] using UniGene build 176, and the official gene symbols for the drug-genes were found manually using Gene and Locus Link. All reporters on our array that according to UniGene build 180 had a gene symbol represented on the resulting list were selected. Among the 14,717 reporters that had less than 10% missing values, 245 matched the 253 initially pre-selected drug genes and 184 matched the 231 van 't Veer genes. We confirmed that the 184 van 't Veer-genes available in our study had similar predictive power in the data set of van 't Veer and co-workers, as had the full set of 231 genes.

3. Results

3.1. False-discovery rates in the different gene sets

The FDR was 6% for the top-100 reporters in the full reporter list containing 4484 reporters (Table 2), but noticeably higher in the two pre-selected reporter sets. Restricting the analysis (both gene ranking and permutations test) to the drug-genes gave a 42% FDR, and correspondingly a 31% FDR for the van 't Veer genes. The top-100 reporters from the different reporter lists can be found in the supplementary information.

When using the top-10 reporters, the FDR:s were lower (Table 2). The unrestricted top 10-list is listed in Table 3a,

Table 2 – Pearson correlation coefficient and number of false positive among 100-top genes for unrestricted, drug, and van 't Veer genes

Reporter set	Pearson correlation	False discovery rate (%)
Unrestricted		
top 10	0.42	2.0
top 100	0.34	6.2
Drug-genes		
top 10	0.30	9.3
top 100	0.15	42
van 't Veer		
top 10	0.34	1.8
top 100	0.15	31

and this list includes genes involved in functions such as signalling, gamma-aminobutyric acid metabolism, RNA processing, N-linked glycosylation via asparagines, electron transport, nucleotide binding, activation of T and natural killer cells, ATP binding, and metalloendopeptidase inhibitor activity. The top drug-genes were according to earlier studies important for resistance mechanisms to doxo- and epirubicin treatment,[10,23] methotrexate [10,40] and docetaxel, [24] cisplatin,[41] 5-FU, [11] vincristine,[10] vindesin,[10] mitomycin C [10] and thiotepa,[10] and they are involved in cell proliferation, RNA processing, DNA-damage response, nucleotide biosynthesis, N-linked glycosylation via asparagines, estrogen receptor signaling pathway, and anti-apoptosis (Table 3b).

3.2. Predictive power of the different gene sets

Using PCA and ANN, the different lists of top-10 and top-100 reporters were used to classify the two groups of patients, with and without distant recurrences after adjuvant CMF, after proper division of data into training and validation tests (see Materials and methods). For each blind test sample, a new ranking of reporters was performed, based on the remaining samples. This was done to avoid information leaks in the analysis. Thus, the resulting predictions were not a test of a specific top reporter list, but rather a test of the full reporter set from which top lists were generated. The result in terms of ROC area (see Materials and methods) was higher for the drug-genes top-10 reporters than for the other two top-10 gene selections, which both had similar results. The ORs were significantly above 1 (> 95% CI, Fisher's exact test) for the drug-genes top-10 reporters and for the unrestricted top-10 reporters. Selecting the top-100 reporters gave worse prediction performance for all three reporter sets, both in terms of ROC area and OR.

3.3. Predictive power of the clinical variables

When using the same ANN procedure to build a classifier, including leave-one-out, NPI parameters and clinical markers yielded ROC areas comparable to the drug-genes top-10 reporter result, and higher ORs than all tested classifiers based on gene expression (Table 4). The classifier using all seven clinical markers performs better than the one using only the three NPI parameters. Using NPI directly, without calibrating any

Table 3 – A list of the top-10 unrestricted genes (a) and drug genes (b) were ranked using Pearson correlation and classified with ANN. +/− indicates if the gene is up or down-regulated in the group with no distant recurrences

