Prognostic and treatment predictive factors for radio- and chemotherapy resistance in breast cancer patients- a step towards personlized medicine

Niméus, Emma

2009

LUND UNIVERSITY
Faculty of Medicine

ISSN 1652-8220
ISBN 978-91-86253-60-8

Lund University, Faculty of Medicine Doctoral Dissertation Series 2009:72

Prognostic and treatment predictive factors for radio- and chemotherapy resistance in breast cancer patients

Emma Niméus-Malmström

Prognostic and treatment predictive factors for radio- and chemotherapy resistance in breast cancer patients

— a step towards personalized medicine

Emma Niméus-Malmström

Department of Oncology
Lund University
Sweden

Prognostic and treatment predictive factors for radio- and chemotherapy resistance in breast cancer patients

— a step towards personalized medicine

Emma Niméus-Malmström, MD

LUND UNIVERSITY
Faculty of Medicine

Department of Oncology,
Clinical Sciences,
Lund University, Sweden

LAYOUT
Johan Albertén

COVER PAGE DRAWING BY
Alexander Chubar

PRINT
Media-Tryck, Lund 2009

ISBN
978-91-86253-60-8

ISSN
1652-8220

Prognostic and treatment predictive factors for radio- and chemotherapy "resistance" in breast cancer patients

— a step towards personalized medicine

Emma Niméus-Malmström, MD

Department of Oncology,
Clinical Sciences,
Lund University, Sweden

Doctoral Dissertation
By due permission of the Faculty of Medicine,
Lund University, Sweden,
to be publicly defended in the lecture hall, Oncology Department,
Lund University Hospital, Lund at 9.00 am,
Friday September 18th, 2009

Faculty Opponent
Professor Ingela Turesson,
Department of Oncology, Radiology, and Clinical Immunology,
Uppsala University, Uppsala, Sweden

Supervisor
Professor Mårten Fernö,
Department of Oncology, Clinical Sciences,
Lund University, Lund Sweden

Co-supervisors
Professor Carsten Peterson,
Department of Theoretical Physics,
Lund University, Lund, Sweden

Associate professor Per Malmström,
Department of Oncology, Clinical Sciences,
Lund University, Lund Sweden

Organization	Document name
LUND UNIVERSITY Clinical Sciences, Department of Oncology Lund University SE-22185 Lund, Sweden	DOCTORAL DISSERTATION
	Date of issue September 18th, 2009
	Sponsoring organization
Author(s) Emma Niméus-Malmström	

Title and subtitle: Prognostic and treatment predictive factors for radio- and chemotherapy "resistance" in breast cancer patients - a step towards personlized medicine

Abstract

Breast cancer is the most common cancer form among women in the Western world. Although treatment has improved during the last decades, there is still a significant proportion of the patients who are not cured. To further improve clinical outcome we need new treatment strategies, new prognostic markers, and new treatment predictive factors for personalized medicine.
In this dissertation I have focused on local relapse and distant recurrences, where the former is a tumor recurring in the same breast and the latter is distant spread in the body. I have focused on high through-put techniques, but have also evaluated one single factor, using immunohistochemistry.
The most promising result is the gene expression profile for "radioresistance" found in a patient cohort consisting of 100 lymph node negative patients operated with breast conservation surgery and either postoperative radiotherapy or not. The samples were analyzed with oligonucleotide array. A gene expression profile was found that clearly separated patients who developed local recurrences despite radiotherapy ("radioresistance") from patients without local recurrences (either with or without radiotherapy). The clinical consequence, if these results can be confirmed, would be that patients with a "radioresistant" gene profile should be offered mastectomy instead of breast conservation surgery and radiotherapy.
In another patient cohort consisting of 85 node positive breast cancer patients treated with CMF (cyclophosphamide, methotrexate, and 5-fluorouracil) we found a gene expression profile, using cDNA microarray, capable of distinguishing patients who developed distant recurrences from non-recurring patients. This profile was compared to a previously published gene list and to drug-associated genes from the literature. Our results were slightly better. However, our gene profile was not able to exceed the performance of conventional clinical markers. From the same patient cohort, we developed a protocol for protein extraction from the same samples used for RNA. We analyzed the protein expression pattern using 2-DE (two-dimensional electrophoresis) and found several differentially expressed proteins, both when comparing distant recurrences to no recurrences and estrogen receptor positive to estrogen receptor negative tumors. Similarities of regulated genes and proteins were also found when comparing the two studies.
Finally, we investigated the prognostic importance of a proliferation marker, cyclin B1, in a case-control study. There were 190 lymph node negative breast cancer patients with no chemotherapy who died from breast cancer and 190 corresponding controls who were alive at the corresponding case's time of death. Cyclin B1 was an independent prognostic proliferation marker and had a high reproducibility. The marker may be useful instead of histological grade, or as a complement, to identify patients in need of adjuvant chemotherapy.
In conclusion, breast cancer is a heterogeneous disease and should be subdivided even more than it is today into different risk groups with the aid of new markers. We used different high through-put techniques to analyze hundreds to thousands of gene expressions and proteins. Our aim was to find new prognostic and predictive gene expressions and protein profiles that may improve individual treatment schemes and help provide personalized medicine. However, so far it is too early to conclude that the use of single markers can be eliminated, because we also found significant prognostic value of the proliferation marker cyclin B1.

Key words: breast cancer, prognosis, treatment prediction, radioresistance, chemotherapy resistance, cyclin B1, gene expression, 2-DE, immunohistochemistry

Classification system and/or index termes (if any):

Supplementary bibliographical information:	Language English
ISSN and key title: 1652-8220	ISBN 978-91-86253-60-8
Recipient's notes	Number of pages / Price
	Security classification

Distribution by (name and address)

I, the undersigned, being the copyright owner of the abstract of the above-mentioned dissertation, hereby grant to all reference sources permission to publish and disseminate the abstract of the above-mentioned dissertation.

Signature E Niméus-Malmström Date 13/8 2009

To all patients, Oncology Ward 86 in Lund
-my heroes

Contents

Included papers

1. *Emma Niméus-Malmström**, Cecilia Ritz*, Patrik Edén, Anders Johnsson, Carina Strand, Görel Östberg, Mårten Fernö, Carsten Petersson. Gene expression profilers and conventional clinical markers to predict distant recurrences for premenopausal breast cancer patients after adjuvant chemotherapy (CMF). Eur J Cancer. 2006 Nov;42 (16): 2729-37.

2. *Emma Niméus*, Johan Malmström, Anders Johnsson, György Marko-Varga and Mårten Fernö. Proteomic analysis identifies candidate proteins associated with distant recurrences in breast cancer after adjuvant chemotherapy. J Pharm Biomed Anal 2007 Feb 19;43 (3): 1086-93.

3. *Emma Niméus-Malmström**, Morten Krogh*, Per Malmström, Carina Strand, Irma Fredriksson, Per Karlsson, Bo Nordenskjöld, Olle Stål, Görel Östberg, Carsten Peterson, Mårten Fernö. Gene expression profiling in primary breast cancer distinguishes patients developing local recurrence after breast conservation surgery, with or without postoperative radiotherapy. Breast Cancer Res 2008; 10(2):R34

4. *Emma Niméus-Malmström*, Anthoula Koliadi, Cecilia Ahlin, Marit Holmqvist, Lars Holmberg, Rose-Marie Amini, Karin Jirström, Fredrik Wärnberg, Carl Blomqvist Mårten Fernö, Marie-Louise Fjällskog. Cyclin B1 in an independent prognostic proliferation marker with a high reproducibility in lymph node negative breast cancer. (Submitted manuscript)

*contributed equally

Not included paper

Emma Niméus, Bo Baldetorp, Pär-Ola Bendahl, Karin Rennstam, Johan Wennerberg, Jan Åkervall and Mårten Fernö. Amplification of the cyclin D1 gene is associated with tumor subsite, non-diploidy and high S-phase fraction in squamous cell carcinoma of the head and neck. Oral Oncol. 2004 Jul;40(6):624-9

Abbreviations

ANN	artificial neural network
ASCO	American Society of Clinical Oncology
ATLAS	Adjuvant Tamoxifen Longer Against Shorter
aTTom	adjuvant tamoxifen treatment-offer more?
BCS	breast conservation surgery
cDNA	complementary DNA
CI	confidence intervall
CGH	comparative genomic hybridization
CMF	cyclophosphamide, methotrexate, and 5-fluorouracil
DDFS	distant disease free survival
2-DE	two-dimensional electrophoresis
DFS	disease free survival
EBCTCG	early breast cancer trialist's collaborative group
EGFR	epidermal growth factor receptor
EUSOMA	European Society of Breast cancer Specialists
ER	estrogen receptor
FAC	5-fluorouracil, anthracyclin, cyclophosphamide
FDA	Food and Drug Administration
FEC	5-fluorouracil, epirubicin, cyclophosphamide
HERs	human epidermal growth factors
HER2	human epidermal growth factor receptor 2
IGF-1R	insulin-like growth factor-1 receptor
IHC	immunohistochemistry
LHRH	luteinizing hormone releasing hormone
LR	local recurrence
mRNA	messenger RNA
NIH	National Institutes of Health
NPI	Nottingham prognostic index
OR	odds ratio
PAI-1	plasminogen activator inhibitor type 1
PCA	principal component analysis
PgR	progesterone receptor
RIN	RNA integrity number
ROC	receiver operating curve
RFS	recurrence free survival
RT	radiotherapy
SDS PAGE	sodium dodecyl sulfate polyacrylamide gel eletrophoresis
SOFT	Suppression of Ovarian Function Trial
SRM	single reaction monitoring
SVM	support vector machine
TNM	tumor size, lymph node status and metastasis
TMA	tissue microarray
uPA	urokinase-type plasminogen activator
VEGF	vascular endothelial growth factor

Abstract

Breast cancer is the most common cancer form among women in the Western world. Although treatment has improved during the last decades, there is still a significant proportion of the patients who are not cured. To further improve clinical outcome we need new treatment strategies, new prognostic markers, and new treatment predictive factors for personalized medicine.

In this dissertation I have focused on local relapse and distant recurrences, where the former is a tumor recurring in the same breast and the latter is distant spread in the body. I have focused on high through-put techniques, but have also evaluated one single factor, using immunohistochemistry.

The most promising result is the gene expression profile for "radioresistance" found in a patient cohort consisting of 100 lymph node negative patients operated with breast conservation surgery and either postoperative radiotherapy or not. The samples were analyzed with oligonucleotide array. A gene expression profile was found that clearly separated patients who developed local recurrences despite radiotherapy ("radioresistance") from patients without local recurrences (either with or without radiotherapy). The clinical consequence, if these results can be confirmed, would be that patients with a "radioresistant" gene profile should be offered mastectomy instead of breast conservation surgery and radiotherapy.

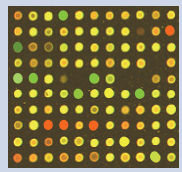
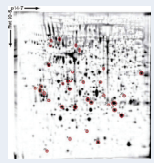
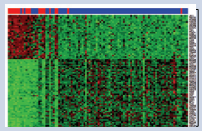
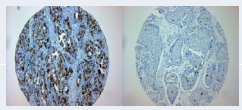
In another patient cohort consisting of 85 node positive breast cancer patients treated with CMF (cyclophosphamide, methotrexate, and 5-fluorouracil) we found a gene expression profile, using cDNA microarray, capable of distinguishing patients who developed distant recurrences from non-recurring patients. This profile was compared to a previously published gene list and to drug-associated genes from the literature. Our results were slightly better. However, our gene profile was not able to exceed the performance of conventional clinical markers.

From the same patient cohort, we developed a protocol for protein extraction from the same samples used for RNA. We analyzed the protein expression pattern using 2-DE (two-dimensional electrophoresis) and found several differentially expressed proteins, both when comparing distant recurrences to no recurrences and estrogen receptor positive to estrogen receptor negative tumors. Similarities of regulated genes and proteins were also found when comparing the two studies.

Finally, we investigated the prognostic importance of a proliferation marker, cyclin B1, in a case-control study. There were 190 lymph node negative breast cancer patients with no chemotherapy who died from breast cancer and 190 corresponding controls who were alive at the corresponding case's time of death. Cyclin B1 was an independent prognostic proliferation marker and had a high reproducibility. The marker may be useful instead of histological grade, or as a complement, to identify patients in need of adjuvant chemotherapy.

In conclusion, breast cancer is a heterogeneous disease and should be subdivided even more than it is today into different risk groups with the aid of new markers. We used different high through-put techniques to analyze hundreds to thousands of gene expressions and proteins. Our aim was to find new prognostic and predictive gene expressions and protein profiles that may improve individual treatment schemes and help provide personalized medicine. However, so far it is too early to conclude that the use of single markers can be eliminated, because we also found significant prognostic value of the proliferation marker cyclin B1.

Dissertation at a glance

	Question	Patients and Method	Result		Conclusion
I	Is it possible to find a gene expression profile for breast cancer patients who develop distant recurrences after adjuvant chemotherapy ("chemotherapy resistance")?	85 patients after adjuvant chemotherapy, with/without distant recurrence. cDNA microarray.	A gene expression profile distinguished the two groups of patients, but it was not superior to conventional clinical markers.		A gene expression profile identified recurring patients after adjuvant CMF ("chemotherapy resistance").
II	Is it possible to find specific proteins for breast cancer patients who develop distant recurrence after adjuvant chemotherapy ("chemotherapy resistance")?	20 patients after adjuvant chemotherapy with/without distant recurrences both ER+ and ER-. 2-D gel electrophoresis.	Several proteins distinguished recurrences from no recurrence and also ER+ from ER- tumors.		Proteins involved in "chemotherapy resistance" and ER associated proteins were found.
III	Is it possible to find a gene profile for patients who develop a local recurrence despite radiotherapy ("radioresistance")?	100 patients with/without local recurrences after breast conservation surgery with/without radiotherapy. Oligonucleotide array.	A gene expression profile identified patients with local recurrences despite radiotherapy, and also patients with no capacity to develop local recurrences		We found a very promising gene expression profile associated with "radioresistance".
IV	Is the proliferation marker, cyclin B1, a prognostic marker for breast cancer death in chemotherapy naïve breast cancer?	A case-control study, 380 patients with no chemotherapy. Cyclin B1 analysis with Immuno-histochemistry.	Cyclin B1 was an independent prognostic factor, adjusted for age, tumor size and endocrine therapy, with a good/very good reproducibility.		Cyclin B1 could be used as a marker for decision whether adjuvant chemotherapy should be given or not.

Populärvetenskaplig sammanfattning

Bröstcancer drabbar var tionde kvinna i Sverige. Patienterna opereras idag antingen med mastektomi, dvs hela bröstet opereras bort, men vanligare är bröstbevarande kirurgi. Efter bröstbevarande kirurgi får patienterna strålbehandling för att minska risken för återfall i samma bröst, sk lokalrecidiv. Majoriteten av patienterna får också medicinsk tilläggsbehandling för att förhindra spridning. Med tilläggsbehandling menas all typ av behandling efter operation och strålbehandling. Exempel på tilläggsbehandling är cytostatika/cellgifter, anti-hormonell och/eller antikroppsbaserad behandling. Dessa kan ges enskilt eller i kombination, bl.a. beroende på tumörens egenskaper. Två tredjedelar av patienterna är botade efter operation. Detta innebär att bara en tredjedel behöver adjuvant behandling – en behandling som emellertid ges till över 80%. Många patienter blir alltså överbehandlade med onödiga biverkningar och kostnader för samhället som följd. De urvalskriterierna, som används idag för att dela in patienterna i olika riskgrupper för återfall, behöver förbättras. Kanske behövs det nya och mer kraftfulla tekniker för att identifiera olika riskgrupper och grupper med känslighet för olika behandlingar. För närvarande används följande patient- och tumör-relaterade faktorer i klinisk rutin: ålder, tumörstorlek, antal sjuka lymfkörltar, histologisk gradering (mikroskopisk undersökning), östrogen och progesteron receptor uttryck och HER2 (uttryck av en specifik gen).

Om bröstcancer sjukdomen återkommer/recidiverar ändras prognosen drastiskt. Genuttryck har visat sig ha stor betydelse inom bröstcancer och därför har jag valt att studera många genuttryck samtidigt för att söka specifika mönster. Varje aktiverad gen kan ses som en mall för proteinnybildning, varför även proteinmönster har studerats.
En orsak till lokalrecidiv trots postoperativ strålbehandling kan vara någon form av strålbehandlingsresistens. I studie III har jag hittat ett genmönster som kan identifiera dessa tumörer. Om fynden kan bekräftas bör patienter med detta uttrycksmönster istället opereras med mastektomi från början och skulle då sannolikt undvika lokalt återfall.

Jag har också jämfört patienter som trots cytostatikabehandling utvecklar fjärrmetastaser med patienter, som fått samma behandling men inte utvecklat fjärrmetastaser. I studie I och II identifierade vi genuttryck och proteiner som kunde särskilja de två grupperna. Vid en jämförelse med traditionella kliniska markörer var genuttrycksprofilen emellertid inte bättre. Orsakerna till detta kan vara många. Brösttumörer är så olika att patientgruppen kanske måste vara mycket större för att finna ett gen/proteinmönster som särskiljer. Även undergruppering av patienter vad gäller östrogen receptor uttryck och studie upplägget kan vara orsaker som haft betydelse.

Eftersom fjärrmetastaser har så stor betydelse för prognosen för bröstcancer så skulle naturligtvis ett test vara optimalt som kunde förutsäga vilka patienter som inte kommer att utveckla metastaser. Dessa patienter skulle då slippa onödig behandling samt relaterade biverkningar. Därför studerade jag i studie IV en tillväxtmarkör cyklin B1 som är delaktig i celldelningen. Ökad förekomst av cyklin B1 visade sig vara korrelerad till död i bröst cancer och skulle kunna användas som ett komplement till dagens markörer.

Background

Breast cancer is the most common malignancy among women in the Western world, and it affects approximately every tenth woman. In the past decades there have been considerable improvements in diagnosis, development of new drugs, and prediction of prognosis and treatment effects, all of which have enhanced survival. Annually in Sweden more than 7 000 women develop breast cancer, and despite improved treatment, approximately 1 500 women die each year from this disease. In the latest meta-analysis from Early Breast Cancer Trialists' Collaborative Group (EBCTCG) from 2000 (published 2005), comprising 194 randomized trials of adjuvant chemotherapy or endocrine therapy worldwide, it was shown that adjuvant systemic treatment significantly reduces the rate of recurrence and mortality[1]. Today patients with primary breast cancer are operated with either breast conservation surgery or mastectomy with the exception of those with metastasized disease. Depending on the type of surgery and tumor/patient characteristics, the patients are treated with postoperative radiotherapy and/or systemic adjuvant treatment. There are several treatment regimes: endocrine, cytostatic, antibody-based, or different combinations. The choice of regime is governed by many prognostic and treatment predictive factors: tumor size, lymph node involvement, histological grade, estrogen- (ER) and progesterone receptor (PgR) status, extensive peritumoral vascular invasion, age, Ki67, and human epidermal growth factor receptor 2 (HER2)[2]. There are promising new factors, such as uPA (urokinase-type plasminogen activator) and PAI-1 (plasminogen activator inhibitor type 1)[3], which have shown grade 1 evidence. Mutations in p53 and topoisomerase IIα are under evaluation, but are not generally accepted markers for clinical use today[2, 4].

Although prognostic and treatment predictive markers have improved treatment decisions, not every patient receives sufficiently effective therapy for a cure. A considerable proportion of the patients are cured by the primary surgery, and they are consequently overtreated when given adjuvant therapy. Breast cancer is a complex heterogeneous disease suggesting that further subdivision of the patients is needed, perhaps down to the individual level. Ideally, the risk of recurrence for the individual patient or sensitivity/resistance to different treatment regimes should be accurately determined. Then a tailor-made treatment plan should be designed for each individual's unique genetic/proteome makeup, so called personalized medicine. For this purpose we probably need to use genetic and proteome information on a wider basis. We will have to analyze several patterns of genes, gene expressions, and proteins rather than single genes or gene products. It is likely that subtle changes over the entire integrated network of genetic and proteome information will be of importance.

During the last decade a large number of studies using gene expression analysis have been published[5]. Gene expression analyses have addressed the clinically important questions such as the identification of patients with a high risk of distant[6-10] and local recurrences[11, 12], and treatment sensitivity[13]. One clear subdivision of breast cancer is based on ER status. ER positive and ER negative tumors have remarkably different gene expression profiles[14]. It has also been shown that breast cancer can be subdivided into different subgroups, e.g. by the use of the "molecular portrait". This classifies the tumors into five subtypes: basal-like, Erb-B2+, normal-like, and luminal-like subtypes A and B[15, 16].

The most common proteomic method when studying breast cancer is two-dimensional electrophoresis (2-DE). This method has been used to describe differences between ductal carcinoma and non-neoplastic tissue[17] and to identify proteins associated with c-erbB-2 (HER2)-overexpression[18]. However, none of the proteins identified with 2-DE or with other proteomic techniques hitherto have been found to be useful for the clinical management of breast cancer patients[4, 19, 20].

In this dissertation, the main focus is on the discovery of new markers, with special focus on high through-put technologies, and to compare their clinical value in relation to established conventional markers. In the following sections, topics related to recurrences and the current status of the markers used clinically, treatment of breast cancer, and different techniques, will be described.

Recurrences

Since recurrence of breast cancer decreases the survival, it is absolutely essential to prevent loco-regional and distant recurrences by selection of appropriate treatment. Historically, there are three theories or paradigms for the progression of breast cancer. The "Halstedian" theory proposes that breast cancer is a local disease with a contiguous spread through the lymphatics dictating local control. In contrast, the "systemic" view proposes that breast cancer is a systemic disease with two distinct subgroups, those with the ability to metastasize to distant sites and those that lack this ability. That theory dictates systemic therapy and predicts that the local control will have little effect on overall survival. More recently, a third hypothesis has been developed, since neither of the prior hypotheses is valid for all tumors. This newer theory proposes that there is a time interval during which tumor cells have not metastasized to distant sites from the primary tumor. At the time of diagnosis it is not known whether this interval has passed. Therefore the greater the likelihood of systemic spread at the time of diagnosis, the lower is the likelihood that local therapy will influence the patient's survival[21].

In the following section the definitions, the incidences in relation to treatment, and the prognosis of the patients in relation to the different recurrence types are summarized. Breast cancer recurrences are divided into three groups: local, regional, and distant. Local recurrences appear as a recurrence in the same breast, regional recurrences appear in the regional lymph nodes (axilla, supra-, infraclavicular fossa and parasternal lymph nodes), and distant recurrences are those that occur elsewhere in the body.

Local recurrence

A local recurrence is defined as a reappearance of the same cancer in the ipsilateral breast, chest wall or overlying skin or scar after initial therapy. Most local recurrences are found by the clinician, the patient herself, or mammography. Most local recurrences occur at a fairly constant rate during the first decade after breast conservation surgery, whereas most local recurrences after mastectomy occur within 4 years[21]. The incidence of a local recurrence is also dependent on the surgical method; there is a higher risk of local recurrences after breast conservation surgery than after mastectomy[22]. With the addition of radiotherapy after breast conservation surgery, the incidence for a local recurrence is reduced by about two thirds[23]. Today breast conservation surgery with radiotherapy is considered an established and safe treatment option for the majority of patients.