Gene name	Gene symbol	Acc number	Up/down
(a)			
4-aminobutyrate aminotransferase	ABAT	BC008990	+
Serum/glucocorticoid regulated kinase-like	SGKL	H98714	+
Thyroid hormone receptor interactor 13	TRIP13	AA630784	−
Interleukin 12A (natural killer cell stimulatory factor 1)	IL12A	AI304577	−
Hypothetical protein FLJ40629	FLJ40629	AA417744	−
Dolichyl-diphosphooligosaccharide-protein glycosyltransferase	DDOST	H96437	−
Arginine-rich, mutated in early stage tumours	ARMET	R91550	−
RNA binding protein with multiple splicing	RBPMS	W67323	+
Chromosome 20 open reading frame 129	C20orf129	R96998	−
ERO1-like (S. cerevisiae)	ERO1L	AA186804	−
(b)			
Dolichyl-diphosphooligosaccharide-protein glycosyltransferase	DDOST	H96437	−
RNA binding protein with multiple splicing	RBPMS	W67323	+
Cell division cycle 27	CDC27	T81764	+
Baculoviral IAP repeat-containing 5 (survivin)	BIRC5	AA460859	−
Oestrogen receptor 1	ESR1	AA291702	+
V-abl Abelson murine leukemia viral oncogene homolog 1	ABL1	H91096	−
Fusion (involved in t(12;16) in malignant liposarcoma)	FUS	W67581	+
X-ray repair complementing defective repair in Chinese hamster cells 1	XRCC1	AA425139	+
V-raf murine sarcoma viral oncogene homolog B1	BRAF	W88566	−
Dihydrofolate reductase	DHFR	N52980	−

+/− indicates if the gene is up or down-regulated in the group with no distant recurrences.

classifier on the data set, improved the results in terms of ROC area further. Using both clinical and gene expression data in hybrid classifiers did, however, not improve the results.

Table 4 – The effectiveness of variables for separating in recurrence versus recurrence-free patient groups is measured using the ROC area and odds ratios (OR), using the top ranked reporters of the unrestricted (unrestr.), drug and van 't Veer reporter sets, respectively

Reporter set	ROC	OR	95% CI (Fischer's exact test)
Unrestricted			
Top 10	0.70	6.5	1.4–62
Top 100	0.60	2.0	0.36–21
Drug-genes			
Top 10	0.78	6.0	1.3–57
Top 100	0.57	2.3	0.42–23
van 't Veer			
top 10	0.69	3.9	0.80–38
top 100	0.65	1.9	0.36–21
Clinical variables and combinations			
All seven	0.78	15	3.1–140
Incl.top 10 unrestr.	0.71	1.2	0.18–14
Incl.top 100 unrestr.	0.66	1.5	0.24–16
Three NPI parameters	0.74	10	2.1–97
Incl.top 10 unrestr.	0.72	5.0	1.0–48
Incl.top 100 unrestr.	0.76	2.1	0.37–140
NPI	0.79	10	2.1–97

As a comparison, the corresponding values for NPI and the seven clinical variables, as well as the combinations of clinical variables and the unrestricted reporter set, are shown.

3.4. Gene ontology

The three top 100 gene lists (unrestricted, drug, and van 't Veer) were functionally classified by annotating the genes with gene ontology followed by clustering into biological processes. Out of the most frequent biological processes, three processes were found on all three gene lists, mitosis, cytokinesis, and regulation of cell cycle, which are all processes related to cell proliferation. Data also indicates that the drug and van 't Veer lists are more similar since several processes such as cell cycle and cell growth maintenance were uniquely common in these two gene lists. Some processes were represented in only one of the gene lists. In the unrestricted gene list several biological processes involving signaling were more common, whereas in the drug gene list, biological processes involving protein modifications and regulation of cell proliferation were found. In the van 't Veer gene list no clear trend could be found due to too few processes present only for this list.

4. Discussion

The present study was focused on trying to explain why certain patients recur in spite of adjuvant chemotherapy (CMF). Currently available conventional factors are not considered sensitive enough for this selection. We constructed classifiers based on conventional markers and gene expression as measured by cDNA microarrays. We found that gene expression data could not improve the predictions. The strength of the conventional markers in relation to the gene expression profile is thus a confirmation of the results from a previous paper from our group,[25] using publicly available data of van 't Veer and coworkers.[20]

The incapacity of gene expression analysis to improve predictive power may simply be due to a too small cohort in the study; one could hypothesise that a marker based on multidimensional gene expression data would benefit more from a larger study than would already established clinical markers. Our studied cohort was large enough to identify genes relevant for development of distant recurrences after adjuvant CMF (6% FDR among top 100 ranked genes), but may still be too small to fully avoid overtraining when building the classifiers. The fact that a combination of clinical parameters and gene expression data failed to improve the results, and sometimes reduced them, points in this direction. The relatively poor performances of the hybrid classifiers should therefore not be seen as any evidence of complete overlap between information from clinical and gene expression based markers. Among the classifiers investigated here, the NPI is the only one that has been calibrated using a large cohort of several thousand samples,[27] while all other classifiers were calibrated on the data set of this paper, consisting of 85 samples. The rather big difference in performance of the NPI (ROC area = 0.79) and the classifier based on the three NPI parameters, but calibrated using ANNs on the current data set (ROC area = 0.74), illustrates the importance of large sample cohorts.

The apparent need for large sample cohorts when using gene expression analysis may be explained by the heterogeneity of breast cancer, with many subpopulations. Among clinical variables, some markers (e.g. ER status) mainly distinguish disease subtypes which correlate to outcome, while other markers (e.g. tumour size) may correlate more directly to the progression of the disease. The huge amount of information embedded in genome-wide studies should, in principle, allow for extraction of both kind of markers in gene expression data, but it is not inconceivable that genome-wide profiling is more related to disease subtypes [16,42] than to progression. If so, gene expression analysis may be better suited for studies aiming at an improved biological insight into the mechanisms behind the studied disease and its subtypes, potentially leading to the discovery of new drug targets and development of new therapeutic protocols. A possible way to improve gene expression analysis (both for direct marker design and for gain of biological insight) is to interpret microarray data not in terms of individual genes, but in a way closer related to the underlying biology, e.g. pathways.[43]

As an initial step in exploiting prior knowledge, we used literature genes and a gene list from a differently selected cohort of breast cancers (van 't Veer). Also, we interpreted the results in terms of gene ontology categories and found some categories in common for the different gene lists. When studying the gene ontology of the three different top 100 lists (unrestricted, drug, and van 't Veer), mitosis, cytokinesis, and regulation of cell cycle existed on all lists. Since all lists are created for use of predicting recurrences/drug resistance this indicates that these well-known tumour genesis processes are also important for recurring tumours. Worth mentioning is that the top 100 drug genes and van 't Veer genes have more processes in common, in comparison to the unrestricted genes.

The design of our study, involving only homogeneously treated premenopausal lymph node positive patients, helps focus on a well-defined medical question, but also implies that the recurrence-free group consists of two subgroups, one with an inherited good prognosis (already being cured by the primary operation and postoperative radiotherapy) and one subgroup with inherited bad prognosis, but also CMF-sensitivity (which without adjuvant CMF would have developed recurrence). The group having developed recurrences may be more homogeneous (inherited bad prognosis and CMF-resistant), but heterogeneity may still be a problem, since drug resistance in many cases is acquired, i.e. changes in gene expression are developed after the administration of the drug. Our study has thus only tried to identify those patients recurring in spite of adjuvant CMF, and for which alternative treatments should be recommended. The design of our study makes it impossible to answer which patients do not need adjuvant systemic therapy and which patients benefit from adjuvant CMF. CMF has nowadays, to a large extent, been replaced by other and more effective cytostatic treatments, e.g. anthracycline or taxane based regimes, but two out of the three drugs included in CMF, cyclophosphamide and 5-fluorouracil, are also included in many anthracycline based regimes. The reasons for including patients treated with CMF in the present study were to obtain a long follow up time and enough cases with frozen tumour tissue available. We furthermore hypothesise that the concept to test gene expression profile as a prognostic marker after adjuvant CMF could be generalised to other cytostatic regimes.

It should be emphasised that we have not pursued a survival analysis, since as discussed above, the objective was to construct a classifier for somewhat extreme cases. In part this choice of procedure was dictated by the limited data set at our disposal for this question of CMF resistance. Our comparisons of different classifiers are not very sensitive to excluding the patients that lacked follow-up to the time threshold.