The presence of a local recurrence implies an increased risk of both distant metastases and mortality[21, 24, 25]. For every four local recurrences avoided with radiotherapy during the first five years after the primary operation, one additional breast cancer patient is alive after 15 years[21, 23].

Besides tumor-involved margins, the generally accepted risk factors for the development of a local recurrence are young age and multicentricity[26-29]. For predicting the occurrence of a local recurrence despite radiotherapy ("radioresistance"), there is currently no useful marker.

Regional recurrence

Regional recurrence refers to metastases in the ipsilateral axillary, supraclavicular and infraclavicular, and/or parasternal lymph nodes, where the supraclavicular fossa is the most commonly affected site after adequate surgery. Regional recurrences are only seen in 1-3% of all breast cancer patients after mastectomy or breast conservation surgery with postoperative radiotherapy to the remaining breast[30]. Regional recurrences significantly enhance the risk of distant metastases[30] and consequently indicate a markedly reduced overall survival.

Distant recurrence

Approximately every third breast cancer patient will develop a distant recurrence, and the disease is then considered as non-curable despite current systemic treatment regimes. A distant recurrence in breast cancer appears most often in the lung, liver, or skeleton. Bone is the most common site of the first metastasis in women with hormone receptor-positive tumors, whereas hormone-receptor-negative tumors tend to develop visceral metastases and/or brain metastases[31, 32]. In comparison to other malignant diseases, breast cancer is unusual because it can recur very late. Normally the distant recurrences appear within the first decade but they can also appear after 20 years[1].

To avoid the development of all kinds of recurrences we need reliable clinical and pathological markers to subgroup each patient to make sure that she receives the appropriate treatment.

Factors in clinical routine

Prognostic factors

Prognostic factors are tools to predict the tumor aggressiveness and to measure the risk of relapse and death in the absence of systemic treatment. The aim of a prognostic factor is to identify patients in need of additional treatment and/or those already being cured by the local therapy. Prognostic factors describe the natural history of a malignant tumor, e.g. the risk for spread from a local tumor to the lymph nodes and to distant sites[33].

Stage	Tumor size (T)	Lymph node status (N)	Distant metastasis (M)
Stage 0	Tis	N0	M0
Stage I	T1	N0	M0
Stage IIA	T0	N1	M0
	T1	N1	M0
	T2	N0	M0
Stage IIB	T2	N1	M0
	T3	N0	M0
Stage IIIA	T0	N2	M0
	T1	N2	M0
	T2	N2	M0
	T3	N1-2	M0
Stage IIIB	T4	N0-2	M0
Stage IIIC	Any	N3	M0
Stage IV	Any	Any	M1

Table 1. *Primary tumor size (T):Tis= Carcinoma in situ, T1= Tumor ≤ 20 mm, T2 = Tumor 21-50 mm in greatest dimension, T3= Tumor >50 mm, T4= Tumor of any size extending to chest wall or skin, and inflammatory carcinoma.*
Regional lymph nodes (N):cN0= No regional lymph node metastasis, cN1= Moveable ipsilateral axillary metastasis, cN2= Fixed ipsilateral metastasis or parasternal lymph nodes, cN3= Metastasis in ipsilateral supra- or infraclavicular lymph nodes, or parasternal lymph nodes + ipsilateral axillary lymph nodes. pN1= 1-3 microscopically positive axillary or parasternal lymph nodes, the latter may not be clinically detectable, pN2=4-9 positive axillary lymph nodes, or clinically detectable parasternal lymph node/s without positive axillary lymph nodes, pN3= ≥10 postive axillary lymph nodes, both clinically detectable parasternal metastases and ≥1 axillary lymph nodes or microscopic parasternal metastases and >3 axillary lymph nodes, or ipsilateral lymph node in fossa supraclavicular.
Distant metastasis (M):M0= No distant metastasis, M1= Distant metastasis

From a prognostic point of view it is important to characterize the spread of the tumor, i.e. the stage. The stage of a tumor is strongly correlated to the prognosis, and that influences the choice of treatment. The TNM classification system is well established and is based on the size of the primary tumor, the occurrence of lymph node metastases, and the presence of distant metastases (Table 1). Based on the number of positive lymph nodes, the patients are usually divided into four groups: negative nodes, 1-3 positive nodes, 4-9 positive nodes, and ≥10 positive nodes. The 5-year survival for breast cancer patients worsens with increasing number of positive lymph nodes[34]. The number of tumor-involved lymph nodes in relation to the number of examined nodes has also been shown to be of clinical importance in node-positive patients. A higher ratio implies worse prognosis[35].

Among tumor associated prognostic factors, histological grade is the most commonly applied. It is based on evaluation of tubular formation, nuclear atypia, and the number of mitosis. Each part is scored from 1 to 3. The scores for the three components are summed and categorized as grade 1 (3-5): well differentiated, grade 2 (6-7): moderately differentiated, or grade 3 (8-9): poorly differentiated[36]. Three parameters, histological grade, axillary lymph node status (1=no positive lymph nodes, 2=1-3 positive lymph nodes, 3=>4 positive lymph nodes) and tumor size, are included in a prognostic index called the Nottingham prognostic index (NPI):
0.2x size (cm) + node stage (1-3) + histological grade (1-3)[37, 38].
Age, extensive peritumoral vascular invasion, and Ki67, according to the St Gallen consensus, are also recommended to be used as prognostic factors in clinical routine[2].

ER, PgR and HER2 are both prognostic factors and treatment predictive factors (see below)[2].

Treatment predictive factors

Treatment predictive factors describe how a group of tumors with defined characteristics will respond to a specific treatment[33]. ER and PgR are strong treatment predictive factors and are used to decide whether endocrine treatment will be beneficial or not. For patients with HER2-positive breast cancer, treatment with the antibody trastuzumab is meaningful both in the metastatic and in the adjuvant situation[39].

Not yet accepted factors under evaluation

There is a constant search for new prognostic and treatment predictive factors. With the introduction of new targeted treatments, the need to predict therapeutic efficiency has increased further.

Single factors

uPA and PAI-1 are considered prognostic factors, since high levels of uPA and PAI-1 are associated with poor RFS and poor overall survival[3]. uPA/PAI-1 can also be considered as a treatment predictive factor, because patients with high levels of uPA/PAI-1 derive

substantial benefit from adjuvant CMF chemotherapy[4, 40]. The tumor suppressor gene, p53, and proliferation markers such as topoisomerase IIα, cyclin D and E, are under investigation but are presently not recommended to be used routinely[4]. Cyclin A and B are also under investigation and are not recommended in clinical routine. Useful clinical markers to detect resistance/sensitivity to chemotherapy have not yet been identified, although some may have clinical usefulness, e.g. thymidilate synthase, thymidine kinase[41-43], c-erbB2[44], multidrug resistance-associated protein[45], p53[46-50], topoisomerase IIα[51], and Tau[52].

Ki-67, cyclins, S-phase fraction, and mitotic activity are proliferation markers. S-phase fraction has been shown to be an independent prognostic factor in a prospective study[53], but it is not used in the clinic. Cyclin D and E are under investigation. Much less attention has been paid to cyclin A and B. The cyclins are effective during the cycling course of events of synthesis and degradation during each cell cycle. The G1 cyclins are required for the binding to the cyclin-dependent-kinases and entry into S-phase. Mitotic cyclins are required for the binding to cyclin-dependent-kinases in G2 and are responsible for the entry into mitosis. Cyclin B1 is a mitotic cyclin. The accumulation of cyclin B1 begins in S-phase; it is essentially restricted to the G2-M transition, reaches its maximal level at mitosis, and then is rapidly degraded at the metaphase-anaphase transition[54].

Multiple factors

Recently the Food and Drug Administration (FDA) approved OncotypeDX®, which is based on 21 selected genes, so called recurrence score. This signature is able to distinguish lymph node negative, tamoxifen treated, ER positive patients not in need of adjuvant chemotherapy from those who would benefit from adjuvant chemotherapy[8]. Among the 21 genes are the genes for ER/PgR, HER2 and the proliferation genes, Ki-67 and cyclin B1.

The gene expression based profile, Mammaprint®, consists of 70 genes, and it predicts the outcome of distant recurrences in young (<55 years old), node-negative breast cancer patients[10]. The authors suggest that those patients with a low-risk gene profile do not need adjuvant systemic therapy, whereas a "poor prognosis" signature would indicate treatment with adjuvant chemotherapy. One possible advantage with the 70-gene profile is that the low-risk group is larger than the low-risk group based on conventional markers, e.g. St Gallen or National Institute of Health (NIH) consensus criteria. As a consequence the number of overtreated patients may be reduced.

Studies to further validate the prognostic gene profiles are under investigation; TaylorX is analyzing OncotypeDX®, and MINDACT (Microarray In Node negative Disease may Avoid ChemoTherapy) is analyzing Mammaprint®.

Prognostic and treatment predictive factors are used daily in the clinic[55, 56] to sub-classify the patients and to assign adequate treatment. The primary treatment of breast cancer can be divided into local and adjuvant systemic treatment. Adjuvant systemic treatment regimes include polychemotherapy, endocrine therapy, and/or trastuzumab, and different combinations thereof. The purpose is to delay or inhibit the development of distant recurrences.

Primary local treatment

Treatment regimes for breast cancer vary both nationally and internationally, and the treatment regimes mentioned in this section are mostly from guidelines of the South Sweden Breast Cancer Group (Sydsvenska bröstcancer gruppen).

The primary local treatment includes surgery and radiotherapy, and the goal is to eradicate the local tumor to reduce recurrences. Today, most breast cancer patients are operated with breast conservation surgery or modified radical mastectomy. In common routine, breast conservation surgery is followed by sentinel node biopsy. That is a surgical technique that uses radioisotope and blue dye to identify axillary lymph node metastases. An axillary lymph node dissection is performed if there is a positive sentinel node. Mastectomy is preferred for T3 (>50 mm) and T4 (overgrowth to adjacent chest wall, skin, multicentric, and inflammatory) tumors. Mastectomy can also be chosen when an acceptable cosmetic result cannot be obtained with breast conservation surgery due to the size of the tumor in relation to the breast volume, or due to the preference of the patient. Furthermore, after the development of a local recurrence, mastectomy is indicated in the majority of the patients. Breast conservation surgery has been shown to be the treatment of choice for women with unifocal breast cancers <3-4 cm. However, the frequency of local recurrences is higher in patients treated with breast conservation surgery than with mastectomy[22]. The meta-analysis from EBCTCG shows that the addition of radiotherapy after breast conservation surgery lowers the risk for a local recurrence by two thirds at 15 years[23], equivalent to the risk of developing a local recurrence after mastectomy. No differences in overall or breast cancer death rates are seen when comparing mastectomy to breast conservation surgery with radiotherapy[22, 23, 57].

According to international guidelines (European Society of Breast Cancer Specialists (EUSOMA) and the American Society of Clinical Oncology (ASCO), radiotherapy is indicated when the risk of developing a local recurrence is >20% in 10 years (http://www.swebcg.roc.se)[23]. Clinically, that is equivalent to patients operated with breast conservation therapy or mastectomy with a tumor >50 mm or >3 lymph node metastases.

There is a universal agreement that patients operated with ≥4 positive lymph nodes after breast conservation surgery or mastectomy should be offered loco-regional radiotherapy. The indications for loco-regional radiotherapy in patients with 1-3 lymph nodes have been the focus of an intense debate during the last few years. According to new guidelines in south Sweden, even patients with 2-3 involved nodes should be offered loco-regional radiotherapy. Among patients with only one involved node there are subgroups that may benefit from loco-regional radiotherapy if they have additional risk factors such as vascular invasion, poorly differentiated tumors and/or age < 40 years. However, such patients do not routinely receive radiotherapy to the regional lymph nodes.

In the EBCTCG meta-analysis from 2005, it was shown that radiotherapy also increases the absolute breast cancer specific survival (5.4%) and the overall survival (5.3%) at 15 years[23]. Standard postoperative radiotherapy after breast conservation surgery

is fractionated, and generally consists of 50 Gy (2 Gy x 25) to the remaining breast parenchyma. A booster dose is an additional dose of 16 Gy radiation that is directed to the tumor bed. The booster dose reduces the risk of local recurrences, especially in patients younger than 50 years of age[58]. However, the cosmetic result has been reported to be less satisfactory when using a booster dose[59].
Similar results, as compared with standard fractionation, have been reported for a shorter course of treatment, i.e. 42.5 Gy in 16 fractions[60].

The acute side effects from loco-regional radiotherapy are usually skin reactions and more rarely, ulcerations. Later the skin may become more fibrotic and stiff. Lymphedema of the arm can also be seen as a later side effect of radiotherapy with chronic swelling of the arm. Other later side effects, which are uncommon, are pneumonitis and worsened lung function.

Even though the risk of a local recurrence is reduced and survival is improved with radiotherapy, many patients are overtreated and would have been cured with only surgery. The benefit for the patients with low risk would be doubtful if we were able to identify this subgroup[61] (Fig.1a). On the other hand, some patients develop local recurrences despite radiotherapy, indicating resistance to irradiation (Fig. 1b).

In summary, there are three patient groups that preferably should be identified prior to choice of surgery: (1) those suitable for breast conservation surgery with no capacity to develop local recurrences (LR) and consequently not in need of radiotherapy (RT), (2) those suitable for breast conservation surgery but with the capacity to develop LR, and in

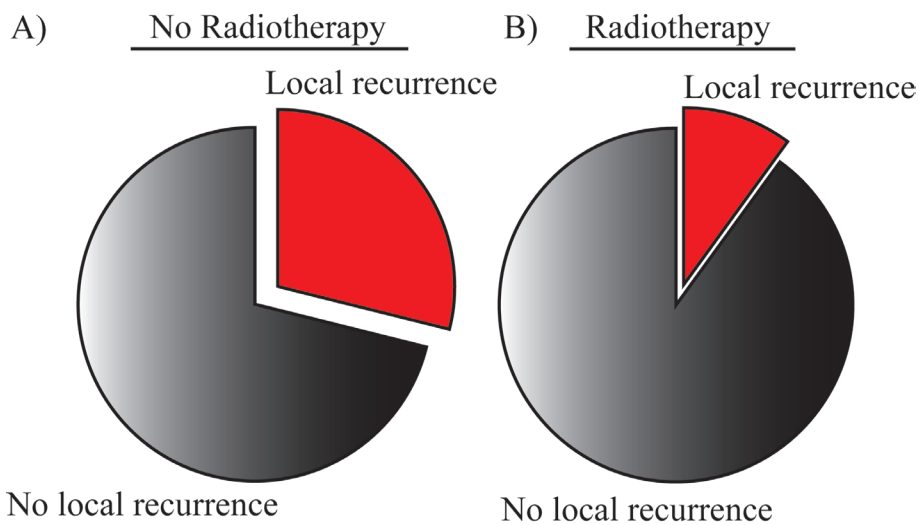


Figure 1. *70% of the patients operated with breast conservation surgery and no additional radiotherapy (RT) will not develop a local recurrence (LR) and do not need RT. 10% of the patients treated with breast conservation surgery and additional RT will develop LR and do need another treatment such as mastectomy or additional adjuvant treatment.*

whom RT will prevent the LR, i.e. "radiosensitive", and (3) those in need of mastectomy, with the capacity to develop LR which are not prevented by RT, i.e. "radioresistant".

"Radioresistance/sensitivity"

Cells in early S-phase and G2-phase are more sensitive to radiotherapy than in other phases. The capacity to repair DNA is inferior or worse in tumor cells as compared with normal cells, which is exploited when designing radiotherapy schedules. One of the many different actions of radiotherapy is to induce double-stranded breaks that prevent mitosis from occurring and that trigger apoptosis. The inability to repair double-stranded breaks and to condense chromatin are the two major reasons for the effect of radiotherapy.

Cells with a proficient double-stranded break-repair mechanism and the capacity to condense the chromatin tend to be less sensitive to radiotherapy[62]. There are several mechanisms that are highly important for the relaxation of chromatin structure and thus increase the accessibility of the damaged sites to the repair machinery. One example is acetylation of histone tails, which induces relaxation of the chromatin structure[62]. Signalling pathways seem to be important for the fate of the irradiated cells. Examples of important pathways are those involving: insulin-like growth factor-1 receptor (IGF-1R), human epidermal growth factors (HERs), vascular endothelial growth factor (VEGF), and genetic variations affecting loss of control over the cell cycle, DNA repair, and apotosis. Hypoxia, which is a major regulator of VEGF, is one important factor for failure of radiotherapy[63].

Adjuvant systemic treatment

The main aim of adjuvant systemic treatment is to remove remaining micro metastases, which are not yet clinically detectable deposits of the disease. This will reduce the recurrence rate and improve long-term survival[1]. Today adjuvant treatment is recommended or indicated for patients with lymph node positive tumors and for most patients with lymph node negative tumors, except to those with a very low risk of recurrence. The advantages with adjuvant therapy should be balanced against the disadvantages, e.g. early and late side effects and costs. Ideally, adjuvant therapy should be given only to patients who will benefit from the treatment. Adjuvant treatment regimes used in the clinical routine today include endocrine therapy, chemotherapy, antibody-based (trastumazab) therapy, and different combinations of these.

Endocrine therapy

In the meta-analysis by EBCTCG, adjuvant tamoxifen treatment was shown to reduce the annual breast cancer death rate by 31% for patients with hormone receptor-positive tumors, irrespective of patient age or use of chemotherapy[1]. The annual recurrence rate was almost halved (recurrence rate ratio 0.59) when 5 years of tamoxifen was compared to no tamoxifen[1]. Five years of tamoxifen also reduced recurrence and mortality as compared

with 2 years[64, 65], which has led to the presently recommended 5-year period of tamoxifen treatment. Longer duration of tamoxifen is being evaluated in the ATLAS (Adjuvant Tamoxifen Longer Against Shorter) and aTTom (adjuvant tamoxifen treatment-offer more?) trials. During the last decade, aromatase inhibitors, letrozole, examestan, and anastrozole have challenged tamoxifen as adjuvant treatment for postmenopausal women with receptor-positive breast cancer. So far, aromatase inhibitors have shown improved recurrence-free survival, but have not shown a difference in mortality as compared with tamoxifen[66]. The present recommendations in Sweden for adjuvant endocrine therapy are as follows: tamoxifen is suggested for the majority of patients, and especially for the lymph node-negative patients. However, patients with a higher risk (lymph node-positive and HER2 positive) are treated with aromatase inhibitors, either alone or in combination with tamoxifen. The "switch strategy" was previously 2-3 years of tamoxifen followed by aromatase inhibitors for the rest of the 5-year period. Since 2009 in Sweden that has been changed to 2-3 years of aromatase inhibitors followed by tamoxifen for the rest of the 5-year period. An aromatase inhibitor is given for five years to patients with ≥4 involved lymph nodes. No benefit from endocrine therapy has been demonstrated in hormone receptor-negative tumors[64].

Ovarian suppression can be used in premenopausal women because it reduces the estrogen level to postmenpausal levels. Three options are available: oophorectomy, ovarian suppression with radiotherapy, and luteinizing hormone releasing hormone (LHRH) analogs.

In premenoupausal women tamoxifen is also effective, but the question regarding the effect of additional ovarian suppression has been raised. A clinical trial, studying the combinations of tamoxifen or aromatase inhibitor with ovarian suppression (surgically, radiotherapy or LHRH analogs), has therefore been initiated (SOFT= Suppression of Ovarian Function Trial).

Chemotherapy

The first polychemotherapy regime established as a standard for adjuvant treatment in women with positive lymph nodes was CMF (cyclophosphamide, methotrexate, and 5-fluorouracil)[1]. In a meta-analysis from EBCTCG it was shown that adjuvant CMF reduces the annual hazard risk of recurrence by 24% and mortality by 14%[67]. As far as 10-year survival is concerned, 7% absolute improvements are seen for patients <50 years with lymph node-negative disease and 11% improvements are seen for lymph node-positive disease. Smaller benefits are shown for patients 50-69 years old (2% for lymph node-negative and 3% for lymph node-positive disease)[67] .

Anthracyclin-based polychemotherapy (FAC or FEC) was later shown more effective than CMF chemotherapy, with a 4% absolute decreased risk for 10-year probabilities of recurrence, for breast cancer mortality, and for overall mortality, as compared with CMF[1]. More recent treatment combinations, with the addition of taxanes to anthracycline-based regimes, have further improved disease-free and overall survival of lymph node-positive breast cancer patients[68].

Antibody-based treatment

HER2 overexpression is reported in 15-30% of primary breast cancer patients[69-73]. Adjuvant trastuzumab combined with adjuvant chemotherapy improves the outcome for women with HER2-positive tumors[39, 74]. Except for endocrine therapy, trastuzumab is so far the only targeted treatment that is approved for adjuvant therapy. Several new targeted treatments have been found to be effective for advanced breast cancer and are currently being tested in randomized trials in the adjuvant setting.

Neoadjuvant treatment

Neoadjuvant treatment is indicated for locally advanced breast cancer or for downstaging before breast conservation surgery.

In Sweden, the incidence of locally advanced breast cancer is low, 5%[75]. Neoadjuvant treatment is recommended as preoperative treatment for locally advanced breast cancer, stage III. The aim of the neoadjuvant treatment is to achieve better surgical conditions to enable mastectomy for more advanced tumors or breast conservation surgery for medium-sized (stage II) tumors. Patients with locally advanced breast cancer are treated with multimodal therapy including preoperative chemotherapy, surgery and/or loco-regional radiotherapy, and additional endocrine treatment and/or trastuzumab is given for ER positive/HER2-positive tumors (http://www.swebcg.roc.se). With multimodal therapy, around 70% of the patients with locally advanced breast cancer will obtain local control[75], and five-year survival figures around 40% have been reported[76]. This figure should be compared with older studies in which 5-year survival without systemic therapy was only 3.5-15%[75].

Palliative treatment

Distant metastases of breast cancer imply a non-curable disease. In this situation palliative treatment can reduce or alleviate symptoms and prolong survival. Tumor resection of breast cancer with distant metastases at diagnosis is not generally indicated, and medical treatment should be based on pathological examination of biopsies from the primary tumor, and if possible, from the dominant metastatic site. The advantages of examining the metastases are to secure the diagnosis and to investigate whether the tumor has preserved its characteristics for ER, PgR, and HER status. The treatment with endocrine and cytotoxic agents is most often the same as used for adjuvant treatment, although there are some cytotoxic drugs, such as vinorelbin and capecitabine, which are only used for palliation.

The goal of palliative treatment is palliation, and quality of life is of great importance. Therefore, some patients are only treated with single drug chemotherapy, even though combination chemotherapy is more effective.

New drugs

Some patients develop recurrences despite different adjuvant treatments. Therefore, new treatment strategies need to be developed. All new drugs are always first evaluated in patients with advanced breast cancer in the palliative setting. If a positive effect is found, the drug is investigated in the adjuvant setting in randomized clinical trials.

Several recently introduced targeted therapies are presently under investigation, e.g. antibodies against HER2, vascular endothelial growth factor (VEGF), and tyrosine kinase inhibitors (i.e. against epidermal growth factor receptor (EGFR), HER2 and multi-tyrosine kinases). HER2-positive tumors that have progressed despite trastuzumab-based therapy, are more effectively treated by lapatinib, a tyrosine kinase inhibitor in combination with capecitabine as compared with monotherapy with only capecitabine[77]. Bevacizumab, an antibody targeting VEGF, in combination with capecitabine or paclitaxel, has shown promising results in metastatic breast cancer and is currently being investigated in the adjuvant setting[111].

Sunitinib is a multitargeted tyrosine kinase inhibitor that is being evaluated in metastatic breast cancer, and has shown promising results, especially with triple negative tumors and HER2-positive, trastuzumab-treated patients[78].

Pertuzumab, a humanized monoclonal antibody against a region of HER2, which blocks heterodimerization and downstream signaling, may have clinical benefits in combination with trastuzumab in patients previously treated with trastuzumab[79].

Fulvestrant is a new endocrine alternative with no agonist effect, i.e. a pure antiestrogen. When complexed with ER, fulvestrant prevents receptor dimerization of ER and prevents binding to ER sites in the nucleus, thus blocking transcription. Fulvestrant has been shown to be as effective as anastrozole in terms of time to progression and time to death in metastatic breast cancer[80].

During the past decades improved local (surgery and radiotherapy) and systemic treatment of primary and metastatic breast cancer have improved the outcome. However, even with the new treatment alternatives, about 1 500 women in Sweden will die annually from this disease. Better and more precise markers would ideally enable more accurate classifications of the patients and thereby increase the proportion of the patients receiving adequate treatment. This could result in more individualized personalized treatment and ultimately improved survival.

Personalized medicine

Even though the use of prognostic and treatment predictive factors and novel treatment regimes has improved prognosis, there is still a great need to further improve the treatment and to find new clinical markers. New methods, such as gene expression analysis, which is used in Mammaprint® and OncotypeDX®, and proteomics techniques, with a high-degree of through-put, may be important for this purpose. The identification of new markers and

improved personalized treatments are the keys for improving health care. The advantage with the new techniques is the possibility to analyze many genes/gene expressions at the same time and to find patterns instead of single genes/gene expressions. Single genes/proteins which are under investigation, such as the proliferation marker cyclin B1, and well known markers such as ER, Ki67, and HER2, are examples of genes found on the OncotypeDX® geneprofile[8]. Note that the geneprofile simultaneously evaluates many factors instead of single parameters. This should enable a more individual-based analysis. This should also enable subdivision of previously large patient groups into smaller and better characterized groups that could benefit from personalized treatment. Looking at the genetic and proteome information in this way may also open up new possibilities for the treatment of other cancer types. A breast cancer patient may have genetic/proteome information corresponding more to some other cancer type, and that other treatment may be useful against her breast cancer.

Established clinical factors are used daily in the clinic[55, 56] to sub-classify patients and assign adequate treatment. Still, a significant proportion of the patients are misclassified and thereby not receiving optimal treatment. The shortcomings of the current prognostic and treatment predictive factors indicate that they are too crude and need to be supplemented with additional information or markers. Better and more precise markers are needed to predict tumors that would not benefit from chemotherapy. Markers are needed, which could better indicate the best type of chemotherapy, the "radioresistant" tumors, and which could more accurately classify the patients. The goal is to increase the proportion of the patients who receive adequate treatment, give more individualized, personalized treatment when appropriate, and to improve survival.

Aims of the thesis

The general aim of this dissertation was to gain more information with regards to chemo- and radiotherapy "resistance" in breast cancer using high through-put techniques, and to find new markers to predict dissemination of breast cancer.
More specifically, the aims were:

- To find a gene expression and a protein profile that are able to distinguish node-positive breast cancer patients developing distant metastases after adjuvant CMF from those not developing distant metastases. The former group ("CMF resistant" subgroup) is a candidate for other adjuvant systemic treatment regimes.
- To find gene expression profiles that identify patients who develop local recurrences despite radiotherapy ("radioresistance" profile) and patients who do not develop local recurrences (not in need of postoperative radiotherapy).
- To verify the prognostic value and reproducibility of the proliferation marker cyclin B1 in a lymph node negative patient cohort without chemotherapy using immunohistochemistry.

Patients and methods

Patients

Sweden has a unique way of handling information from cancer patients: the personal social security number, the Swedish cancer registry, and the cause of death registry. In addition, the very low number of private hospitals in Sweden assures the collection of population based patient data and tumor material.
Every Swedish citizen has a personal social security number making it possible to track every living and deceased person in Sweden. It is requested that all malignant tumors are reported to the Swedish cancer registry. Also, the cause of death is requested to be reported to the cause of death registry by the doctor in charge of the patient. This registry is linked to the population statistics registry.

Since the end of 1970, fresh frozen tumor material has been sent to the Department of Oncology in Lund for analysis of ER and PgR. Remaining tumor material has been

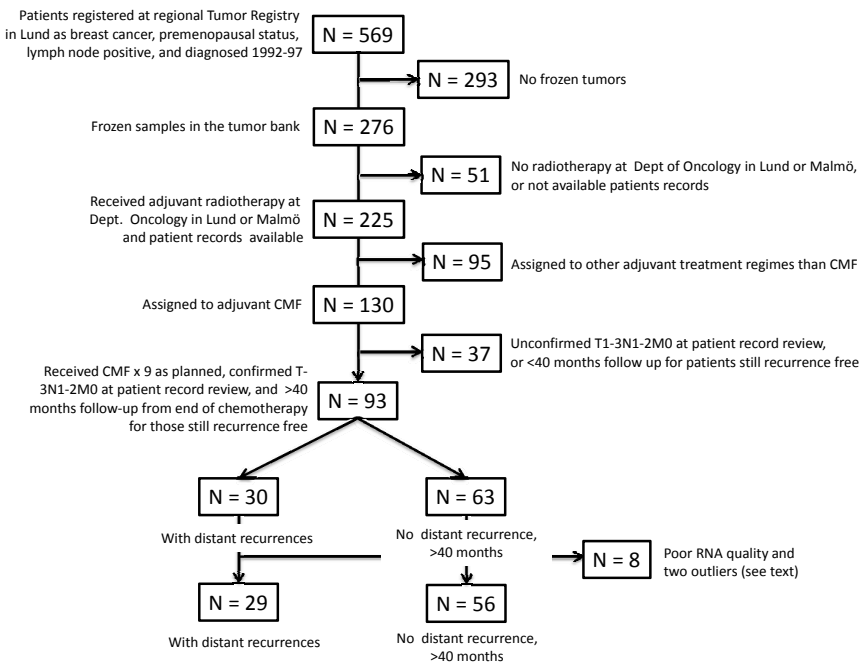


Figure 2. *The step-wise selection of patients included in Paper I.*

stored in the tumor bank of the South Sweden Breast Cancer Group. This tumor bank, and also the tumor banks in Stockholm, Gothenburg, and Linköping, have provided us with fresh frozen tumor material for our studies.

In Paper I the patients were diagnosed in 1992-97 with primary breast carcinoma stage T1-3N1-2M0. The patients were selected according to the following criteria: premenopausal women, referral to the Department of Oncology in Lund or Malmö for postoperative adjuvant radiotherapy, treatment with nine cycles of CMF, either distant recurrences within 40 months after completion of CMF or free from distant recurrences for 40 months or longer. Technical criteria were: frozen primary tumor samples still available, good quality of extracted RNA, and successful hybridization. Initially 93 patients were selected, but 8 were excluded due to poor RNA quality and large changes in the reference intensities, so called outliers (Fig. 2). This left 85 useable patient cases. Histopathological re-evaluation of all primary tumors was performed.

In Paper II 20 of the 85 patients from Paper I were selected. Ten of these patients developed distant recurrences within 40 months (5 ER+/5 ER-) and 10 did not (5 ER+/5 ER-).

In Paper III the patients were selected in two different inclusion steps (Fig. 3). Initially, 102 patients from a randomized clinical trial in the South and West Health Care Regions in Sweden[81] and 19 patients from a population based cohort study with a nested case-control study (Stockholm and South Sweden)[27, 82] were included. The patients included were all operated with breast conservation surgery and axillary clearance, without lymph-

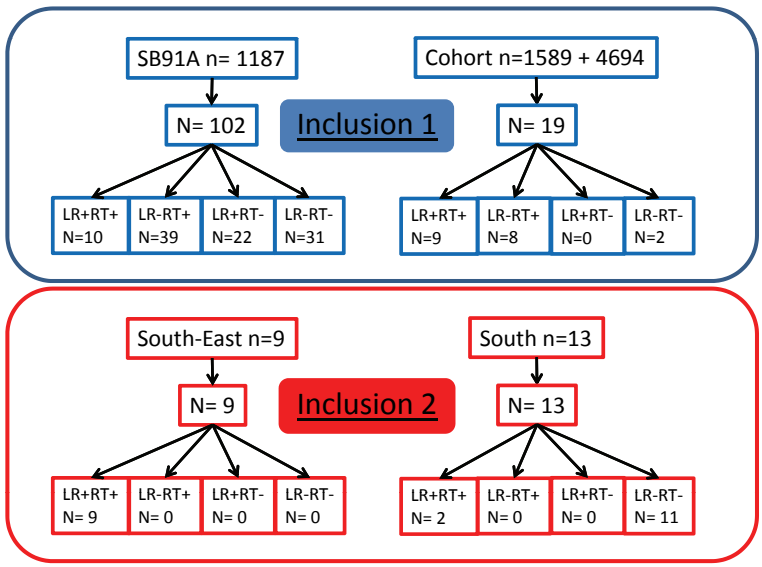


Figure 3. *In Paper III, 143 patients in total were included. The inclusion was done in two steps to obtain a larger homogenous subgroup of only ER positive tumors. The tumors were collected from a randomized study, a case-control and a cohort study in the first inclusion step. In the second inclusion step the patients were collected from a local recurrence registry.*

node involvement, tumor size < 30 mm (two patients had tumors measuring 32 and 40 mm, respectively, and one was T2 without any further information of size), tumor-free margins (> 1 mm), no multicentricity, and with frozen tumor tissue with good RNA quality available. Patients with a recurrence in the contralateral breast, or distant metastases prior or simultaneous with a local recurrence, were excluded. Twenty-two additional patients with ER positive tumors from the South-East and South Health Care Regions were also included in order to enable analysis in ER positive and ER negative subgroups separately (Fig. 3).

In total, 143 patients were included, and of those, 127 had received no adjuvant treatment. The patients, which received adjuvant treatment, had received tamoxifen and/or chemotherapy. Histopathological re-evaluation of all primary tumors was performed.

In Paper IV we studied a defined cohort of women diagnosed with breast cancer in the Uppsala-Örebro region during 1993-2004. Inclusion-criteria were tumor size ≤ 50 mm, no lymph node metastases, and no adjuvant chemotherapy. Originally 900 women fulfilled these criteria (Fig. 4). In this cohort two hundred and forty patients died from breast cancer and were defined as cases. Controls were women alive at the time for the corresponding case's death. After reviewing the medical records fifty cases and corresponding controls were excluded because they did not fulfill the inclusion criteria. The average age was 66 years for cases and 61 years for controls. The average size was 20 mm for cases and 16 mm for controls. All patients had undergone surgery consisting of either modified radical mastectomy with axillary dissection or breast conservation surgery with axillary dissection and post-operative radiotherapy of the breast. 53 (28%) cases and 48 (25%) controls received anti-hormonal therapy.

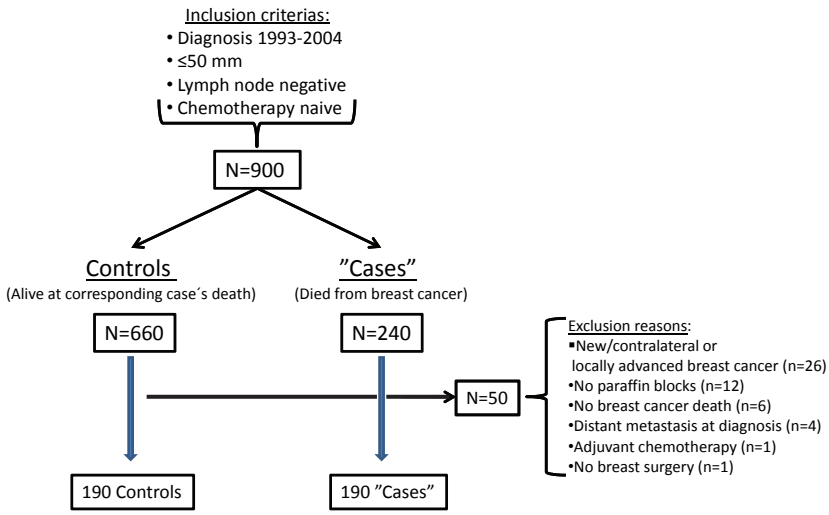


Figure 4. *The selection of 190 cases who died in breast cancer and 190 corresponding controls alive at the corresponding case's time of death in Paper IV.*

Methods

For sporadic breast cancer, genetic variations seem to be more complex than one single gene aberration, and the interaction of several genes is most likely of importance. The genetic information is stored in the DNA, and the human DNA is transcribed into messenger RNA (mRNA), which is the messenger carrying the genetic information to the ribosomes for translation into a proteins. Today 20 000-25 000 genes are known, with an even more complex proteome due to mRNA splicing, and post-translational modifications.

Gene expression profiling

The microarray technology consists of an array with several thousand single-stranded nucleotide sequences/genes spotted on a glass slide. The mRNA is extracted with Trizol from fresh frozen tumor tissue, and the tissue suspension is separated into three phases, a water- , an inter- , and an organic-phase. The RNA is found in the water phase and the proteins in the organic- and inter- phases. The proteins can be detected on a 2-DE gel, see below.

In this doctoral work two types of microarray technologies were used: cDNA microarray in Paper I and oligonucleotide array in Paper III. For the oligonucleotide array, shorter fragments (approximately 60-70 base-pairs) are used, whereas in the cDNA microarray, larger fragments (approximatelly 1 500 base-pairs) are used. Each spot in both a cDNA microarray and in an oligonucleotide array is called a reporter. The isolated mRNA is converted into single stranded fluorescently labeled cDNA and is then combined with the single-stranded genes/or fragments of genes on the chip. This process is called hybridization. To enable comparison of the patient samples each sample is hybridized together with a reference RNA. The reference RNA and the tumor mRNA both have the ability to bind to the same binding site. The reference RNA is labeled with Cy 5 (green) and the tumor tissue is labeled with Cy 3 (red) (Fig. 5). The reference RNA used was Stratagene reference, which is a pool of RNA from different cell lines from various tumors. The point is that the reference RNA binds to as many spots as possible making a comparison to the tumor RNA possible, giving us a ratio. The interesting comparison is thus not between the reference and the tumor, but rather between the difference in the ratios from the samples, since the same reference is used for all the samples.

The RNA quality is of great importance for the results[83]. In Paper I the RNA quality was measured using a Bioanalyzer in which the ratio of the two peaks from the ribosomes in the eletropherogram can be measured, 28S and 18S. In Paper I, 6 patients were excluded due to degraded RNA. In Paper III a new method for checking the RNA quality was implemented, using an algorithm, RNA integrity number (RIN) based on the ratio of 28S:18S in rRNA and also other region points in the eletropherogram[84]. Twenty-one samples were excluded in Paper III due to poor RNA quality, RIN values <6[83].

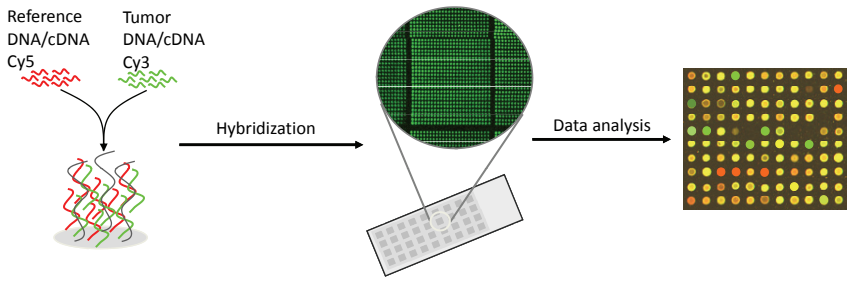


Figure. 5. *Gene expression is analyzed on a chip with several thousand gene segments (reporters). The RNA from the tumor is transcribed into fluorescently labeled cDNA with red (Cy3), the reference RNA is transcribed with green (Cy5), and both are hybridized onto a microarray-slide. The microarrays are then scanned and a color image is generated. Genes upregulated in the tumor appear red, whereas those with a decreased expression appear green in relation to the reference. Genes with similar expression of tumor and reference appear yellow. The ratio between the tumor sample and the reference sample is then calculated, which makes it possible to compare.*

Proteomics

The proteins are dynamic compounds that can be modified and changed in quantity. They are responsible for carrying out the molecular functions in cells or organs. Proteins are clinically and biologically interesting since they are often the targets for different targeted therapies. Proteins can be detected by antibodies for treatment prediction and diagnosis. The total number of proteins, including all the different modifications, has been estimated to be several million.

Proteomics is the broad definition for a collection of technologies used to study proteomes, i.e. the protein complement of a genome. Recent improvements in proteomic technology have enabled the implementation of proteomics into biological and clinical research areas. Especially interesting is the development of targeted proteomic techniques that allow sensitive and reproducible measurements on a pre-defined set of target proteins. Traditional proteomics was initially performed with a technique based on 2-D gels, but later also using relative quantitative mass spectrometry methods (shotgun proteomics). When analyzing proteins, all proteomic techniques share three common steps: protein/peptide separation, protein identification, and protein quantification. In Paper II, 2-D gels were used on proteins extracted from Trizole samples from fresh frozen tumor. In 2-D gels, the proteins are separated in two dimensions, first according to the isoelectric point (pI) and then according to size on an SDS PAGE gel, creating a two-dimensional pattern or spots. The integrated optical density of the spots can then be compared between the different groups with a computer algorithm, and the induced and repressed proteins can be found. The proteins of interest can then be identified with mass spectrometry. 2-DE is a time-consuming technique with restrictions so that only a limited part of the proteome can be detected. Membrane proteins and low abundant proteins are hard to measure.

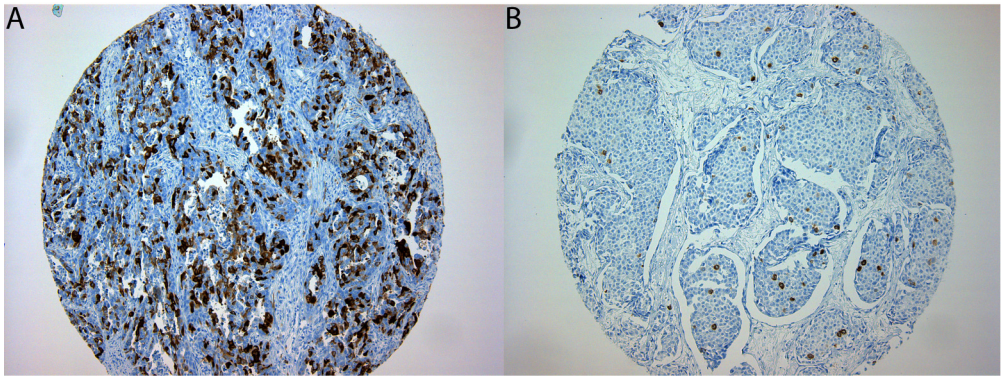


Figure 6: *Positive (A) and negative (B) staining of cyclin B1 (Epitomics Inc.) on tissue microarray slides from breast cancer cells with 100 x magnification.*

Immunohistochemistry

Immunohistochemistry was used in Paper IV and performed on formalin-fixed paraffin-embedded tumors. Representative areas from each tumor were punched out and tissue microarrays (TMAs) were constructed using two cores (diameter 1 mm) from each tumor. A monoclonal antibody was used for Cyclin B1 (Y106, Epitomics Inc, Burlingame, CA, USA), (Fig. 6). Tissue samples from tonsils were used as positive control.

Conventional factors

Histological grade was evaluated according to Elston and Ellis[85]. ER and PgR were analyzed routinely after primary operation with enzyme immunoassay or validated with immunohistochemistry. Receptor values ≥ 25 fmol/mg protein (enzyme immunoassay) or >10% (immunohistochemistry) were considered positive.

All patients with HER2-amplified tumors or an IHC score of 3+ were considered HER2-positive.

Statistical analyses

Endpoints

Different endpoints are used for evaluation of the prognosis of breast cancer patients. Since survival is worse after development of a distant recurrence than after a regional or local recurrence, it is important to separate different end-points. Recurrence-free survival (RFS) includes local, regional and distant recurrences, whereas distant disease free survival (DDFS) includes only distant recurrences.

Disease free survival (DFS) is the length of time after treatment with no sign of the disease and without other events such as death due to other diseases or causes.
Ipsilateral breast recurrence includes only local recurrences.

Overall survival includes all causes of death. 5-year and 10-year survival rates are frequently used for breast cancer.

Death from breast cancer is also used when the endpoint only includes death in breast cancer, whereas overall survival also includes other causes of death.

Statistical analysis of microarray data

Statistical analysis of microarray results in Papers I and III was done in cooperation with the Department of Theoretical Physics. In brief, the statistical evaluation was performed as follows: Given the results from the hybridization, the tumor samples are compared in order to identify the genes of interest. First the error model is applied to avoid uncertain measurements, which are called outliers. The error model moves uncertain measurements towards the mean across the assay, giving a smaller influence. However, certain measurements will not be moved in the same way. The proportion of certainty of a sample is evaluated from the signal-to-noise ratio (signal/background). Then the genes are filtered. If a gene is missing in more than 10% of the patient samples, then that gene is removed ("filtered out") from further analysis from all of the patient samples due to "insufficient data". If there is no variation of a gene in all of the patient samples then that gene is also removed from further analysis as "uninteresting".

Thereafter a supervised classification was applied: Artificial Neural Network (ANN) in Paper I and Support Vector Machines (SVM) in Paper III, and the genes were then ranked according to their importance. As far as ANN is concerned, variables (expressions of genes for each patient) were reduced using principal component analyses (PCA). When SVM was used there was no need for reducing the variables with PCA. PCA finds the linear combination of genes in gene space that gives the dominant directions. The 100 genes were usually reduced to 2 or 4 PCA/patient. From each of the samples, the PCAs were put into an ANN. The ANN predicted if the sample was from a patient with or without a distant recurrence (Paper I). If the prediction was in agreement with the recurrence status then the system was strengthened, and when the prediction was not in agreement, the system was adjusted. In addition, the clinical variables could be used in an ANN, instead of using the genes, but since the clinical variables were so few (seven in Paper I) no PCA was needed. After creating a training set, the results needed to be cross-validated. Then the classifier needed to be validated using a dataset not included in the construction of the classifier. One commonly applied approach is the leave-one-out procedure. There, one of the samples is left out, the genes/clinical variables are adjusted for the rest of the samples, and then the variables are re-tested to see if the previously left-out sample could be correctly classified or not. The results could be presented in different ways: the receiver operating curve (ROC)-area and Odds ratio were used in Papers I and III. The ROC shows the sensitivity (y-axes) and specificity (1-specificity on the x-axes) of the prediction using no specific cut-off. An ROC curve 0.5 equals a random guess while an ROC curve 1.0 is a perfect classification.