In conclusion, we have confirmed the strength of conventional markers compared to gene expression profilers for prognostic considerations, shown by similar performance in predicting clinical outcome after adjuvant cytostatic (CMF) therapy. We have also stressed important issues when interpreting gene expression data, including gene selection, overtraining, and study design.

Conflict of interest statement

None declared.

Acknowledgement

Appendix A.   Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ejca.2006.06.031.

REFERENCES

1. Early Breast Cancer Trialists' Collaborative Group. Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *Lancet* 2005;**365**:1687–717.
2. Clark JL, Berger SH, Mittelman A, et al. Thymidylate synthase gene amplification in a colon tumor resistant to fluoropyrimidine chemotherapy. *Cancer Treat Rep* 1987;**71**:261–5.
3. Romain S, Martin PM, Klijn JG, et al. DNA-synthesis enzyme activity: a biological tool useful for predicting anti-metabolic drug sensitivity in breast cancer? *Int J Cancer* 1997;**74**:156–61.
4. Washtien WL. Increased levels of thymidylate synthetase in cells exposed to 5-fluorouracil. *Mol Pharmacol* 1984;**25**:171–7.
5. Di Leo A, Chan S, Paesmans M, et al. HER-2/neu as a predictive marker in a population of advanced breast cancer patients randomly treated either with single-agent doxorubicin or single-agent docetaxel. *Breast Cancer Res Treat* 2004;**86**:197–206.
6. Konecny GE, Thomssen C, Luck HJ, et al. Her-2/neu gene amplification and response to paclitaxel in patients with metastatic breast cancer. *J Natl Cancer Inst* 2004;**96**:1141–51.
7. Muss HB, Thor AD, Berry DA, et al. c-erbB-2 expression and response to adjuvant therapy in women with node-positive early breast cancer. *N Engl J Med* 1994;**330**:1260–6.
8. el-Deiry WS. Role of oncogenes in resistance and killing by cancer therapeutic agents. *Curr Opin Oncol* 1997;**9**:79–87.
9. Geisler S, Borresen-Dale AL, Johnsen H, et al. TP53 gene mutations predict the response to neoadjuvant treatment with 5-fluorouracil and mitomycin in locally advanced breast cancer. *Clin Cancer Res* 2003;**9**:5582–8.
10. MacGrogan G, Mauriac L, Durand M, et al. Primary chemotherapy in breast invasive carcinoma: predictive value of the immunohistochemical detection of hormonal receptors, p53, c-erbB-2, MiB1, pS2 and GST pi. *Br J Cancer* 1996;**74**:1458–65.
11. Maxwell PJ, Longley DB, Latif T, et al. Identification of 5-fluorouracil-inducible target genes using cDNA microarray profiling. *Cancer Res* 2003;**63**:4602–6.
12. Burger H, Foekens JA, Look MP, et al. RNA expression of breast cancer resistance protein, lung resistance-related protein, multidrug resistance-associated proteins 1 and 2, and multidrug resistance gene 1 in breast cancer: correlation with chemotherapeutic response. *Clin Cancer Res* 2003;**9**:827–36.
13. Fazeny-Dorner B, Piribauer M, Wenzel C, et al. Cytogenetic and comparative genomic hybridization findings in four cases of breast cancer after neoadjuvant chemotherapy. *Cancer Genet Cytogenet* 2003;**146**:161–6.
14. Nooter K, Brutel de la Riviere G, Look MP, et al. The prognostic significance of expression of the multidrug resistance-associated protein (MRP) in primary breast cancer. *Br J Cancer* 1997;**76**:486–93.
15. Hedenfalk I, Duggan D, Chen Y, et al. Gene-expression profiles in hereditary breast cancer. *N Engl J Med* 2001;**344**:539–48.
16. Gruvberger S, Ringner M, Chen Y, et al. Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns. *Cancer Res* 2001;**61**:5979–84.
17. Sotiriou C, Neo SY, McShane LM, et al. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proc Natl Acad Sci U S A* 2003;**100**:10393–8.
18. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 2004;**351**:2817–26.
19. van de Vijver MJ, He YD, van't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 2002;**347**:1999–2009.
20. van 't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002;**415**:530–6.
21. Wang Y, Klijn JG, Zhang Y, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* 2005;**365**:671–9.
22. Kihara C, Tsunoda T, Tanaka T, et al. Prediction of sensitivity of esophageal tumors to adjuvant chemotherapy by cDNA microarray analysis of gene-expression profiles. *Cancer Res* 2001;**61**:6474–9.
23. Kudoh K, Ramanna M, Ravatn R, et al. Monitoring the expression profiles of doxorubicin-induced and doxorubicin-resistant cancer cells by cDNA microarray. *Cancer Res* 2000;**60**:4161–6.
24. 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* 2003;**362**:362–9.
25. Eden P, Ritz C, Rose C, et al. "Good Old" clinical markers have similar power in breast cancer prognosis as microarray gene expression profilers. *Eur J Cancer* 2004;**40**:1837–41.
26. Elston CW, Ellis IO. 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* 1991;**19**:403–10.
27. Blamey RW, Davies CJ, Elston CW, et al. Prognostic factors in breast cancer – the formation of a prognostic index. *Clin Oncol* 1979;**5**:227–36.
28. Sigurdsson H, Baldetorp B, Borg A, et al. Flow cytometry in primary breast cancer: improving the prognostic value of the fraction of cells in the S-phase by optimal categorisation of cut-off levels. *Br J Cancer* 1990;**62**:786–90.
29. Andersson A, Eden P, Lindgren D, et al. Gene expression profiling of leukemic cell lines reveals conserved molecular signatures among subtypes with specific genetic aberrations. *Leukemia* 2005;**19**:1042–50.
30. Saal LH, Troein C, Vallon-Christersson J, et al. BioArray Software Environment (BASE): a platform for comprehensive management and analysis of microarray data. *Genome Biol* 2002;**3**. SOFTWARE0003.
31. Yang YH, Dudoit S, Luu P, et al. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res* 2002;**30**:e15.
32. Fernebro J, Francis P, Eden P, et al. Gene expression profiles relate to SS18/SSX fusion type in synovial sarcoma. *Int J Cancer* 2006;**118**:1165–72.
33. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 1995;**57**:289–300.
34. Gruvberger-Saal SK, Eden P, Ringner M, et al. Predicting continuous values of prognostic markers in breast cancer from microarray gene expression profiles. *Mol Cancer Ther* 2004;**3**:161–8.
35. Khan J, Wei JS, Ringner M, et al. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat Med* 2001;**7**:673–9.