The Odds ratio is the risk-difference between two groups, in Paper I recurrence *vs.* no recurrence. A ratio was calculated using the following: the risk of developing recurrence

in the high-risk group using our predictor divided by the risk of developing recurrence in the low-risk group using our predictor. The ratio can then be compared using different predictors, e.g. the comparison of our gene-expression predictor to NPI. The ratio is dependent on the chosen cut-off. On the other hand, when using the ROC curve, one advantage is that no cut-off is needed. Fisher´s exact test and Wilcoxon test were used for evaluation of the clinical variables. Fisher´s exact test was used for categorized variables, such as ER, which are evaluated as groups (positive or negative). For continuous variables, e. g. age, Wilcoxon's was used.

Statistical analysis was done to rule out potential confounding factors in the patient and tumor material, described below. In Paper III, a new way of evaluating confounding factors (age, and different health care regions) was applied. A test set was created with patients who behave exactly opposite to the usual connection, e.g. between age and a local recurrence (<50 years and LR-RT-/LR-RT+ or >50 years and LR+RT+). The ROC area for our profile was 0.88, and since low age normally is a predictor for a local recurrence, age as a confounding factor is neglectable. Health care regions were evaluated using the same procedure, but in this case samples from health care regions in West Sweden and Stockholm were mixed and compared to samples from the South and South-East regions. This test set also showed no sign of a confounding factor (ROC 0.87).
In Paper I the clinical markers performed slightly better than the gene expression profile making it unnecessary to rule out confounding factors.

Statistical analysis of proteomic data

The spot analysis in Paper II including detection and matching of the spots, and statistical analyses were done in Ludesi Interpreter™, http://www.ludesi.com using Student t-test, and the level of significance was set to 5%. Thereafter significantly differently expressed spots were further filtered based on spot quality with a manual control to include only visually clear differences. The significantly regulated proteins were further reduced after masspectometric analysis, where about 80% were identified.

Evaluation of clinical markers was done with Mann-Whitney U-test, a test for continuous variables.

Statistical analysis of immunohistochemistry data

In Paper IV, prognostic factors such as age, tumor size, hormone receptors, histological grade, mitotic count, tubuli, nuclear atypia, HER2, Ki-67, cyclin A and cyclin B1 were analyzed univariatly, using conditional logistic regression analysis to estimate odds ratios. Since histological grade, mitotic count, nuclear atypia, Ki-67, cyclin A and cyclin B1 were highly correlated in Paper IV, multivariate conditional logistic regression analysis, including these factors simultaneously, was not considered to be appropriate. In addition, models adjusted for age, tumor size and ER/endocrine therapy were performed. Correlations between cyclin B1 and other clinicopathological parameters were assessed with Spearman´s correlation test.

Results

Paper I: *Gene expression profilers and conventional clinical markers to predict distant recurrences for premenopausal breast cancer patients after adjuvant chemotherapy (CMF)*

In Paper I, 85 premenopausal, lymph node positive, CMF (Cyclophosphamide, Methotrexate, 5-fluorouracil) adjuvantly treated breast cancer patients with distant recurrences (within 40 months) and without distant recurrences (for at least 40 months) were compared regarding gene expression profiles. Using ANN, the gene expression data were able to distinguish between the two patient groups (OR 6.5 CI 1.4-62: ROC 0.70), but conventional clinical variables and NPI were still slightly better predictors (OR 15 CI 3.1-140, OR 10 CI 2.1-97, respectively, and ROC 0.78 and 0.79, respectively). The top-10 list included genes involved in functions such as signaling, 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.

We also investigated preselected genes known to be of importance for drug resistance (see supplement in Paper I), and discriminating genes from a publicly available microarray data set, the 70-gene profile (van´t Veer et al.[10]) in comparison to our gene expression profile. With both of these approaches lower ORs were obtained (OR 6.0 CI 1.3-57, OR 3.9 CI 0.80-38, respectively). The ROC for the drug resistance genes showed a slightly higher ROC, whereas the 70-gene profile showed a slightly lower ROC (0.78 and 0.69, respectively).

Paper II: *Proteomic analysis identifies candidate proteins associated with distant recurrences in breast cancer after adjuvant chemotherapy*

In Paper II, we developed a protocol to extract proteins and mRNA from the same tissue. For twenty of the 85 patients from Paper I, the proteins were separated with 2-DE as a complementary approach. Several proteins distinguishing the recurrence group from the non-recurrence group were thereby detected. When comparing only ER/PgR positive tumors (recurrence/no recurrences), we found 7 significantly regulated proteins. The proteins with higher concentrations in the recurrence group were involved in translation/folding, iron ion binding and protease inhibition, whereas proteins involved in signaling, ubiquitination, and splicing showed a lower concentration. In addition, two proteins were identified in the ER negative subgroup with higher expression in the recurrence group. These proteins were involved in cytoskeletal processes and extra cellular matrix.

The regulated proteins were also compared to the findings in gene expression with a somewhat supporting result. When comparing a ranked gene list from Paper I to the identified proteins, some correlations were seen. One example is the thioredoxin domain containing protein 5 (similar to glucose-regulated protein) that was increased, and the

corresponding gene was overexpressed and ranked 59th of 4484 most important genes. The gene for thioredoxin domain containing protein 5 was also upregulated in the patients with distant recurrences in Paper I.

Paper III: *Gene expression profiling in primary breast cancer distinguishes patients developing local recurrence after breast conservation surgery, with or without postoperative radiotherapy*

In Paper III, local recurrences in breast cancer were studied using oligonucleotide microarrays. In this study we included 121 patients homogenously treated with breast conservation surgery with/or without post-operative radiotherapy with lymph-node negative tumors and tumor-free margins. Radiotherapy, with 50 Gy/25 fractions, was administered without a booster dose. Among the ER positive tumors we found a promising gene expression profile which was able to distinguish patients that developed local recurrences despite radiotherapy (n=20) from patients which did not develop local recurrences with/or without radiotherapy (n=80), with an ROC area 0.91. With 90% sensitivity the clinical situation would result in 18 of 20 LR+RT+ patients and 70 of 80 LR–RT+/LR–RT– patients correctly classified (87.5% specificity). 5 237 of 26824 reporters had a p-value below 0.001 (false discovery rate of 0.005). Among the ER negative tumors a somewhat weaker but still significant profile was found with an ROC area of 0.74 (n=21). The ER positive subgroup was by far larger than the ER negative subgroup, n=100 and n= 21, respectively, which may be one explanation for the superior performance in the ER positive group. It has previously been found that ER negative tumors tend to be more heterogeneous resulting in difficulties finding gene profiles[56]. Furthermore, we demonstrated that the gene expression profile provides substantially additive value to established clinical markers, e.g. age, histological grade, and tumor size in predicting a local recurrence despite radiotherapy.

We also investigated the capacity of the gene profile to predict a local recurrence by comparing patients not treated with postoperative radiotherapy with and without local recurrences. The ROC in the combined ER positive/negative group was 0.66 (n=66). The ER positive and negative subgroups were too small to analyze separately (n= 52 and n=14, respectively).

Paper IV: *Cyclin B1 is an independent prognostic proliferation marker with a high reproducibility in a population based lymph node negative breast cancer cohort*

In Paper IV, the proliferation marker cyclin B1 was evaluated using immunohistochemistry in a case-control study. 190 cases were selected and defined as breast cancer patients who died from breast cancer and controls were alive at the corresponding case's death (n=190), (Fig. 4). The inclusion criteria were tumor size <50 mm, no lymph node metastases, and no adjuvant chemotherapy. Cyclin B1 was found to be a significant factor (OR>2) for breast cancer death in both univariate and multivariate analyses, adjusted for tumor size, age and endocrine therapy. Also, the reproducibility between two different investigators was good to very good (kappa values 0.74-0.82) when counting different numbers of cells.

Discussion

Breast cancer, today, is a major threat that kills more than 800,000 women each year worldwide. With the aim to improve treatment results, the number of studies of translational research in breast cancer has increased during the last decades. To bring the translational research closer to clinical use, a large web-based consultation of breast cancer professionals was carried out worldwide. The aim was to identify the topics of highest priority. Four hundred and twenty clinicians, academics and researchers or scientists participated and voted on 70 topics of breast cancer[86]. The top priorities, seen in Figure 7, included molecular signatures for selection of patients not in need of chemotherapy and to choose an optimal chemotherapy regime (Fig. 7).

The top priority, to spare patients from chemotherapy, and the seventh priority, to identify low-risk patients not in need of adjuvant therapy, are closely related to Paper IV in which cyclin B1 expression using immunohistochemistry is validated as a prognostic factor in a chemotherapy naïve patient cohort. The primary aim of Paper IV was to evaluate the

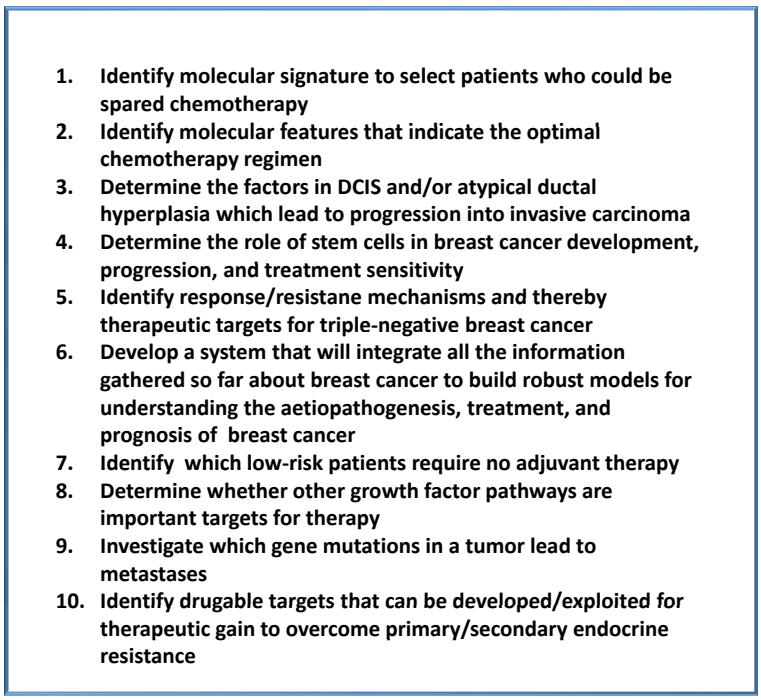


Figure 7. *Top ten priorities of breast cancer research, conducted from a web-based consultations of breast cancer professionals[86].*

prognostic importance of cyclin B1 in a population-based homogenous chemotherapy naïve patient cohort, and the second aim was to investigate the reproducibility of the evaluation of cyclin B1 staining with immunohistochemistry. We found that a higher percentage of positively stained cyclin B1 was significantly associated with breast cancer death. Therefore, cyclin B1-positivity could be useful as a marker to decide whether a patient should be treated with chemotherapy or not - right in line with the top priority. Since the prognostic importance of cyclin B1 also remained after adjustment for adjuvant endocrine therapy, cyclin B1 may also be useful to identify patients not in need of any adjuvant systemic therapy. Cyclin B1 has previously been found to be of great prognostic interest[87-90], and it is one of the genes included in the 21 genes in Oncotype DX®[8]. Furthermore, the most important gene expressions in many of the profiles are associated with cell cycle regulation and proliferation[91]. Today the selection of patients in need of chemotherapy is based on conventional clinical markers such as histological grading, which is partly composed of a marker for proliferation, i.e. mitotic count. Histological grading is a well established method but is hampered by difficulties with interpersonal reproducibility[92].

The reproducibility of cyclin B1 was a specific focus in Paper IV. There we used two observers using two different types of manual methods: one observer used a normal light microscope and the other used a computer analysis of the microscope picture. The reproducibility was good to very good depending on the number of cells counted. The decision about which factor should be used in routine clinical management of breast cancer patients should be based on: (1) the prognostic strength of the factor, and also (2) more practical issues such as reproducibility and costs of the analyses. Since histological grade and cyclin B1 are associated, they tend to contain overlapping information. Cyclin B1 analyses could be used as a complement to histological grade to better identify patients with low risk of recurrence not in need of adjuvant chemotherapy, or not in need of any adjuvant systemic treatment.

The second priority of the top ten survey was to select the optimal chemotherapy regime.

Clinical outcome after polychemotherapy CMF (Cyclophosphamide, Methotrexate and 5-FU) was studied in Paper I and II. For this purpose we used high through-put technologies, gene expression and 2-DE for proteomic analyses. The aim of Paper I was to find a gene expression profile to distinguish adjuvantly CMF treated patients with and without distant recurrences. Eighty-five node positive homogenously treated patients were included and analyzed with cDNA microarray. In Paper II we studied a subgroup (n=20) of the patients from Paper I with regard to protein expression using 2-DE. In Paper II we also subdivided patients with regards to ER status.

CMF is a poly-chemotherapy which has been used for decades, but today newer generations of polychemotherapy regimens with anthracyclins and/or taxanes have been shown to be more effective and are used more frequently[93]. The reasons for studying a patient cohort treated with CMF were: (1) the long follow-up period, and (2) the phenomenon of drug resistance found in CMF treated patients, which could also be relevant for newer treatment regimes, (3) the newer treatment regimens such as FAC/FEC (5-fluorouracil, anthracyclin/epirubicin, cyclophosphamide), which both contain

cyclophosphamide and 5-fluorouracil, and (4) CMF is less toxic than FAC/FEC and thus CMF is still an option for older patients.

In Paper I we were able to distinguish CMF treated breast cancer patients with distant recurrences from recurrence-free patients with a gene expression profile using cDNA microarray. Our gene list was also compared to the well-known gene profile of van´t Veer[10] and gene expressions associated to drug resistance from the literature. Our gene list gave a slightly better prediction. However, the performance of our profile was not superior to already clinically established prognostic markers. Reasons for our failure to outperform conventional clinical markers could be: (1) intratumor heterogeneity, (2) study design, (3) distinct differences in gene expression between ER positive and ER negative tumours, (4) size of study, (5) evaluation of RNA quality, (6) acquired drug resistance, (7) different statistical methods, and (8) different array platforms. These reasons are discussed in more detail below.

(1) Breast cancers are known to be very heterogeneous due to the presence of a number of different cell types such as in situ and invasive cancer cells, stroma cells, and lymphocytes. Perhaps this problem could be reduced using micro-dissection, where only cancer cells are taken out. However, it has also been showed that the stroma cells surrounding the tumor cells are important, maybe due to secretion of several factors important for the behavior of the tumor cells[94]. Farmer et al.[95] recently found that a set of stroma related genes could predict resistance to chemotherapy in ER negative patients.

(2) Study design. We included 29 distant recurrences within 40 months and 56 recurrence-free for 40 months or longer, all receiving adjuvant CMF treatment. Adjuvant CMF improves the 10-year survival in patients <50 years by 11% (from 42% to 53%) for lymph node positive[67], meaning that many patients actually would have been recurrence-free also without CMF treatment. A significant proportion of the recurrence-free patients are thus cured by surgery and radiotherapy alone and do not contribute to the identification of responders. The problem is that with such a study design we will never know who are the responders and who are already cured by the local therapy. Therefore it would be necessary to perform a randomized study of chemotherapy vs. a control group which is not treated with adjuvant chemotherapy.

(3) ER status. Other studies have also implemented a subdivision of the patients with regard to ER status[7, 56] when trying to develop gene expression markers for distant recurrences. Maybe this has also influenced the results in Paper I. However, the sample size was too small in our study for a subdivision based on ER status (46 ER positive and 39 ER negative cases).

(4) Sample size. With a greater number of samples, the statistical power will be increased, and the analysis can be performed in more homogeneous subgroups. In Paper I the relatively few patients with ER positive tumors in each subgroup was not sufficient to obtain a significant prediction. The importance of a larger number of cases was exemplified when using the NPI (Nottingham prognostic index). NPI was constructed in a large study[38], and it gives an index which includes the number of tumor-involved lymph nodes, tumor size, and histological grade. In our study we used the same clinical parameters as

in NPI but calibrated with ANN in our own patient cohort. We obtained a better result from the original index (ROC 0.79 for original NPI, and ROC 0.74 for NPI calibrated in our material). This result shows the benefit of using a cohort consisting of thousands of patients. The limited size of our patient cohort could also explain the lower predictive value, and that could be one reason for not finding a gene profile working better than the clinical parameters.

(5) A suboptimal method for evaluation of the RNA quality. During my doctoral work another method (RNA Integrity Number, RIN) has been introduced in our laboratory, and it shows somewhat better results[83].

(6) Acquired drug resistance. The tumor material used in Papers I and II was taken at primary surgery before CMF treatment. Drug resistance can either be inherited or acquired. If the drug resistance is inherited it may be possible to find a gene profile from the primary tumor, but if drug resistance develops during the CMF treatment we should not be able to find a gene profile from the primary tumor. In either case, the clinical markers would have the same problem.

(7) The development of newer statistical methods.

(8) The development of better gene expression microarray platforms. In Paper I we used one of the early generation array platforms.

The latter two reasons may be possible explanations for why one recent study was able to identify a 6-gene profile that predicts recurrence after adjuvant CMF therapy, although only two ratios of three genes were able to give additional information to conventional markers[96]. When comparing their top 25 genes to our top 10, no overlap was found.
Another recent study has found a gene expression profile that functions in combination with clinicopathological variables, to stratify breast cancer patients into different risk groups, including sensitivity to cytotoxic therapies[97].

In Papers I and II we used the same patient cohort but with different methods, cDNA microarray and 2-DE measuring RNA and proteins in a complementary way. There are drawbacks and advantages with each method. The advantages with microarray are the great amount of data and the quality of the data. The gene sequences on the microarray chip could be compared with the spots on the 2-D gel. The separation of the 2-DE spots is more difficult to interpret, and the matching is not as exact as the gene sequences on the microarray chip. The advantages with 2-DE is that the proteins are evaluated instead of the mRNA, and the proteins are the effectors which actually carry out the functions. This means that additional information can be assigned to such things as posttranslational modifications, splicing, truncations and localization.

In Paper II, we adapted a protocol for 2D-analysis from mRNA-extracted tumor samples. We found several significantly regulated proteins that distinguish tumors with and without recurrence in the whole patient group, as well as within ER positive and ER negative subgroups, separately. We also found regulated proteins when comparing ER positive and ER negative tumors. The regulated proteins were also compared to the 4484

ranked genes in Paper I. Thioredoxin domain containing protein 5 (similar to glucose-regulated protein) was increased in the group of patients with distant recurrences, and the corresponding gene was ranked 59th on the gene list. This gene was also upregulated in the group with distant recurrences. Two proteins involved in the initiation of translation, eukaryotic translation initiation factor 4A-II and 1A, were found to be increased in tumors with distant recurrences. Genes with similar functions were also upregulated in the tumors with distant recurrences in the gene expression data set. For example, three different eukaryotic translation initiation factors (factors 5, 2, and 4A-I), which were ranked 125th, 288th, and 367th, respectively, and eukaryotic translation elongation factor 1, ranked 76th, were elevated in the tumors with distant recurrences. Eukaryotic translation initiation factor 1A was downregulated in the tumors with distant recurrences and ranked 1596th. Note that, as has previously been shown, there is no absolute correlation between mRNA and protein expression[98], which may explain why not all proteins were detected on the gene list.

2-DE has previously been used in other breast cancer studies. Somiari et al.[17] compared four breast tumors to normal breast tissue and found several (n=737) regulated proteins. Interestingly, fewer proteins (n=41) displayed an up- or down-regulation in at least one sample occurring on all four gels, demonstrating the heterogeneity of the tumor samples and the limitations of the 2-DE technology. The relatively few numbers of candidate proteins, in comparison to gene expression results, demonstrate the drawbacks with 2-DE and its limitations.

Many proteins are expressed at such low levels that they will escape detection, e.g. ER is not found on the gel when comparing the ER positive to ER negative samples. There are also proteins not suitable for the 2-DE, e.g. hydrophobic membrane proteins. Such proteins may be of special importance, since they can be secreted and detected in peripheral blood. The limitations using 2-DE have been partly overcome in more recent proteomic techniques referred to as shotgun proteomics and targeted proteomics. Shotgun proteomics is a method to identify proteins in complex mixtures using a combination of high performance liquid chromatography and mass spectrometry. Shotgun proteomics can be combined with other labeling strategies such as ITRAQ in which up to four different stable isotopes are used for labeling four samples in parallel.

Despite the recent improvements in mass spectrometry, the dynamic range of the best instruments is limited. This results in a tendency of identifying the most abundant proteins, e.g. albumin and immunglobulins in serum. This is an inherent problem with shotgun proteomics, which can be circumvented by different approaches to achieve a deeper proteomic analysis. One example is the glycocapture technique[99] in which only the glycosylated peptides are selected. Since albumin and immunoglobulins are removed, there is a drastic reduction of the complexity. The limitations with low-abundant proteins in shotgun proteomics has led to the development of a new branch within proteomics called targeted proteomics[100]. In targeted proteomics, only an exclusive, targeted set of proteins is analyzed. This allows the detection and quantification of identical, non-redundant sets of proteins in multiple repeat analyses. This process is referred to as single reaction monitoring (SRM)[101]. Since the mass spectrometer only detects the proteins of

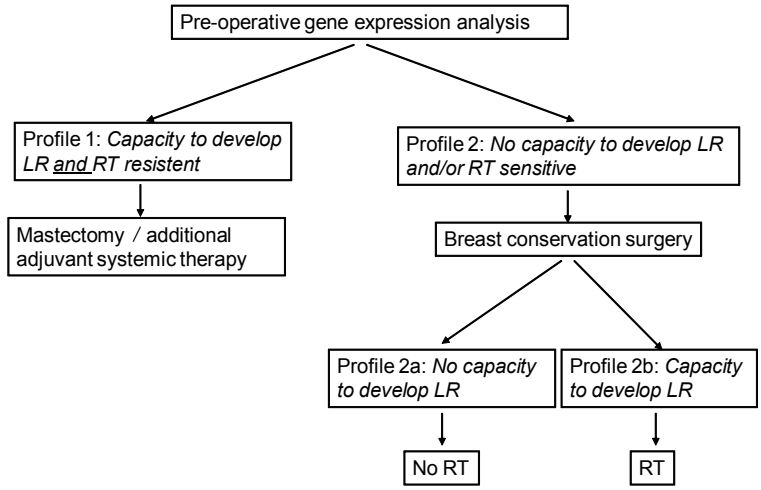


Figure 8. *Schematic way to preoperatively select surgery (mastectomy or breast conservation surgery) and radiotherapy according to the geneprofile.*

interest in this approach, it can be successfully applied even if the proteins of interest are present within a large background of other proteins. These traits may be critically important for the application of proteomics to biomarker discovery and validation.
In this doctoral work, I believe that the gene expression profile found in Paper III is the most important discovery. To my knowledge, results of such research dignity have never been presented before for local recurrences in breast cancer. In this study we identified a gene expression profile which was capable of predicting patients who developed local recurrences despite postoperative radiotherapy from patients who did not develop local recurrences, with or without radiotherapy. The possibility to distinguish the "radioresistant" tumors would enable improved treatment strategies, where patients with "radioresistant" tumors could be offered mastectomy as primary surgery and /or additional or different adjuvant systemic treatment (Fig. 8).

Today, besides tumor margins, the only clinically used markers for local recurrences are young age and multicentricity[26-29]. Today there is no marker for "radioresistance". Our gene profile provides additional information beyond the information now available from age, histological grade, and tumor size to predict a local recurrence despite radiotherapy. Recently, two other gene expression profiles for prediction of a local recurrence have been published[12, 11]. Nuyten et al. studied a predefined list "Wound-response" originating from Chang et al.[102]. That profile was originally found from serum-stimulated fibroblasts in vitro. In many aspects the wound-response is similar to invasive tumor growth and could also be implemented when predicting local recurrences after radiotherapy *vs*. no

local recurrences. However, they did not find any significant gene expression profile when all genes were included in the analysis. One possible explanation may be that their patient cohort was rather heterogeneous with regards to margin-, and lymph-node-status, variable adjuvant systemic treatment, and radiotherapy with and without booster dose. In addition, the analyses were not reported for ER positive and ER negative subgroups separately, which previously has shown to be of importance[7].

Our prediction profile was found to be superior when dividing the patients into ER positive and ER negative tumors resulting in very promising ROC areas only using ER positive tumors (ROC 0.91) but lower when combining the whole cohort (ROC 0.83). This emphasizes the importance to study breast cancers with different ER status separately, which previously has been reported for distant recurrences[7, 56, 103]. Also, it has been showed that different gene expression patterns are of importance in ER positive and ER negative tumors. Of particular importance are proliferation genes for ER positive tumors and immunological genes in the ER negative tumors[104]. In our study we only included patients treated with standard doses of radiotherapy, whereas Nuyten et al. also included patients with booster doses (n=25). A booster dose of 16 Gy reduces the risk of local recurrences, especially in patients younger than 50 years old[58], but with less satisfying cosmetic results[105]. This could mean that "radioresistant" tumors might need higher doses of radiotherapy to be cured. However, this stresses the fact that including patients treated with a booster dose might confuse the identification of the patients. Some patients treated with a booster might be "radioresistant" at the standard dose of 50 Gy, but the recurrence was prevented or delayed by the booster. Therefore, homogenously treated patient cohorts are preferable when evaluating the importance of gene expression analyses.

The wound-response-gene profile was also able to distinguish local recurrence patients in our material. However, the distinction was less stringent than when using our gene profile, with a ROC area 0.75 (p=0.007) within the ER positive group, 0.75 (p=0.08) within the ER negative group, and 0.61 (p=0.10) within the combined ER positive/ ER negative group. This should be compared to our list 0.91 ($p=9x10^{-6}$) within the ER positive group, 0.74 (p=0.08) within the ER negative group, and 0.83 ($p=9x10^{-5}$) within the combined ER positive/ER negative group.

The other gene profile which distinguishes a local recurrence after mastectomy was published by Cheng et al.[11]. Since breast conservation therapy with additional radiotherapy has shown the same prognosis as mastectomy, breast conservation therapy is the more commonly used operation today. Although the profile is of interest, 40-50% of breast cancer patients are still operated with mastectomy.

In Paper III, a number of improvements were introduced as compared with the results in Paper I. The local recurrence-free patients were subdivided into two groups: either with or without radiotherapy. The advantage with this study design is the possibility both to identify a prognostic profile (the risk of developing a local recurrence) and to identify radiotherapy sensitivity. The LR-RT- group included only patients with an inherent good prognosis. These patients are also present in the LR-RT+ group, but some patients in this group may also have the capacity to develop a local recurrence, but are radiosensitive. These two groups were then compared to the LR+RT+ patients, where the patients were

resistant to radiotherapy. The LR+RT- group was not included in the comparison since this was a mixed group containing both "radioresistant" and "radiosensitive" tumors. A weakness in the study design was the selection of the patients. The vast majority of the patients in Paper III were part of a randomized study which reduces selection bias, but the patients included in our study were only a very limited part of the total randomized trial. Furthermore, some patients (n=19) were from a case-control study or identified from the Tumor registry.

The study design is one of the changes made in Paper III as compared with Paper I with superior results. Another change made in Paper III, as compared with Paper I, was the subdivision of the patients with regards to ER status, with superior results compared to analyzing ER positive and ER negative tumors together. The RNA quality was evaluated in a different way, where the new method using RIN algorithms[83, 84] was used in Paper III as compared with the ratio (28S/18S) used in Paper I resulting in a better evaluation of RNA quality in Paper III. The statistical analysis ANN, used in Paper I, was changed to the better performing SVM. Also a newer array platform was used in Paper III.

The seventh priority of the top ten list in the web-based consultation of breast cancer professionals was to identify low-risk patients not in need of any adjuvant treatment. If you include local recurrences and radiotherapy, this question is one of the aims in Paper III. In Paper III we also sought a gene profile able to distinguish the patients not capable of developing a local recurrence and thus not candidates for post-operative radiotherapy. Side effects for the patients and costs for the hospital would thereby be avoided. A profile for patients not capable of developing a local recurrence was found, by comparing LR+RT- vs. LR-RT-, with a ROC-area of 0.66 (p=0.04) within the combined ER positive/ER negative group. Unfortunately, the numbers of patients were too low to allow investigation of the ER positive and ER negative subgroups separately.

The conclusion from Paper III was that we found a very promising gene profile among the ER positive tumors to distinguish "radioresistant" tumors. We also found a somewhat less predictive, but still significant, gene profile to distinguish tumors not in need of radiotherapy. Both profiles need to be validated in a larger study.

The priorities of breast cancer professionals are permeated with the importance of new molecular signatures. This statement most likely refers to the gene expression profiles available today. In the past years a large number of gene lists have been published of which some add independent prognostic value to clinical and histopathological factors. However, no gene expression profile is yet generally recommended for clinical use due to the lack of confirming studies. The recurrence score has shown promising results in other patient cohorts[106, 107], and it is frequently used in the United States outside of clinical trials. OncotypeDX® and Mammaprint® are currently being evaluated in prospective, randomized clinical trials, TaylorX (Oncotype DX®) and MINDACT (Mammaprint®).

It became evident that the largest studies[7, 10, 16] showed very little overlap in genes. Still it is obvious that gene expression data is of great importance although it is needed to be critically evaluated. New gene-expression signatures that are based on biological differences such as hypoxia and tumorigenic breast cancer cells[108, 109] may provide complementary

information where several profiles are evaluated simultaneously[110] and may finally lead to new tools and better understanding.

The main aim of my dissertation is related to the discovery of prognostic markers and treatment predictive factors. The use of better prognostic markers will allow improved selection of patients for treatment. This doctoral work represents a step in the direction towards personalized medicine with a vision of curing a larger proportion of breast cancer patients due to better treatment selection. Furthermore, overtreatment with adjuvant therapy will be reduced. More specifically, the aim was to find gene expression profiles and protein patterns which could distinguish the patients with regard to chemotherapy "resistance" and "radioresistance". The most important discovery was a gene expression profile for "radioresistance", which may have a great clinical importance. Also, we found a gene expression profile and proteins to predict patients who develop distant recurrences after adjuvant chemotherapy. Finally, we aimed at finding prognostic markers for patients not in need of chemotherapy, and found cyclin B1 to be a prognostic marker in a population-based case-control study. We have used and evaluated several different high through-put methods, and found advantages and drawbacks with all of them. However, we found advantages studying both genes and proteins which give complementary information. Of critical importance is a proper study design and preferably tumor material from randomized trials. In my opinion all patients, whether included in trials or not, should have fresh tumor tissue stored, which would be very important for further investigations. In conclusion, we have taken a small step towards personalized medicine, a development that is needed for further progress.

My future perspectives

The goal of breast cancer research is to find a cure for all patients, a task not easily fulfilled. However, it is important to remember how the prognosis for the patients has improved during the last decades. That improvement has occurred partly thanks to research. The goal with my future research is to continue with the ideas which came up during my doctoral work and hopefully to implement them in the clinic. I strongly believe that the improved knowledge about the subgroups within breast cancer may also be of relevance for other cancer forms. Treatments based on specific genetic aberrations may be applicable to other cancer types harboring the same aberrations. Ideally, targeted treatment could be offered to patients with tumors of different origin with activation of the relevant molecular pathways.

The gene expression profile found for the prediction of "radioresistant" tumors in Paper III will be highly prioritized in my future research. A new cohort of patients is being collected for verification of the gene expression results and for using array CGH (comparative genomic hybridization). One of the advantages with array CGH is the possibility of analyzing paraffin-embedded tumor material, which will increase the number of available samples from breast cancers patients. If the profile can be verified in a new study, the next step would be to validate the profile in a larger prospective patient cohort. A targeted list of the genes from our gene profile is also extremely interesting when searching for markers in serum from patients developing local recurrences despite radiotherapy. It would be much simpler if we could detect treatment predictive markers in serum, rather than in tumor extracts. For serum-based procedures, targeted proteomics and SRM would be used.

Kinases are key control elements in all cellular processes and are the most prominent class of oncogenes. The kinases' activity is controlled by reversible phosphorylation and the abnormal regulation of the kinase activity is a common event in malignancies. Analyzing the informative parts of this kinase activity network in a cancerous tissue would be highly informative for classification of cancer sub-types and determination of treatment targets based on the kinases. This will be of special interest for combined drug treatments or multi-kinase inhibitors. At present, the tyrosine kinases are analyzed separately using conventional technologies such as ELISA. However, recent improvements of proteomic technologies have allowed the detection of several kinases simultaneously. In an ongoing project we aim at quantifying a significant proportion of the human kinases in breast cancer to correlate the kinase protein abundance to the likelihood of having a distant recurrence. A goal would also be to validate candidate markers directly in blood due to easy access of blood. In an ongoing project we are using a technology to specifically enrich for glycosylated peptides to decrease sample complexity. Newer techniques such as single nucleotide polymorphisms (SNP), methylations and miRNA should also be taken into considerations in future studies, since they may also be of importance for further improving the concept of personalized therapy.

Acknowledgement

The journey I have made throughout the years as a PhD student has given me the opportunities to meet a lot of interesting people in Lund, in Seattle and in Zürich. It has been a wonderful time. The research has given me opportunities to collaborate with people with different backgrounds than my own clinical background, and it has opened my mind to the world of research. I believe I have been unusually lucky to work with three supervisors representing three different areas of research: Mårten Fernö, my main supervisor, Carsten Peterson and Per Malmström, my co-supervisors. All of you are excellent scientists, experts in your fields and also have wonderful personalities. The combination of expertise, skills, viewpoints, and pleasant personalities that you three possess has resulted in a perfect trio and collectively, the sum of the input is greater than the sum of the parts alone.

I would like to thank a lot of people who have made this journey possible. First of all, my supervisor, Mårten Fernö, for guiding me, but also for letting me believe in my own ideas and visions. The german word for PhD supervisor is "Doktorvater" and I believe it suits you very well. You have supported me in a very caring way, shared your knowledge, and given me so much of your time and patience. We have often worked individually in different ways, and the combination has been very productive. We have also shared a lot of happy memories, and you have become a very good friend.

My special thanks to my two co-supervisors, Carsten Peterson and Per Malmström; Carsten Peterson, for your enthusiasm and guidance into the world of theoretical physics. It has been a pleasure working with you and sharing our totally different research fields.
Per Malmström, for being my clinical mentor. Your never-ending enthusiasm to teach me both science and clinical experience has been very valuable, and you are always the most updated.
To Bo Baldetorp, Dick Killander and Håkan Ohlsson, for running the Department of Oncology and creating a scientifically stimulating work atmosphere. Bo Baldetorp, for support and help during the years, and your optimism.
My research collaborator, Ruedi Aebersold, professor at ETH, Zürich, Switzerland, for giving me the opportunity to work with you and your group both in Seattle and in Zürich, and a special thanks for great skiing with you!
To all co-authors for stimulating and fruitful collaborations.
To Morten Krogh, Cecilia Ritz and Patrik Edén at the Department of Theoretical Physics for interesting discussions and letting me be part of your abstract world.
To Marie-Louise Fjällskog and Anthoula Koliadi, at the Department of Oncology in Uppsala, for introducing me to cyclin B1.
Für Ruth Hüttenhain, für eine gute Zeit in Zürich mit dir, und vor allem, dass du immer Deutsch mit mir gesprochen hast, du warst mein Stern an der ETH!

To Reto Ossola, for good times both in Seattle and Zürich, and for introducing me to Chäsfondue.
To Susanne André, for being the best coordinator, and for great help with computer problems.
To Carina Strand, for sharing your experience in the lab and technical support. And being a good friend.
To my room-mates, Karin Rennstam and Susanne Magnusson, for lighting up even a bad day, and for your friendship.
To past and present PhD students, for interesting discussions and non-scientific chats and laughs throughout the years!
To my clinical colleagues, for being wonderful doctors and good friends.
To all my friends, I am honored to have you as my friends!
To my Malmström family, for letting me have the luxury of having an extra family, and for introducing me into your uncomplicated way of handling life.
To my parents, Ingrid and Anders, for endless encouragement and support for me and my family. You are fantastic role models to which I could never live up to, but I am grateful, and you are wonderful!
To my sister, Frida, Söstra-mi, for being such a colorful person and the best sister in the world!
To Nora and Elly, "Galningen 1 och 2", my two daughters and sunshines, for making me smile when I am angry and for making me realize what is most important in life.
Most importantly, to Johan, my everything. For being the most optimistic person and colleague I have ever met, for putting perspective into my life, and for being the funniest and most loving husband and father. I cannot thank you enough!

I am indebted to the departments which participate in the South, West, South-East Sweden, Stockholm Breast Cancer Group, and Uppsala/Örebro regions for providing samples and clinical follow-up. This work was supported by funds from the Swedish Cancer Society, the Swedish Research Council, the Swedish Foundation for Strategic Research, the Gunnar Nilsson Cancer Foundation, the Mrs Berta Kamprad Foundation, the Anna and Edwin Bergers Foundation, the University Hospital of Lund Research Foundation, the Knut Alice Wallenberg Foundation through the Swegene consortium, the Strategic Science Foundation CREATE Health Centre, the Skane County Council's Research and Development Foundation, the Governmental Funding of Clinical Research within the National Health Service, the John and Augusta Persson Foundation for Medical Science, the Royal Physiographic Society in Lund, the Maggie Stephens Foundation, the Lennanders Foundation, for travel stipends from AstraZeneca, Roche, Bristol-Myers Squibb and The Swedish Society for Medical research.

Very last, I would like to thank the patients who participated in this work.

References

1. Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. Lancet. May 14-20 2005;365(9472):1687-1717.
2. Goldhirsch A, Ingle JN, Gelber RD et al. hresholds for therapies: highlights of the St Gallen International Expert Concensus on the Primary Therapy of Early Breast Cancer 2009, Ann Oncol. Jun 17 2009; 20(8):1319-29.
3. Look MP, van Putten WL, Duffy MJ, et al. Pooled analysis of prognostic impact of urokinase-type plasminogen activator and its inhibitor PAI-1 in 8377 breast cancer patients. J Natl Cancer Inst. Jan 16 2002;94(2):116-128.
4. Harris L, Fritsche H, Mennel R, et al. American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. J Clin Oncol. Nov 20 2007;25(33):5287-5312.
5. Miller LD, Liu ET. Expression genomics in breast cancer research: microarrays at the crossroads of biology and medicine. Breast Cancer Res. 2007;9(2):206.
6. 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. Dec 19 2002;347(25):1999-2009.
7. Wang Y, Klijn JG, Zhang Y, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet. Feb 19-25 2005;365(9460):671-679.
8. 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. Dec 30 2004;351(27):2817-2826.
9. 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. Sep 2 2003;100(18):10393-10398.
10. van 't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature. Jan 31 2002;415(6871):530-536.
11. Cheng SH, Horng CF, West M, et al. Genomic prediction of locoregional recurrence after mastectomy in breast cancer. J Clin Oncol. Oct 1 2006;24(28):4594-4602.
12. Nuyten DS, Kreike B, Hart AA, et al. Predicting a local recurrence after breast-conserving therapy by gene expression profiling. Breast Cancer Res. 2006;8(5):R62.

13. 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. Aug 2 2003;362(9381):362-369.
14. Gruvberger S, Ringner M, Chen Y, et al. Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns. Cancer Res. Aug 15 2001;61(16):5979-5984.
15. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. Nature. Aug 17 2000;406(6797):747-752.
16. Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A. Sep 11 2001;98(19):10869-10874.
17. Somiari RI, Sullivan A, Russell S, et al. High-throughput proteomic analysis of human infiltrating ductal carcinoma of the breast. Proteomics. Oct 2003;3(10):1863-1873.
18. Gharbi S, Gaffney P, Yang A, et al. Evaluation of two-dimensional differential gel electrophoresis for proteomic expression analysis of a model breast cancer cell system. Mol Cell Proteomics. Feb 2002;1(2):91-98.
19. Hondermarck H, Tastet C, El Yazidi-Belkoura I, Toillon RA, Le Bourhis X. Proteomics of breast cancer: the quest for markers and therapeutic targets. J Proteome Res. Apr 2008;7(4):1403-1411.
20. Umar A, Luider TM, Foekens JA, Pasa-Tolic L. NanoLC-FT-ICR MS improves proteome coverage attainable for approximately 3000 laser-microdissected breast carcinoma cells. Proteomics. Jan 2007;7(2):323-329.
21. Punglia RS, Morrow M, Winer EP, Harris JR. Local therapy and survival in breast cancer. N Engl J Med. Jun 7 2007;356(23):2399-2405.
22. Veronesi U, Cascinelli N, Mariani L, et al. Twenty-year follow-up of a randomized study comparing breast-conserving surgery with radical mastectomy for early breast cancer. N Engl J Med. Oct 17 2002;347(16):1227-1232.
23. Clarke M, Collins R, Darby S, et al. Effects of radiotherapy and of differences in the extent of surgery for early breast cancer on local recurrence and 15-year survival: an overview of the randomised trials. Lancet. Dec 17 2005;366(9503):2087-2106.
24. Touboul E, Buffat L, Belkacemi Y, et al. Local recurrences and distant metastases after breast-conserving surgery and radiation therapy for early breast cancer. Int J Radiat Oncol Biol Phys. Jan 1 1999;43(1):25-38.
25. Fortin A, Larochelle M, Laverdiere J, Lavertu S, Tremblay D. Local failure is responsible for the decrease in survival for patients with breast cancer treated with conservative surgery and postoperative radiotherapy. J Clin Oncol. Jan 1999;17(1):101-109.
26. Voogd AC, Nielsen M, Peterse JL, et al. Differences in risk factors for local and distant recurrence after breast-conserving therapy or mastectomy for stage I and II

breast cancer: pooled results of two large European randomized trials. J Clin Oncol. Mar 15 2001;19(6):1688-1697.

27. Fredriksson I, Liljegren G, Palm-Sjovall M, et al. Risk factors for local recurrence after breast-conserving surgery. Br J Surg. Sep 2003;90(9):1093-1102.
28. Leopold KA, Recht A, Schnitt SJ, et al. Results of conservative surgery and radiation therapy for multiple synchronous cancers of one breast. Int J Radiat Oncol Biol Phys. Jan 1989;16(1):11-16.
29. Kurtz JM, Jacquemier J, Amalric R, et al. Breast-conserving therapy for macroscopically multiple cancers. Ann Surg. Jul 1990;212(1):38-44.
30. Harris EE, Hwang WT, Seyednejad F, Solin LJ. Prognosis after regional lymph node recurrence in patients with stage I-II breast carcinoma treated with breast conservation therapy. Cancer. Nov 15 2003;98(10):2144-2151.
31. Maki DD, Grossman RI. Patterns of disease spread in metastatic breast carcinoma: influence of estrogen and progesterone receptor status. AJNR Am J Neuroradiol. Jun-Jul 2000;21(6):1064-1066.
32. Dent R, Hanna WM, Trudeau M, Rawlinson E, Sun P, Narod SA. Pattern of metastatic spread in triple-negative breast cancer. Breast Cancer Res Treat. Jun 10 2008.
33. Hayes DF, Trock B, Harris AL. Assessing the clinical impact of prognostic factors: when is "statistically significant" clinically useful? Breast Cancer Res Treat. 1998;52(1-3):305-319.
34. Fisher B, Bauer M, Wickerham DL, et al. Relation of number of positive axillary nodes to the prognosis of patients with primary breast cancer. An NSABP update. Cancer. Nov 1 1983;52(9):1551-1557.
35. Karlsson P, Cole BF, Price KN, et al. The role of the number of uninvolved lymph nodes in predicting locoregional recurrence in breast cancer. J Clin Oncol. May 20 2007;25(15):2019-2026.
36. 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. Nov 1991;19(5):403-410.
37. Haybittle JL, Blamey RW, Elston CW, et al. A prognostic index in primary breast cancer. Br J Cancer. Mar 1982;45(3):361-366.
38. Blamey RW, Davies CJ, Elston CW, Johnson J, Haybittle JL, Maynard PV. Prognostic factors in breast cancer -- the formation of a prognostic index. Clin Oncol. Sep 1979;5(3):227-236.
39. Romond EH, Perez EA, Bryant J, et al. Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. N Engl J Med. Oct 20 2005;353(16):1673-1684.
40. Janicke F, Prechtl A, Thomssen C, et al. Randomized adjuvant chemotherapy trial in high-risk, lymph node-negative breast cancer patients identified by urokinase-type

plasminogen activator and plasminogen activator inhibitor type 1. J Natl Cancer Inst. Jun 20 2001;93(12):913-920.

41. 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. Apr 22 1997;74(2):156-161.
42. Washtien WL. Increased levels of thymidylate synthetase in cells exposed to 5-fluorouracil. Mol Pharmacol. Jan 1984;25(1):171-177.
43. Clark JL, Berger SH, Mittelman A, Berger FG. Thymidylate synthase gene amplification in a colon tumor resistant to fluoropyrimidine chemotherapy. Cancer Treat Rep. Mar 1987;71(3):261-265.
44. 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. May 5 1994;330(18):1260-1266.
45. 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(4):486-493.
46. el-Deiry WS. Role of oncogenes in resistance and killing by cancer therapeutic agents. Curr Opin Oncol. Jan 1997;9(1):79-87.
47. 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. Nov 1996;74(9):1458-1465.
48. Maxwell PJ, Longley DB, Latif T, et al. Identification of 5-fluorouracil-inducible target genes using cDNA microarray profiling. Cancer Res. Aug 1 2003;63(15):4602-4606.
49. 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. Nov 15 2003;9(15):5582-5588.
50. Zhang CC, Yang JM, Bash-Babula J, et al. DNA damage increases sensitivity to vinca alkaloids and decreases sensitivity to taxanes through p53-dependent repression of microtubule-associated protein 4. Cancer Res. Aug 1 1999;59(15):3663-3670.
51. Tinari N, Lattanzio R, Natoli C, et al. Changes of topoisomerase IIalpha expression in breast tumors after neoadjuvant chemotherapy predicts relapse-free survival. Clin Cancer Res. Mar 1 2006;12(5):1501-1506.
52. Rouzier R, Rajan R, Wagner P, et al. Microtubule-associated protein tau: a marker of paclitaxel sensitivity in breast cancer. Proc Natl Acad Sci U S A. Jun 7 2005;102(23):8315-8320.
53. Malmstrom P, Bendahl PO, Boiesen P, Brunner N, Idvall I, Ferno M. S-phase fraction and urokinase plasminogen activator are better markers for distant recurrences than Nottingham Prognostic Index and histologic grade in a prospective

study of premenopausal lymph node-negative breast cancer. J Clin Oncol. Apr 1 2001;19(7):2010-2019.