36. Michiels S, Koscielny S, Hill C. Prediction of cancer outcome with microarrays: a multiple random validation strategy. *Lancet* 2005;**365**:488–92.

37. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;**143**:29–36.

38. Pepe MS, Janes H, Longton G, et al. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *Am J Epidemiol* 2004;**159**:882–90.

39. Ringner M, Veerla S, Andersson S, et al. ACID: a database for microarray clone information. *Bioinformatics* 2004;**20**:2305–6.

40. Zhao SC, Banerjee D, Mineishi S, et al. Post-transplant methotrexate administration leads to improved curability of mice bearing a mammary tumor transplanted with marrow transduced with a mutant human dihydrofolate reductase cDNA. *Hum Gene Ther* 1997;**8**:903–9.

41. Nakamura M, Tsuji N, Asanuma K, et al. Survivin as a predictor of cis-diamminedichloroplatinum sensitivity in gastric cancer patients. *Cancer Sci* 2004;**95**:44–51.

42. Zhao H, Langerod A, Ji Y, et al. Different gene expression patterns in invasive lobular and ductal carcinomas of the breast. *Mol Biol Cell* 2004;**15**:2523–36.

43. Breslin T, Krogh M, Peterson C, et al. Signal transduction pathway profiling of individual tumor samples. *BMC Bioinformatics* 2005;**6**:163.