54. Smits VA, Medema RH. Checking out the G(2)/M transition. Biochim Biophys Acta. May 28 2001;1519(1-2):1-12.
55. Nimeus-Malmstrom E, Ritz C, Eden P, et al. Gene expression profilers and conventional clinical markers to predict distant recurrences for premenopausal breast cancer patients after adjuvant chemotherapy. Eur J Cancer. Nov 2006;42(16):2729-2737.
56. Eden P, Ritz C, Rose C, Ferno M, Peterson C. "Good Old" clinical markers have similar power in breast cancer prognosis as microarray gene expression profilers. Eur J Cancer. Aug 2004;40(12):1837-1841.
57. Fisher B, Anderson S, Bryant J, et al. Twenty-year follow-up of a randomized trial comparing total mastectomy, lumpectomy, and lumpectomy plus irradiation for the treatment of invasive breast cancer. N Engl J Med. Oct 17 2002;347(16):1233-1241.
58. Bartelink H, Horiot JC, Poortmans P, et al. Recurrence rates after treatment of breast cancer with standard radiotherapy with or without additional radiation. N Engl J Med. Nov 8 2001;345(19):1378-1387.
59. Vrieling C, Collette L, Fourquet A, et al. The influence of the boost in breast-conserving therapy on cosmetic outcome in the EORTC "boost versus no boost" trial. EORTC Radiotherapy and Breast Cancer Cooperative Groups. European Organization for Research and Treatment of Cancer. Int J Radiat Oncol Biol Phys. Oct 1 1999;45(3):677-685.
60. Whelan T, MacKenzie R, Julian J, et al. Randomized trial of breast irradiation schedules after lumpectomy for women with lymph node-negative breast cancer. J Natl Cancer Inst. Aug 7 2002;94(15):1143-1150.
61. Liljegren G, Lindgren A, Bergh J, Nordgren H, Tabar L, Holmberg L. Risk factors for local recurrence after conservative treatment in stage I breast cancer. Definition of a subgroup not requiring radiotherapy. Ann Oncol. Mar 1997;8(3):235-241.
62. Szumiel I. Intrinsic radiation sensitivity: cellular signaling is the key. Radiat Res. Mar 2008;169(3):249-258.
63. Jameel JK, Rao VS, Cawkwell L, Drew PJ. Radioresistance in carcinoma of the breast. Breast. Dec 2004;13(6):452-460.
64. Randomized trial of two versus five years of adjuvant tamoxifen for postmenopausal early stage breast cancer. Swedish Breast Cancer Cooperative Group. J Natl Cancer Inst. Nov 6 1996;88(21):1543-1549.
65. Tamoxifen for early breast cancer: an overview of the randomised trials. Early Breast Cancer Trialists' Collaborative Group. Lancet. May 16 1998;351(9114):1451-1467.

66. Baum M, Budzar AU, Cuzick J, et al. Anastrozole alone or in combination with tamoxifen versus tamoxifen alone for adjuvant treatment of postmenopausal women with early breast cancer: first results of the ATAC randomised trial. Lancet. Jun 22 2002;359(9324):2131-2139.

67. Polychemotherapy for early breast cancer: an overview of the randomised trials. Early Breast Cancer Trialists' Collaborative Group. Lancet. Sep 19 1998;352(9132):930-942.

68. De Laurentiis M, Cancello G, D'Agostino D, et al. Taxane-based combinations as adjuvant chemotherapy of early breast cancer: a meta-analysis of randomized trials. J Clin Oncol. Jan 1 2008;26(1):44-53.

69. Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. Science. Jan 9 1987;235(4785):177-182.

70. Paik S, Hazan R, Fisher ER, et al. Pathologic findings from the National Surgical Adjuvant Breast and Bowel Project: prognostic significance of erbB-2 protein overexpression in primary breast cancer. J Clin Oncol. Jan 1990;8(1):103-112.

71. Owens MA, Horten BC, Da Silva MM. HER2 amplification ratios by fluorescence in situ hybridization and correlation with immunohistochemistry in a cohort of 6556 breast cancer tissues. Clin Breast Cancer. Apr 2004;5(1):63-69.

72. Press MF, Sauter G, Bernstein L, et al. Diagnostic evaluation of HER-2 as a molecular target: an assessment of accuracy and reproducibility of laboratory testing in large, prospective, randomized clinical trials. Clin Cancer Res. Sep 15 2005;11(18):6598-6607.

73. Ryden L, Haglund M, Bendahl PO, et al. Reproducibility of human epidermal growth factor receptor 2 analysis in primary breast cancer - A national survey performed at pathology departments in Sweden. Acta Oncol. Apr 7 2009:1-7.

74. Piccart-Gebhart MJ, Procter M, Leyland-Jones B, et al. Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. N Engl J Med. Oct 20 2005;353(16):1659-1672.

75. Bergh J, Jonsson PE, Glimelius B, Nygren P. A systematic overview of chemotherapy effects in breast cancer. Acta Oncol. 2001;40(2-3):253-281.

76. Palangie T, Mosseri V, Mihura J, et al. Prognostic factors in inflammatory breast cancer and therapeutic implications. Eur J Cancer. 1994;30A(7):921-927.

77. Geyer CE, Forster J, Lindquist D, et al. Lapatinib plus capecitabine for HER2-positive advanced breast cancer. N Engl J Med. Dec 28 2006;355(26):2733-2743.

78. Burstein HJ, Elias AD, Rugo HS, et al. Phase II study of sunitinib malate, an oral multitargeted tyrosine kinase inhibitor, in patients with metastatic breast cancer previously treated with an anthracycline and a taxane. J Clin Oncol. Apr 10 2008;26(11):1810-1816.

79. Portera CC, Walshe JM, Rosing DR, et al. Cardiac toxicity and efficacy of trastuzumab combined with pertuzumab in patients with [corrected] human epidermal growth factor receptor 2-positive metastatic breast cancer. Clin Cancer Res. May 1 2008;14(9):2710-2716.

80. Dodwell D, Coombes G, Bliss JM, Kilburn LS, Johnston S. Combining Fulvestrant (Faslodextrade mark) with Continued Oestrogen Suppression in Endocrine-sensitive Advanced Breast Cancer: the SoFEA Trial. Clin Oncol (R Coll Radiol). Apr 1 2008.

81. Malmstrom P, Holmberg L, Anderson H, et al. Breast conservation surgery, with and without radiotherapy, in women with lymph node-negative breast cancer: a randomised clinical trial in a population with access to public mammography screening. Eur J Cancer. Aug 2003;39(12):1690-1697.

82. Fredriksson I, Liljegren G, Arnesson LG, et al. Time trends in the results of breast conservation in 4694 women. Eur J Cancer. Aug 2001;37(12):1537-1544.

83. Strand C, Enell J, Hedenfalk I, Ferno M. RNA quality in frozen breast cancer samples and the influence on gene expression analysis--a comparison of three evaluation methods using microcapillary electrophoresis traces. BMC Mol Biol. 2007;8:38.

84. Schroeder A, Mueller O, Stocker S, et al. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. BMC Mol Biol. 2006;7:3.

85. Elston C, Ellis I. Assessment of histological grade. In Systemic Pahology. 3rd Ed. ed. Edinburgh, London, New York, Philadelphia, San Francisco, Sydney, Toronto: Churchill Livingstone; 1998.

86. Dowsett M, Goldhirsch A, Hayes DF, Senn HJ, Wood W, Viale G. International Web-based consultation on priorities for translational breast cancer research. Breast Cancer Res. 2007;9(6):R81.

87. Kuhling H, Alm P, Olsson H, et al. Expression of cyclins E, A, and B, and prognosis in lymph node-negative breast cancer. J Pathol. Apr 2003;199(4):424-431.

88. Rudolph P, Kuhling H, Alm P, et al. Differential prognostic impact of the cyclins E and B in premenopausal and postmenopausal women with lymph node-negative breast cancer. Int J Cancer. Jul 10 2003;105(5):674-680.

89. Suzuki T, Urano T, Miki Y, et al. Nuclear cyclin B1 in human breast carcinoma as a potent prognostic factor. Cancer Sci. May 2007;98(5):644-651.

90. Aaltonen K, Amini RM, Heikkila P, et al. High cyclin B1 expression is associated with poor survival in breast cancer. Br J Cancer. Apr 7 2009;100(7):1055-1060.

91. Wirapati P, Sotiriou C, Kunkel S, et al. Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. Breast Cancer Res. 2008;10(4):R65.

92. Boiesen P, Bendahl PO, Anagnostaki L, et al. Histologic grading in breast cancer--reproducibility between seven pathologic departments. South Sweden Breast Cancer Group. Acta Oncol. 2000;39(1):41-45.

93. 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. May 14-20 2005;365(9472):1687-1717.

94. Finak G, Bertos N, Pepin F, et al. Stromal gene expression predicts clinical outcome in breast cancer. Nat Med. May 2008;14(5):518-527.

95. Farmer P, Bonnefoi H, Anderle P, et al. A stroma-related gene signature predicts resistance to neoadjuvant chemotherapy in breast cancer. Nat Med. Jan 2009;15(1):68-74.

96. Specht K, Harbeck N, Smida J, et al. Expression profiling identifies genes that predict recurrence of breast cancer after adjuvant CMF-based chemotherapy. Breast Cancer Res Treat. Oct 17 2008.

97. Acharya CR, Hsu DS, Anders CK, et al. Gene expression signatures, clinicopathological features, and individualized therapy in breast cancer. Jama. Apr 2 2008;299(13):1574-1587.

98. Gygi SP, Rochon Y, Franza BR, Aebersold R. Correlation between protein and mRNA abundance in yeast. Mol Cell Biol. Mar 1999;19(3):1720-1730.

99. Zhang H, Li XJ, Martin DB, Aebersold R. Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. Nat Biotechnol. Jun 2003;21(6):660-666.

100. Malmstrom J, Lee H, Aebersold R. Advances in proteomic workflows for systems biology. Curr Opin Biotechnol. Aug 2007;18(4):378-384.

101. Hager JW, Yves Le Blanc JC. Product ion scanning using a Q-q-Q linear ion trap (Q TRAP) mass spectrometer. Rapid Commun Mass Spectrom. 2003;17(10):1056-1064.

102. Chang HY, Sneddon JB, Alizadeh AA, et al. Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. PLoS Biol. Feb 2004;2(2):E7.

103. Teschendorff AE, Naderi A, Barbosa-Morais NL, et al. A consensus prognostic gene expression classifier for ER positive breast cancer. Genome Biol. 2006;7(10):R101.

104. Teschendorff AE, Miremadi A, Pinder SE, Ellis IO, Caldas C. An immune response gene expression module identifies a good prognosis subtype in estrogen receptor negative breast cancer. Genome Biol. 2007;8(8):R157.

105. Bartelink H, Horiot JC, Poortmans PM, et al. Impact of a higher radiation dose on local control and survival in breast-conserving therapy of early breast cancer: 10-year results of the randomized boost versus no boost EORTC 22881-10882 trial. J Clin Oncol. Aug 1 2007;25(22):3259-3265.

106. Goldstein LJ, Gray R, Badve S, et al. Prognostic utility of the 21-gene assay in hormone receptor-positive operable breast cancer compared with classical clinicopathologic features. J Clin Oncol. Sep 1 2008;26(25):4063-4071.

107. Mamounas EP, Tang G, Bryant J, et al. Association between the 21-gene recurrence score assay (RS) and risk of locoregional failure in node-negative, ER-positive breast cancer: results from NSABP B-14 and NSABP B-20. 28th Annual San Antonio Breast Cancer Symposium. San Antonio; 2005.

108. Chi JT, Wang Z, Nuyten DS, et al. Gene expression programs in response to hypoxia: cell type specificity and prognostic significance in human cancers. PLoS Med. Mar 2006;3(3):e47.

109. Liu R, Wang X, Chen GY, et al. The prognostic role of a gene signature from tumorigenic breast-cancer cells. N Engl J Med. Jan 18 2007;356(3):217-226.

110. Lauss M, Kriegner A, Vierlinger K, et al. Consensus genes of the literature to predict breast cancer recurrence. Breast Cancer Res Treat. Jul 2008;110(2):235-244.

111. Miller K, Wang M, Gralow J, et al. Paclitaxel plus bevacizumab versus paclitaxel alone for metastatic breast cancer. N Engl J Med. Dec 27 2007;357(26):2666-2676.

Paper I

EUROPEAN JOURNAL OF CANCER 42 (2006) 2729–2737

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejconline.com

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

Article history:
Received 18 February 2006
Received in revised form
30 May 2006
Accepted 2 June 2006
Available online 4 October 2006

Keywords:
Breast cancer
cDNA microarray
Drug resistance
Prognostic markers

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.

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

* *Corresponding author.*
E-mail address: marten.ferno@med.lu.se (M. Fernö).
[d] These authors have contributed equally.
0959-8049/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ejca.2006.06.031

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 $<3.0 \times 10^9$/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-

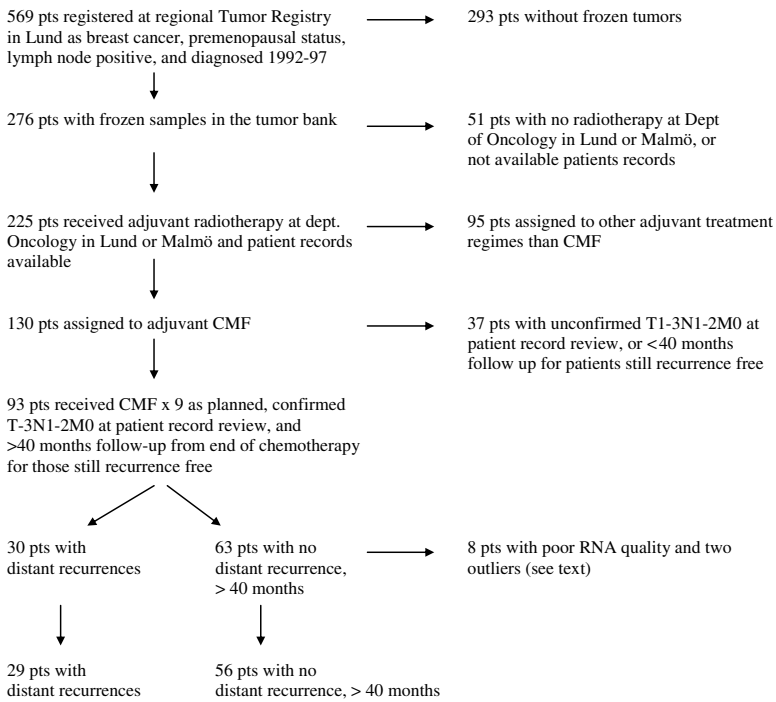


Fig. 1 – The selection of patients included in the study.

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 × 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 outcome 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, 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).

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

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

This study was supported by Swedish Cancer Society, Gunnar, Arvid and Elisabeth Nilsson Foundation, Mrs Berta Kamprad Foundation, Swedish Research Council and the Swedish Foundation for Strategic Research through the Lund Center for Stem Cell Biology and Cell Therapy. We are also indebted to participating departments of surgery, oncology, and pathology of the South Sweden Breast Cancer Group, and Regional Tumour Registry for primary care and follow-up of the patients, and for handling of frozen breast cancer tissue.

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.

Paper II

Available online at www.sciencedirect.com
ScienceDirect
Journal of Pharmaceutical and Biomedical Analysis 43 (2007) 1086–1093
JOURNAL OF PHARMACEUTICAL AND BIOMEDICAL ANALYSIS
www.elsevier.com/locate/jpba

Proteomic analysis identifies candidate proteins associated with distant recurrences in breast cancer after adjuvant chemotherapy

Emma Niméus [a], Johan Malmström [b], Anders Johnsson [a], György Marko-Varga [c], Mårten Fernö [a,*]

[a] *Department of Oncology, Clinical Sciences, University Hospital, Lund, Sweden*
[b] *Department of Cell and Molecular Biology, BMC C13, Lund, Sweden*
[c] *Department of Analytical Chemistry and AstraZeneca R&D, Lund, Sweden*

Received 17 March 2006; received in revised form 10 September 2006; accepted 11 September 2006
Available online 7 November 2006

Abstract

Breast cancer is a heterogenous disease and it is of importance to select patients with regard to different prognosis and treatment sensitivity to individualize treatment regimes. In this study we successfully adapted a protein extraction protocol from mRNA extracted tumor samples enabling two-dimensional gel electrophoresis (2-DE) analysis of samples previously analyzed by cDNA microarray. The aim was to find candidate proteins that distinguish breast cancer patients with or without recurrences after adjuvant CMF (cyclophosphamide, methotrexate and 5-FU) treatment within four years to follow-up. We identified several proteins distinguishing the recurrence group from the non-recurrence group, especially in the ER and PgR positive subgroup ($n = 7$). The induced proteins were involved in translation/folding, iron ion binding, and protease inhibition, whereas proteins involved in signaling, ubiquitination, and splicing were decreased in expression. These results show that it is possible to use 2-DE to separate high abundant proteins in breast cancer tissue and to find discriminating proteins to identify patients with different prognosis after adjuvant CMF treatment.

Keywords: Breast cancer; Two-dimensional gel electrophoresis; Drug resistance; Prognostic markers

1. Introduction

Breast cancer is the most common malignancy among women in the Western world, affecting approximately every tenth woman. After the primary local treatment the patients are typically divided into risk groups based on prognostic factors, such as stage (tumor size, lymph node status, and metastases), histological grade, age, and estrogen (ER) and progesterone receptor (PgR) status. Markers of proliferation, i.e. S-phase fraction (SPF), and invasive factors within the urokinase plasminogen activator system are sometimes also used. Based on these prognostic factors, patients with a high risk of relapse receive adjuvant systemic therapy, either cytotoxic, endocrine, monoclonal antibodies and/or combination. Examples of adjuvant cytotoxic treatments are CMF (cyclophosphamide, methotrexate and 5-fluorouracil), antracyclin- and taxane-based drug combinations. The overall positive effect of adjuvant cytotoxic therapy is limited with only an increased survival of approximately 10% [1]. The remaining patients are either already cured by the primary local treatment or recur in spite of the treatment given and thus do not benefit from the adjuvant cytotoxic therapy. Possible mechanisms for recurrence despite treatment are low initial drug sensitivity or an acquired drug resistance, which are common clinical problems in cancer treatment.

Useful markers for chemotherapy resistance and/or sensitivity have not so far successfully been found, even though some markers show promising results in a limited number of studies, such as thymidylate synthase, thymidine kinase [2–4], c-erbB-2 [5], multidrug resistance-associated protein [6], and

Abbreviations: CMF, cyclophosphamide, methotrexate, 5-fluorouracil; 2-DE, two-dimensional gel electrophoresis; ER, estrogen receptor; PgR, progesterone receptor; SPF, S-phase fraction; ECM, extracellular matrix

* Corresponding author at: Department of Oncology, Clinical Sciences, Lund University Hospital, SE 221 85 Lund, Sweden. Tel.: +46 46 17 75 65; fax: +46 46 14 73 27.
E-mail address: Marten.Ferno@med.lu.se (M. Fernö).

0731-7085/$ – see front matter © 2006 Published by Elsevier B.V.
doi:10.1016/j.jpba.2006.09.019

p53 [7–10]. The development of gene expression analyses and techniques within proteomics enables extensive characterization of malignant tumors, which may help us to understand treatment resistance and/or treatment sensitivity. Gene expression studies on mRNA level have shown to be able to detect differences between sporadic and hereditary breast cancer [11], and between ER positive and ER negative breast cancer [12].

Promising results for predicting clinical outcome have also been obtained [13–16]. However, several aspects in tumor biology cannot be captured by gene expression analysis only, such as protein expression levels, protein degradation and posttranslational modifications, emphasizing the need for complementary analysis at the protein level. Proteomic studies using two-dimensional electrophoresis (2-DE) analysis of breast cancer have found differences between ductal carcinoma and non-neoplastic tissue [17], the identification of proteins associated with c-erbB-2-expression [18], and evaluation of certain known prognostic factors [19]. In other malignancies (ovarian, prostate, vaginal, and cervical cancer), 2-DE has been used to discriminate between normal/benign and cancer tissue [20–22]. Chemotherapy resistance has also been studied in 2-DE using cell lines from melanoma [23]. To achieve more effective chemotherapeutic treatment of breast cancer patients it is essential to define reliable indicators of response to treatment in individual patients and to establish which mechanisms are responsible for drug resistance. In this study, our aim was to identify proteins that can be used to distinguish tumors from patients later developing distant recurrences after adjuvant CMF from patients without distant recurrence during the follow-up period.

2. Method and patients

2.1. Patients

According to treatment guidelines in the Regional Care program for breast cancer in Southern Sweden issued in 1991, premenopausal lymph node positive (N+) breast cancer patients were recommended radiotherapy and postoperative adjuvant chemotherapy. Patients in the present study were selected in a stepwise manner to fulfill the following criteria: premenopausal women with primary breast carcinoma, stage T1-3N1-2M0, diagnosed 1992–97, for whom frozen primary tumor samples were still available, referred to the Department of Oncology in Lund or Malmö for adjuvant radiotherapy, treatment with nine cycles of CMF, and either distant recurrence within 40 months after completion of CMF or remaining free from distant recurrence for 40 months or longer. This cohort consisted of 85 patients (29 recurrences and 56 recurrence-free patients). Out of these, 20 patients were selected based on recurrence status and ER/PgR status, thus making up four groups with five patients in each: (1) distant recurrence and ER−/PgR−; (2) distant recurrence and ER+/PgR+; (3) no distant recurrence and ER−/PgR−; and (4) no distant recurrence and ER+/PgR+. The study was approved by the ethics committee at Lund University. The data from time to recurrence and the conventional clinical markers (e.g. ER, PgR, SPF, DNA ploidy status, histological grade, histological type, tumor size, number of tumor-involved lymph nodes, age at diagnosis, and location of distant recurrence) are summarized in Table 1.

2.2. Treatment

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 9 cycles.

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 for five weeks. During the five-week radiotherapy session, cyclophosphamide was given at a dose of 850 mg/m^2 every three weeks, while methotrexate and 5-fluorouracil were omitted.

2.3. Methods

2.3.1. Conventional clinical markers

ER and PgR were analyzed 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.

Flow cytometric DNA analysis was also performed routinely at the time of the primary operation in an Ortho Cytoron Absolute flow cytometer (Ortho Diagnostic Systems, Raritan, NJ, USA). Ploidy status was defined as follows: one DNA cell population is equal to diploid and two or more cell populations are equal to non-diploid. Samples with an SPF $\geq$ 12% were classified as high SPF, and those samples with values below these levels as low SPF [24]. Histological grade was re-evaluated for all the samples by the same observer according to Elston and Ellis [25]. 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 to 3 points. The overall histological 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. Histological type was re-evaluated according to WHO [26].

2.3.2. Protein isolation for 2-DE

The tumor tissue was obtained from the tumor bank at the Department of Oncology, consisting of residual tumor samples after routine analyses of ER, PgR, DNA ploidy status, and SPF. From this tissue, the mRNA pool was isolated from the top layer of a Trizol extraction. The layers beneath the mRNA pool (interphase and organic phase) contained the extracted proteins. The DNA was precipitated from the interphase and organic phase with 40% ethanol without precipitating the proteins, and the proteins were then precipitated from the supernate with isopropyl alcohol. The supernatant was removed and the protein pellet was washed in 0.3 M guanidine hydrochloride in 95% ethanol followed by a final wash in 75% ethanol. Extensive washing proved to be necessary to remove interfering substances from the protein pool, such as lipids and large insoluble particles. The protocol for protein extraction was optimized using only

Table 1
Conventional clinical parameters for 20 breast cancer patients, subdivided as follows

Time[a]	ER[a]	PgR[a]	Lymph nodes[a]	SPF[a]	Ploidy[a]	Size[a]	Hist grade[a]	Hist type	Age[a]	Rec.-location
(A) Group 1, distant recurrences, ER/PgR negative										
14	0.9	1.5	1	13	Non-dip	25	3	Ductal-UNS	47	Retina/lungs
−3	1.9	3.3	2	8.6	Non-dip	35	3	Ductal-UNS	48	Lungs/liver
26	20	12	1	30	Non-dip	22	3	Ductal-UNS	43	Lungs/bone
35	0	0	21	7.3	Non-dip	21	3	Tubuloductal	52	Pleura
10	1.4	0	21	23	Non-dip	15	3	Ductal-medullar	37	CNS
14[b]	*1.4*	*1.5*	*2*	*13*		*22*	3		*47*	
(B) Group 2, distant recurrences, ER/PgR positive										
19	89	250	6	5.8	Diploid	37	3	Ductal-UNS	45	Bone
30	200	280	12	24	Non-dip	50	2	Lobular	49	Bone
38	42	100	1	6.2	Diploid	11	1	Tubuloductal	37	Liver/bone
16	47	150	2	14	Non-dip	21	3	Ductal-UNS	45	Liver
30	160	26	2	18	Diploid	25	3	Ductal-UNS	46	Bone/pleura/liver
30	*89*	*150*	*2*	*14*		*25*	*3*		*45*	
(C) Group 3, no distant recurrences, ER/PgR negative										
55	6.6	6.5	5	16	Non-dip	35	3	Ductal-UNS	50	
94	0.7	1.3	5	21	Non-dip	18	3	Ductal-medullar	41	
58	0	0	2	21	Non-dip	36	3	Ductal-UNS	48	
69	1.1	2.8	2	28	Non-dip	22	3	Ductal-UNS	48	
58	0	0	2	14	Diploid	25	3	Ductal-medullar	46	
58	*0.7*	*1.3*	*2*	*21*		*25*	*3*		*48*	
(D) Group 4, no distant recurrences, ER/PgR positive										
54	210	340	2	4.6	Non-dip	15	2	Ductal-UNS	49	
85	190	1300	8	8.4	Non-dip	21	3	Ductal-UNS	50	
55	100	330	1	9	Non-dip	12	1	Tubuloductal	50	
59	350	420	2	12	Non-dip	21	3	Ductal-UNS	50	
82	210	370	5	2.8	Diploid	20	2	Ductal-UNS	48	
59	*210*	*370*	*2*	*8.4*		*20*	*2*		*50*	

Time is the number of months to recurrence, evaluated from the day when the CMF treatment was accomplished, or follow-up time for the patients with no recurrences. The median for each parameter is also calculated and shown in italics.

[a] =Time to recurrence (0 = after 6 months' treatment) or follow-up time for patients in the non-recurrence groups (months), ER and PgR (fmol/mg protein), lymph nodes (number of tumor-involved lymph nodes, SPF (%), ploidy (diploid population, non-diploid population), size (tumor size, mm), hist grade (histological grade 1–3), hist type (histological type), age (age at diagnosis, years).

[b] Median value.

15% of the mRNA extracted leftover, resulting in the possibility of running multiple gels from the same sample.

2.3.3. Sample preparations and gel electrophoresis

Immobiline Dry strips (180 mm, pH 4–7, non-linear) were rehydrated in 350 μl of the solubilization solution containing 8 M urea and 2% CHAPS (3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate), 10 mM DTT (dithiothreitol), and 0.5% immobilized pH gradient (IPG) 4–7 buffer. The isoelectrophofocusing (IEF) step was performed at 20 °C in an IPGphor™ (Amersham Pharmacia Biotech, Uppsala, Sweden) and run according to the following gradient schedule: (1) 0–300 V for 1 min; (2) 300–3500 V for 1.5 h; (3) 3500 V until approximately 45,000 V h were reached. The strips were equilibrated for 10 min in a solution containing 65 mM DTT, 6 M urea, 30% (w/v) glycerol, 2% (w/v) SDS (sodium dodecyl sulfate), and 50 mM Tris–HCl, pH 8.8. A second equilibration step was also carried out for 10 min in the same solution except for DTT, which was replaced by 259 mM iodoacetamide. The strips were soaked in electrophoresis buffer (24 mM Tris base, 0.2 M glycine and 0.1% SDS) just before the molecular weight separation, and applied on 14% homogeneous Duracryl slabgel and overlaid with a solution of 1% agarose in electrophoresis buffer (kept at 60 °C). Electrophoresis was carried out in a Hoefer™ DALT gel apparatus (Amersham Pharmacia Biotech, San Francisco, CA, USA) at 20 °C and constant 100 V for 18 h.

2.3.4. Gel staining and spot analysis

Gels were silver stained [27] and scanned using a Fluor-S™ MultiImager (Bio-Rad Laboratories, Sundbyberg, Sweden) and Quantity One (version 4.0.3, Bio-Rad Laboratories, Sundbyberg, Sweden). Spot analysis was performed using the PDQUEST (version 6.1.0) two-dimensional gel analysis system (Bio-Rad discovery series, Bio-Rad Laboratories, Sundbyberg, Sweden). After spot detection and matching, every spot on the gel was given an integrated optical density (IOD) value by the software program. This value was compared to the total IOD of all of valid spots and thus every spot is shown as ppm (parts per million) of the total IOD of the valid spots. The average spot intensity of every spot on the gels from the early distant recurrences group was compared to the average spot intensity to corresponding spots on the gels from the no recurrences group. The data sets were analyzed using Ludesi Interpreter™, http://www.ludesi.com. The significant differentially expressed

spots were further filtered based on spot quality. Several comparisons were made both including all samples in recurrence and no recurrence group or subdivided into ER/PgR positive or negative subgroups. In addition the ER/PgR positive/negative samples were compared when including all samples as well as after further subdivision on to the recurrence/no recurrence group.

2.3.5. Identification of the protein spots

Thirty-nine spots with a *p*-value less than 0.05 and eleven landmark proteins were sliced out and transferred to small Eppendof tubes and washed three times with a wash-solution (40% acetonitrile, 60 mM ammonium hydrogen carbonate, pH 7.8). The protein spots were dried down in a vacuum concentrator for 15 min and digested with trypsin (Promega Porcine) in 25 mM ammonium bicarbonate and incubated overnight at 37 °C. The digest was stopped by adding 0.2% TFA (trifluoro acetic acid) and Ziptips were used to concentrate and desalt the protein digests according to the manufacture's instructions (Millipore, Bedford, MA, USA). The peptides were thereafter spotted on polished stainless steel target plates together with 7.5 mg/mL α-cyano-4-hydroxycinnamic acid dissolved in 60:40 acetonitrile–water. The MALDI (matrix assisted laser desorption ionization) plates were analyzed in automated mode on the AB4700 Proteomics Analyzer (Applied Biosystems, Framingham, MA) with 1000 laser shots in MS mode and with internal two-point calibration on trypsin peptides with a resulting mass accuracy of <10 ppm. Peaks with a signal-to-noise ratio above 50 passing the exclusion filter of trypsin autolysis peaks and matrix peaks were subjected to MS/MS analysis using up to 3000 laser shots/precursor unless the pre-defined signal-to-noise level in the MS/MS acquisitions was achieved sooner. The MS/MS data were submitted for data base to Mascot (http://www.matrixscience.com/) with a parent mass error tolerance of 50 parts per million and mass fragments with an error tolerance of 0.2 Da.

2.3.6. Statistics of conventional markers

The statistical analysis of the conventional clinical markers was performed in Stata 7.0 (StataCorp. 2001. Stata Statistical Software: Release 7.0. College Station, TX: Stata Corporation). Mann–Whitney *U*-test and Kruskall–Wallis was used to test the null hypothesis of equal distribution in two subgroups. The level of significance was set to 5%.

3. Results

3.1. Description of patient cohort

Twenty patients were selected all with measured clinical makers summarized in Section 2 (see Table 1). Patients were selected to be as similar as possible with regard to the conventional clinical markers to rule out any influence associated with these markers. However, small but statistically significant differences were found in the ER/PgR positive subgroup after comparison recurring to non-recurring tumors. The expression of ER and PgR was higher in the non-rercurring subgroup, (89 vs. 210, $p = 0.047$, 150 vs. 370 $p = 0.009$) also the age at diagnosis was statistically significant lower in the recurring group (45 vs. 50, $p = 0.02$). In the ER/PgR negative subgroup, the corresponding comparisons showed no statistical significant difference (all *p*-values >0.24).

3.2. The development of assay conditions

We then developed an extraction protocol allowing isolation of both mRNA and proteins from the tumor samples. The proteins were separated by 2-DE using both p*I* (isoelectric point) strips 3–10 and 4–7 to determine which p*I* range was the most suitable for analyzing the tumor material. Samples from six patients (three from the recurring group and three from the non-recurring group) were analyzed with both p*I* ranges, and samples from two patients were re-analyzed twice to study the reproducibility of the 2-DE. From these gel sets it was then possible to determine a correlation coefficient, which is a rough estimate of the reproducibility, of both the sample reproducibility within the same patient group, including sample similarity, extraction reproducibility and experimental reproducibility, as well as the reproducibility of experimental protocol from the repetitive analysis of the same sample. A correlation coefficient of 1 equals 100% reproducibility of the expression levels between two sets of sample and we found the intra sample reproducibility to be 0.9 whereas inter sample correlation coefficient was around 0.8 for both p*I* ranges, which is consistent with previously published studies on tumor material [28]. The 2-D gels from samples from the six patients, in both p*I* ranges, matched separately, resulted in approximately 800 matched spots in the 3–10 range (see Supplement 1 for comparison) and 1000 in the 4-7 range (Fig. 1). Since the p*I* range 4–7 contains a higher number of matched spots with better resolution, this p*I* range was used throughout the remaining analysis of the extended study of 20 breast cancer patients.

After spot matching and statistical analysis, spots of interest were analyzed by tandem mass spectrometry and the identified proteins are shown in Fig. 1 and corresponding protein identifications and expression levels in Tables 2–4 (see Supplement 2 for zoom-segments of gel spots of interest). In addition landmarks were identified in order to have reference points concerning the p*I* and molecular weight.

3.3. Recurrences versus no recurrences

Thioredoxin domain containing protein 5 (similar to glucose regulated protein) was significantly increased ($p < 0.05$) in the recurrence group ($n = 10$) compared to the group without recurrences ($n = 10$; Table 2A).

Comparison after subdivision of the recurring ($n = 5$)/no recurring ($n = 5$) group with regard to ER/PgR status, resulted in the identification of seven differentially expressed proteins ($p < 0.05$) from the ER/PgR positive subgroup (Table 2B). Proteins with increased expression in the recurrence group were involved in translation/folding, iron ion binding, and protease inhibitor, whereas those with a lower expression were involved in signaling, ubiquitination, and splicing. Additional

Table 2
Proteins with different expression in breast cancer samples from patients with recurrences vs. no recurrences

Spot#	Acc#	Protein name	Mwt[a]	Score[b]	#pep	Functional group	Ratio
(A) All tumors							
1	Q9BVH9	Thioredoxin domain containing protein 5 (Similar to glucose-regulated protein)	36725	90	2	Unknown	2.4
(B) ER/PgR positive tumors							
2	P09525	ANNEXIN IV	35729	245	5	Signaling	0.6
3	Q14240	EUKARYOTIC INITIATION FACTOR 4A-II	46593	177	6	Translation/protein folding	2.4
4	P15374	UBIQUITIN CARBOXYL-TERMINAL HYDROLASE ISOZYME L3	26337	50	1	Ubiquitination	0.5
5	Q07955	PRE-MRNA SPLICING FACTOR SF2, P33 SUBUNIT	27711	83	2	Splicing	0.2
6	P47813	EUKARYOTIC TRANSLATION INITIATION FACTOR 1A	16433	58	1	Translation/protein folding	5.7
7	P02792	FERRITIN LIGHT CHAIN	19933	89	1	Iron ion binding	6.8
8	P01009	ALPHA-1-ANTITRYPSIN PRECURSOR	46878	72	1	Protease inhibitor	1.1
(C) ER/PgR negative tumors							
9	P08670	VIMENTIN	53579	68	1	Cytoskeletal	2.0
10	P20774	OSTEOINDUCTIVE FACTOR PRECURSOR	34243	44	2	ECM	1.1

The spot number is correlated to the numbers found marked on the gel in Fig. 1. Accession numbers from Swissprot (http://us.expasy.org/sprot/). A score >50 was considered a significant hit by the search engine. Number of peptides is matched peptides to the corresponding protein. The column "functional group" represents the function of the proteins. The ratio is calculated from the recurrence/no recurrence.
[a] Mwt (molecular weight) (Da).
[b] Mascot (www.matrixscience.com).

two proteins were identified in the ER negative subgroup with higher expression in the recurrence group. These proteins are involved in cytoskeletal processes and extracellular matrix (Table 2C).

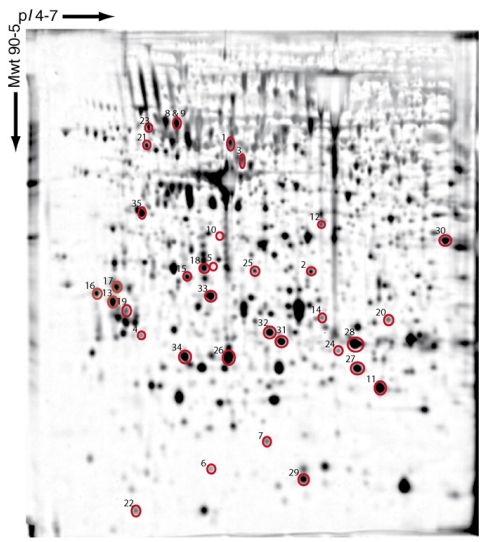


Fig. 1. Proteins were separated by 2-DE and differentially expressed proteins and landmarks are marked with a spot number, corresponding to the spot number in Tables 2–4. Spots number 8 and 9 (alpha-1-antitrypsin precursor and vimentin) were identified as a co-migration in the same spot. The actual gel used in this picture is from a patient with an early distant recurrence and an ER/PgR positive tumor.

3.4. ER/PgR negative versus ER/PgR positive tumors

An apoptosis-associated speck-like protein showed a lower expression in the ER negative group ($n = 10$) than in the ER positive group ($n = 10$; Table 3A).

After subdividing with regard to recurrence status, and in the same way comparing ER/PgR negative and ER/PgR positive tumors, we found six proteins (see Table 3B) with significantly different expression in the recurrence group and eight proteins (see Table 3C) in the non-recurrence group. Proteins with a higher expression in the ER/PgR negative subgroup were found to be involved in translation/protein folding, signaling, and *N*-acetylglucosamine metabolism. Proteins with a lower expression were found to be involved in cytoskeleton, DNA repair, ECM (extracellular matrix), signaling, translation/folding, protease inhibitor, and cytochrome C oxidase.

4. Discussion

Our aim was to identify candidate proteins to predict the clinical outcome after adjuvant CMF treatment. It has previously been shown that tumors with different receptor status have large differences in gene expression patterns [12,14]. In order to obtain more homogeneous groups we therefore divided the total series of 20 patients into 4 subgroups with different combinations of distant recurrence (yes or no) and ER/PgR status (negative or positive). We used an extraction method allowing purification of both mRNA and proteins for the analysis of cDNA microarray and 2-DE. Two different p*I* ranges were investigated in order to establish expression maps with the highest number of uniquely resolved spots. Even though the 3–10 p*I* range is broader, the 4–7 range resulted in a higher number of matched spots, most likely since only few proteins are present in the extreme edges of the 3–10 p*I*, and that an increased separation of the more

Table 3
Proteins with different expression when comparing ER/PgR positive breast cancer samples to ER/PgR negative

Spot#	Acc#	Protein name	Mwt[a]	Score[b]	#pep	Functional group	Ratio
(A) All tumors							
11	Q9ULZ3	APOPTOSIS-ASSOCIATED SPECK-LIKE PROTEIN	21670	76	3	Apoptosis	0.8
(B) Tumors from patients with distant recurrences							
12	Q9BVP0	N-ACETYLGLUCOSAMINE KINASE	37694	152	4	*N*-Acetylglucosamine metabolism	2.0
13	P42655	14-3-3 PROTEIN EPSILON	29155	113	5	Signaling	0.7
14	P30040	ENDOPLASMIC RETICULUM PROTEIN ERP29 PRECURSOR	29032	122	3	Translation/protein folding	2.2
15	Q99426	CYTOSKELETON-ASSOCIATED PROTEIN CKAPI	27594	84	1	Cytoskeletal	0.7
16	P24534	ELONGATION FACTOR 1-BETA	24788	84	2	Translation/protein folding	0.6
17	P07226	TROPOMYOSIN, FIBROBLAST NON-MUSCLE TYPE	28619	376	14	Cytoskeletal	0.5
(C) Tumors from patients with no recurrences							
18	P20774	OSTEOINDUCTIVE FACTOR PRECURSOR	34243	154	4	ECM	0.2
19	P29312	14-3-3 PROTEIN ZETA/DELTA	27899	215	4	Signaling	3.3
20	P30040	ENDOPLASMIC RETICULUM PROTEIN ERP29 PRECURSOR	29054	56	1	Translation/protein folding	1.8
21	P01009	ALPHA-1-ANTITRYPSIN PRECURSOR	46878	56	2	Protease inhibitor	0.3
22	P00167	CYTOCHROME B5	15189	234	3	Cytochrome *c* oxidase activity	0.4
23	P54727	UV EXCISION REPAIR PROTEIN PROTEIN RAD23 HOMOLOG B	43202	75	3	DNA repair	0.6
24	P29354	GROWTH FACTOR RECEPTOR-BOUND PROTEIN 2	25304	61	2	Signaling	1.9
25	Q9H3J8	My027 protein	33554	112	2	Unknown	0.5

The spot number is correlated to the numbers found marked on the gel in Fig. 1. Accession numbers from Swissprot (http://us.expasy.org/sprot/). A score >50 was considered a significant hit by the search engine. Number of peptides is matched peptides to the corresponding protein. The column "functional group" represents the function of the proteins. The ratio is calculated from the ER/PgR negative/ER/PgR positive.

[a] Mwt (molecular weight) (Da).

[b] Mascot (www.matrixscience.com).

crowded area 4–7 interval resulted in a increase of the number of matched spots.

In general, the number of proteins able to distinguish the different sample groups was rather small. However, this is not surprising since only the most highly expressed proteins are detected in the 2D gels and previous analysis of tumor samples have found similar numbers of discriminating proteins [20–22]. One protein, thioredoxin domain containing protein 5 (similar to glucose-regulated protein), was found to be increased in the tumors with distant recurrences, when comparing tumors from patients with recurrences versus. no recurrences. This protein is present in the endoplasmic reticulum lumen, although its function is not known in detail. When the tumors were furthermore subdivided in the ER/PgR positive subgroup, seven proteins showed significant differences in expression levels between the recurrence and non-recurrence group. These proteins were involved in translation/protein folding, splicing, ubiquitination, and iron ion binding (ferritin), which are of importance in tumorgenesis. For example, recent findings have indicated that the ubiquitin conjugation leads to selective degradation of nuclear oncoproteins and suppressor gene products [29,30], and that ferritin is important for proliferation in many different neoplasms [31].

In the ER/PgR negative subgroup fewer differentially expressed proteins were found. This lower number of distinguishing in the ER/PgR negative is consistent with the findings in the comparison of conventional factors between the recurrence and the non-recurrence group. In the ER/PgR positive subgroup, ER, PgR, and age at diagnosis differed, whereas none of the factors differed in the ER/PgR negative subgroup. In another study

Table 4
Landmarks were identified as reference points for the p*I* and molecular weight

Spot#	Acc#	Protein name	Mwt[a]	Score[b]	#pep
26	P02647	APOLIPOPROTEIN A-I PRECURSOR	30759	288	7
27	P30048	MITOCHONDRIALTHIOREDOXIN-DEPENDENT PEROXIDE REDUCTASE PRECURSOR	28017	185	3
28	P04792	HEAT SHOCK 27 KD PROTEIN	22427	290	5
29	P00441	SUPEROXIDE DISMUTASE	16023	174	3
30	P04083	ANNEXIN I	38787	416	8
31	P04792	HEAT SHOCK 27 KD PROTEIN	22427	201	3
32	P02743	SERUM AMYLOID P-COMPONENT PRECURSOR	25485	99	2
33	O00299	CHLORIDE INTRACELLULAR CHANNEL PROTEIN 1	27249	125	1
34	P52565	RHO GDP-DISSOCIATION INHIBITOR 1	23250	109	2
35	P08865	40S RIBOSOMAL PROTEIN SA	32947	232	6

The spot number is correlated to the numbers found marked on the gel in Fig. 1. Accession numbers from Swissprot (http://us.expasy.org/sprot/) A score >50 was considered a significant hit by the search engine. Number of peptides is matched peptides to the corresponding protein.

[a] Mwt (molecular weight) (Da).

[b] Mascot (www.matrixscience.com).

from our group, we also found it easier to predict clinical outcome for the ER positive than for the ER negative cohort, based on gene expression data or conventional factors [32]. A possible explanation for this could be that the ER positive tumors are a more homogenous group than ER negative tumors. We have previously analyzed the same samples with cDNA microarray (Nimeus et al. in press in European Journal of Cancer, 2006). A list of genes distinguishing patients with distant recurrences from patients with no recurrences was created and the 4484 genes included were ranked according to their prognostic importance. When comparing the most important genes to the proteins with different expression, similarities to this study were found. As mentioned above, thioredoxin domain containing protein 5 (similar to glucose-regulated protein) was increased in the group of patients with distant recurrences and the corresponding gene was ranked the 59th most important gene and was also induced in the group with distant recurrences. Two proteins involved in the initiation of translation, eukaryotic translation initiation factor 4A-II and 1A, were found to be increased in tumors with distant recurrences. Genes with similar functions were also upregulated in the tumors with distant recurrences in the gene expression data set, exemplified by three different eukaryotic translation initiation factors (factor 5, 2 and 4A-I) ranked 125th, 288th and 367th, respectively and eukaryotic translation elongation factor 1 ranked 76th. Eukaryotic translation initiation factor 1A was downregulated in the tumors with distant recurrences and ranked 1596th most important gene. As has previously been shown [33], there is no absolute correlation between mRNA and protein expression, which may explain why not all proteins were detected on the gene list. In addition the 2-DE approach is hampered by the fact that only a limited number of proteins can readily be detected and identified. However, even though the cDNA microarray generates a more complete list of distinguishing genes, the 2-D gel approach allows quantification at the protein level as well as detection of posttranslational modifications, corroborating that these two techniques may be complementary.

Comparing the ER/PgR positive tumors and ER/PgR negative, one protein, apoptosis-associated speck-like protein, was found in lower amounts in the ER/PgR positive subgroup. This protein promotes caspase-mediated apoptosis and has previously been shown to be a target of methylation-induced gene silencing in human breast cancers [34]. In the subgroup with distant recurrences several proteins involved in translation/protein folding and proteins associated with cytoskeletal functions were differentially expressed, which indicates reorganization of highly abundant proteins in these tumors. Among the differentially expressed proteins in the subgroup with no recurrences, a UV-excision repair protein was found in lower amounts in the ER/PgR negative tumors than in the ER/PgR positive tumors. This protein is involved in DNA repair. Previously it has been shown that impaired DNA repair has been associated with poor clinical prognosis [35]. It is noteworthy that neither ER nor PgR were detected, which most likely depends on the low expression levels of these proteins, far below the sensitivity of the staining procedure.

In summary, by the use of 2-DE we were able to find candidate proteins involved in several different biological functions linked to tumorgenesis that were differentially expressed in primary tumors from patients later developing distant recurrences compared to those not developing recurrences. Even though the number of patients included in this study was relatively few we still may have found processes of importance for drug resistance/sensitivity in breast cancer. Independent verifications of these markers need to be accomplished in larger patient samples.

Acknowledgments

This study was supported by Swedish Cancer Society, Swedish Research Council, Gunnar, Arvid and Elisabeth Nilsson Foundation, and Mrs Berta Kamprad Foundation. We thank Görel Östberg for re-evaluation of histological grade and histological type. We are also indebted to participating departments of surgery, oncology, and pathology of the South Sweden Breast Cancer Group, and Regional Tumor Registry for primary care and follow-up of the patients, and for handling of frozen breast cancer tissue.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jpba.2006.09.019.

References

[1] Lancet. Early Breast Cancer Trialistsï Collaborative Group. Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. 365 (2005) 1687–1717.
[2] S. Romain, P.M. Martin, J.G. Klijn, W.L. van Putten, M.P. Look, O. Guirou, J.A. Foekens, Int. J. Cancer 74 (1997) 156–161.
[3] W.L. Washtien, Mol. Pharmacol. 25 (1984) 171–177.
[4] J.L. Clark, S.H. Berger, A. Mittelman, F.G. Berger, Cancer. Treat. Rep. 71 (1987) 261–265.
[5] H.B. Muss, A.D. Thor, D.A. Berry, T. Kute, E.T. Liu, F. Koerner, C.T. Cirrincione, D.R. Budman, W.C. Wood, M. Barcos, et al., N. Engl. J. Med. 330 (1994) 1260–1266.
[6] K. Nooter, G. Brutel de la Riviere, M.P. Look, K.E. van Wingerden, S.C. Henzen-Logmans, R.J. Scheper, M.J. Flens, J.G. Klijn, G. Stoter, J.A. Foekens, Br. J. Cancer 76 (1997) 486–493.
[7] W.S. el-Deiry, Curr. Opin. Oncol. 9 (1997) 79–87.
[8] G. MacGrogan, L. Mauriac, M. Durand, F. Bonichon, M. Trojani, I. de Mascarel, J.M. Coindre, Br. J. Cancer 74 (1996) 1458–1465.
[9] P.J. Maxwell, D.B. Longley, T. Latif, J. Boyer, W. Allen, M. Lynch, U. McDermott, D.P. Harkin, C.J. Allegra, P.G. Johnston, Cancer Res. 63 (2003) 4602–4606.
[10] S. Geisler, A.L. Borresen-Dale, H. Johnsen, T. Aas, J. Geisler, L.A. Akslen, G. Anker, P.E. Lonning, Clin. Cancer Res. 9 (2003) 5582–5588.
[11] I. Hedenfalk, D. Duggan, Y. Chen, M. Radmacher, M. Bittner, R. Simon, P. Meltzer, B. Gusterson, M. Esteller, O.P. Kallioniemi, B. Wilfond, A. Borg, J. Trent, N. Engl. J. Med. 344 (2001) 539–548.
[12] S. Gruvberger, M. Ringner, Y. Chen, S. Panavally, L.H. Saal, A. Borg, M. Ferno, C. Peterson, P.S. Meltzer, Cancer Res. 61 (2001) 5979–5984.
[13] L.J. van't Veer, H. Dai, M.J. van de Vijver, Y.D. He, A.A. Hart, M. Mao, H.L. Peterse, K. van der Kooy, M.J. Marton, A.T. Witteveen, G.J. Schreiber, R.M. Kerkhoven, C. Roberts, P.S. Linsley, R. Bernards, S.H. Friend, Nature 415 (2002) 530–536.

[14] C. Sotiriou, S.Y. Neo, L.M. McShane, E.L. Korn, P.M. Long, A. Jazaeri, P. Martiat, S.B. Fox, A.L. Harris, E.T. Liu, Proc. Natl. Acad. Sci. U.S.A. 100 (2003) 10393–10398.
[15] S. Paik, S. Shak, G. Tang, C. Kim, J. Baker, M. Cronin, F.L. Baehner, M.G. Walker, D. Watson, T. Park, W. Hiller, E.R. Fisher, D.L. Wickerham, J. Bryant, N. Wolmark, N. Engl. J. Med. (2004).
[16] Y. Wang, J.G. Klijn, Y. Zhang, A.M. Sieuwerts, M.P. Look, F. Yang, D. Talantov, M. Timmermans, M.E. Meijer-van Gelder, J. Yu, T. Jatkoe, E.M. Berns, D. Atkins, J.A. Foekens, Lancet 365 (2005) 671–679.
[17] R.I. Somiari, A. Sullivan, S. Russell, S. Somiari, H. Hu, R. Jordan, A. George, R. Katenhusen, A. Buchowiecka, C. Arciero, H. Brzeski, J. Hooke, C. Shriver, Proteomics 3 (2003) 1863–1873.
[18] S. Gharbi, P. Gaffney, A. Yang, M.J. Zvelebil, R. Cramer, M.D. Waterfield, J.F. Timms, Mol. Cell Proteomics 1 (2002) 91–98.
[19] K. Roberts, K. Bhatia, P. Stanton, R. Lord, Proteomics 4 (2004) 784–792.
[20] A.A. Alaiya, B. Franzen, A. Hagman, B. Dysvik, U.J. Roblick, S. Becker, B. Moberger, G. Auer, S. Linder, Int. J. Cancer 98 (2002) 895–899.
[21] A.A. Alaiya, M. Oppermann, J. Langridge, U. Roblick, L. Egevad, S. Brindstedt, M. Hellstrom, S. Linder, T. Bergman, H. Jornvall, G. Auer, Cell. Mol. Life Sci. 58 (2) (2001) 307–311.
[22] K. Hellman, A.A. Alaiya, K. Schedvins, W. Steinberg, A.C. Hellstrom, G. Auer, Br. J. Cancer 91 (2004) 319–326.
[23] J. Poland, D. Schadendorf, H. Lage, M. Schnolzer, J.E. Celis, P. Sinha, Clin. Chem. Lab. Med. 40 (2002) 221–234.
[24] H. Sigurdsson, B. Baldetorp, A. Borg, M. Dalberg, M. Ferno, D. Killander, H. Olsson, J. Ranstam, Br. J. Cancer 62 (1990) 786–790.
[25] C.W. Elston, I.O. Ellis, Histopathology 19 (1991) 403–410.
[26] WHO, in: F.A. Tavassoli, P. Devilee (Eds.), Pathology and genetics of tumours of the breast and female genital organs, IARC Press, Lyon, 2003.
[27] A. Shevchenko, M. Wilm, O. Vorm, M. Mann, Anal. Chem. 68 (1996) 850–858.
[28] B. Franzen, G. Auer, A.A. Alaiya, E. Eriksson, K. Uryu, T. Hirano, K. Okuzawa, H. Kato, S. Linder, Int. J. Cancer. 69 (1996) 408–414.
[29] K. Iwaya, H. Ogawa, Y. Mukai, A. Iwamatsu, K. Mukai, Cancer. Sci. 94 (2003) 864–870.
[30] A. Ciechanover, A. Orian, A.L. Schwartz, Bioessays 22 (2000) 442–451.
[31] J.C. Kwok, D.R. Richardson, Crit. Rev. Oncol. Hematol. 42 (2002) 65–78.
[32] P. Eden, C. Ritz, C. Rose, M. Ferno, C. Peterson, Eur. J. Cancer 40 (2004) 1837–1841.
[33] S.P. Gygi, Y. Rochon, B.R. Franza, R. Aebersold, Mol. Cell. Biol. 19 (1999) 1720–1730.
[34] K.E. Conway, B.B. McConnell, C.E. Bowring, C.D. Donald, S.T. Warren, P.M. Vertino, Cancer Res. 60 (2000) 6236–6242.
[35] M. Arai, J. Utsunomiya, Y. Miki, Int. J. Clin. Oncol. 9 (2004) 270–282.

Paper III

Research article

Open Access

Gene expression profiling in primary breast cancer distinguishes patients developing local recurrence after breast-conservation surgery, with or without postoperative radiotherapy

Emma Niméus-Malmström[1], Morten Krogh[2], Per Malmström[1], Carina Strand[1], Irma Fredriksson[3], Per Karlsson[4], Bo Nordenskjöld[5], Olle Stål[5], Görel Östberg[6], Carsten Peterson[2] and Mårten Fernö[1]

[1]Institute of Clinical Sciences, Department of Oncology, University Hospital, SE 221 85 Lund, Sweden
[2]Computational Biology and Biological Physics, Department of Theoretical Physics, Lund University, SE 223 68 Lund, Sweden
[3]Department of Surgery, Karolinska University Hospital in Solna, SE 171 76 Stockholm, Sweden
[4]Department of Oncology, Sahlgrenska University Hospital, SE 413 45 Gothenburg, Sweden
[5]Department of Clinical and Experimental Medicine, Division of Oncology, Linköping University, University Hospital, SE 581 85 Linköping, Sweden
[6]Department of Pathology, Halmstad Hospital, SE 302 33 Halmstad, Sweden

Corresponding author: Mårten Fernö, Marten.Ferno@med.lu.se

Received: 8 Nov 2007 Revisions requested: 29 Nov 2007 Revisions received: 26 Feb 2008 Accepted: 22 Apr 2008 Published: 22 Apr 2008

Breast Cancer Research 2008, **10**:R34 (doi:10.1186/bcr1997)
This article is online at: http://breast-cancer-research.com/content/10/2/R34

Abstract

Introduction Some patients with breast cancer develop local recurrence after breast-conservation surgery despite postoperative radiotherapy, whereas others remain free of local recurrence even in the absence of radiotherapy. As clinical parameters are insufficient for identifying these two groups of patients, we investigated whether gene expression profiling would add further information.

Methods We performed gene expression analysis (oligonucleotide arrays, 26,824 reporters) on 143 patients with lymph node-negative disease and tumor-free margins. A support vector machine was employed to build classifiers using leave-one-out cross-validation.

Results Within the estrogen receptor-positive (ER^+) subgroup, the gene expression profile clearly distinguished patients with local recurrence after radiotherapy (n = 20) from those without local recurrence (n = 80 with or without radiotherapy). The receiver operating characteristic (ROC) area was 0.91, and 5,237 of 26,824 reporters had a *P* value of less than 0.001 (false discovery rate = 0.005). This gene expression profile provides substantially added value to conventional clinical markers (for example, age, histological grade, and tumor size) in predicting local recurrence despite radiotherapy. Within the ER^- subgroup, a weaker, but still significant, signal was found (ROC area = 0.74). The ROC area for distinguishing patients who develop local recurrence from those who remain local recurrence-free in the absence of radiotherapy was 0.66 (combined ER^+/ER^-).

Conclusion A highly distinct gene expression profile for patients developing local recurrence after breast-conservation surgery despite radiotherapy has been identified. If verified in further studies, this profile might be a most important tool in the decision making for surgery and adjuvant therapy.

Introduction

The addition of postoperative radiotherapy to breast-conservation surgery in patients with lymph node-negative breast cancer has been shown to reduce the 10-year risk of local recurrence from 29.2% to 10% [1]. However, more than half of the patients will never develop local recurrence whether given radiotherapy or not and a small proportion of the patients will develop local recurrence despite being given radiotherapy. Besides tumor-involved margins, generally accepted risk factors for the development of local recurrence are young age

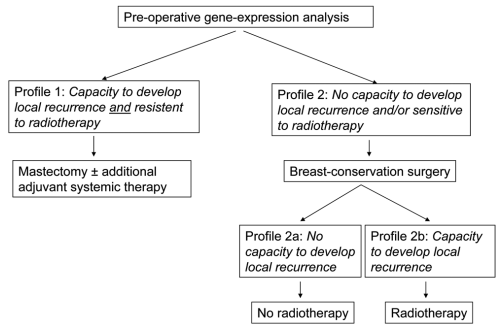
ER = estrogen receptor; ER^- = estrogen receptor-negative; ER^+ = estrogen receptor-positive; GO = Gene Ontology; LR^-RT^- = no local recurrence, no radiotherapy given; LR^-RT^+ = no local recurrence after radiotherapy; LR^+RT^- = local recurrence developed, no radiotherapy given; LR^+RT^+ = local recurrence developed after radiotherapy; RIN = RNA integrity number; ROC = receiver operating characteristic; SVM = support vector machine.

and multicentricity [2-5]. A number of other risk factors have been reported (for example, extensive intraductal component [6], family history [7], lymphovascular invasion [2,8-10], lobular cancer [11], and estrogen receptor-negative (ER^-) status [10]), but their clinical usefulness so far is limited. If the patients who develop local recurrence despite radiotherapy can be identified, other treatment strategies should be considered. No factor hitherto has been found to be clinically useful for the identification of patients developing local recurrence after radiotherapy.

Gene expression analyses have been found to be useful for molecular subclassification of breast cancer and also have shown promising results for predicting distant recurrence [12-17]. Concerning prediction of local recurrence, only a few studies have been reported. Cheng and colleagues [18] demonstrated two sets of gene expression profiles to be associated with local recurrence after mastectomy. Today, however, the majority of patients with breast cancer are operated on with breast-conservation surgery. As conventional risk factors for local recurrence after mastectomy differ from those after breast-conservation surgery, these findings may not be applicable when using less extensive surgery. Two recent studies included only patients treated with breast-conservation surgery: one was unable to find a distinguishing gene expression profile [19], whereas the other could only separate patients developing local recurrence after radiotherapy from patients not developing local recurrence by means of a predefined gene list, the wound-response signature [20]. This signature has been suggested to provide a possible link between cancer progression and wound healing and originally was defined as the fibroblast core serum response [21]. The material in the study by Nuyten and colleagues [20] was heterogeneous with regard to margin status, ER status, lymph node status, adjuvant systemic treatment (47% with and 53% without), and radiotherapy (including both standard and boost treatment). This heterogeneity might be the reason for not finding a significant gene profile in this study when using the whole set of genes. As far as the importance of considering ER status in gene expression analyses, today it is generally accepted that ER^+ and ER^- breast tumors have remarkably distinct gene expression profiles [22,23] and this subdivision of ER status has been successfully applied when predicting distant recurrence [14,24].

Our study aimed at elucidating whether gene expression analysis is useful in predicting tumor sensitivity to radiotherapy and capacity to develop local recurrence in a patient material homogenous with regard to tumor-free margins, lymph node status, and radiotherapy (only standard doses). A predictive gene expression profile might impact the choice of both surgery and radiotherapy. A hypothetical clinical routine scheme, demonstrating three treatment options, is outlined in Figure 1. After a preoperative analysis of the gene expression profile, the first step is to identify the patients who will develop local recurrence despite radiotherapy. For this group, mastectomy might be a better choice. The second step is to separate those patients with no capacity to develop local recurrence and therefore not in need of radiotherapy after breast-conservation surgery from those with a capacity to develop local recurrence and in need of radiotherapy.

Figure 1



A hypothetical clinical routine scheme for the choice of surgery and radiotherapy after preoperative gene expression analysis.

Materials and methods

Patients

Study design, inclusion criteria, and sample collection

Frozen tumor samples were collected from patients representing the following four groups: (a) LR^+RT^+ (local recurrence developed after radiotherapy), (b) LR^-RT^+ (no local recurrence after radiotherapy), (c) LR^+RT^- (local recurrence developed, no radiotherapy given), and (d) LR^-RT^- (no local recurrence, no radiotherapy given). All patients were operated on with breast-conservation surgery and axillary clearance with no lymph node involvement, tumor size of less than or equal to 30 mm (two patients had tumors measuring 32 and 40 mm, respectively, and one was T2 without any further information on size), tumor-free margins (>1 mm), no multicentricity, and with frozen tumor tissue with good RNA quality available. Local recurrence was defined as the appearance of a new breast tumor in the ipsilateral residual breast parenchyma in the overlying skin or in the scar. Patients with recurrence in the contralateral breast or with distant metastases prior or simultaneous to local recurrence were excluded. In the first inclusion, 102 patients from a randomized clinical trial in the South and West health care regions in Sweden [25] and 19 patients from a population-based cohort study with a nested case-control study (Stockholm and South Sweden) [3,26] were included (Tables 1 and 2). To perform gene expression profiling in a more homogenous material with regard to ER status and yet with a sufficient number of cases in all four subgroups, 22 additional ER^+ tumors from the South-East and South health care regions were included in a second inclusion (Tables 1 and 2). The study was approved by the Ethics Committee at Lund

Table 1

Clinical and pathological characteristics of the 77 patients receiving radiotherapy, with or without the development of local recurrence

All	LR^+RT^+		LR^-RT^+
Time to local recurrence, months	n = 30		n = 47
Median			
Range	50		-
Follow-up, months	8–149		-
Median			
Range	-		85
Tamoxifen	-		62–231
Chemotherapy	6		1
Tamoxifen and chemotherapy	0		0
No adjuvant treatment	0		0
	24		46
Inclusion 1 and 2	Inclusion 1	Inclusion 2	Inclusion 1
Menopause			
Pre	12	5	13
Post	7	1	33
Not available	0	5	1
Age at operation, years			
Median	48	47	57
Range	27–63	34–73	33–73
Size, millimeters			
Median	14	15	15
Range	2–32	10–20	4–22
Not available	0	2[a]	0
Grade			
1	3	3	22
2	8	6	13
3	7	1	10
Not available	1	1	2
Estrogen receptor			
Positive	9	11	42
Negative	10	0	5
Progesterone receptor			
Positive	5	9	31
Negative	13	2	11
Not available	1	0	5
Health care region			
South	8	2	30

Table 1 *(Continued)*

Clinical and pathological characteristics of the 77 patients receiving radiotherapy, with or without the development of local recurrence

West	2	0	9
South-East	0	9	0
Stockholm	9	0	8

[a]One T1 and one T2. LR^-RT^+ = no local recurrence after radiotherapy; LR^+RT^+ = local recurrence developed after radiotherapy.

University (Lund, Sweden). Patient and primary tumor characteristics and follow-up information were collected from the patients' medical records.

Treatment

Postoperative radiotherapy with a median absorbed dose of 50 Gy (range 48 to 54 Gy) was given in 24 to 27 fractions in one series to the remaining breast parenchyma. Adjuvant systemic therapy was given to 16 patients (Tables 1 and 2).

Conventional factors

Histological grade was re-evaluated according to Elston and Ellis [27]. ER and progesterone receptor content were analyzed routinely after primary operation with enzyme immunoassay according to kit instructions (Abbott Laboratories, Diagnostics Division, Chicago, IL, USA) and expressed as femtomoles per milligram of cytosol protein. Receptor values greater than or equal to 25 fmol/mg protein were considered positive.

Gene expression analysis

RNA was extracted from freshly frozen invasive breast tumors as previously described [28]. The RNA quality was assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA), and the RNA integrity number (RIN) method [29] was used to validate the RNA quality. Twenty-one samples were excluded due to RIN values of below 6. Labeling and hybridization were performed as previously described [28]. By means of Human Genome Oligo Set Version 2.1 (containing 21,329 70-mer probes) and Human Genome Oligo Set Version 2.1 Upgrade 27 (containing 5,462 70-mer probes), oligonucleotide arrays were produced by the Swegene DNA Microarray Resource Centre, Lund University [30]. In inclusion step 1, probes were printed in duplicate, creating 55 K slides, and in inclusion step 2, in single, creating 27 K slides.

Statistics

Wilcoxon rank sum tests, Sammon maps, and support vector machine (SVM) classifications [31] were performed with the statistical language R [32] using the libraries MASS (Sammon) and e1071 (SVM). For the SVM, only genes with no missing values were used. For the LR^+RT^+ versus LR^-RT^+/LR^-RT^- groups, the numbers of genes with no missing values were 9,128, 13,362, and 8,834 for the ER^+, ER^-, and combined ER groups, respectively. For the LR^+RT^- versus LR^-RT^- groups, they were 11,209, 13,547, and 10,658, respectively. For the wound-response genes, the corresponding numbers were 93, 120, and 91 for the LR^+RT^+ versus LR^-RT^+/LR^-RT^- groups and 105, 122, and 99 for the LR^+RT^- versus LR^-RT^- groups. Leave-one-out cross-validation was used. When a sample was left out, the SVM was trained on the remaining samples, and the distance to the maximal margin hyperplane (the decision value) was calculated for the left-out sample. A linear kernel was used and the cost of constraints violation (C constant) was fixed to one. No parameter tuning was performed even if the use of another layer of cross-validation might have improved the results. The goal of this study was to prove that gene expression profiles can distinguish the groups, not to find the optimal classifier. Actually, the optimal classifier does not even need to be an SVM. We also minimized potential suspicions about information leak by restricting the parameters of the SVM to the default values of the R function svm. A receiver operating characteristic (ROC) curve and area were calculated using the decision values. The expected average value of the ROC area is 0.5 if there is no discrimination between the groups. Due to random variations, ROC areas above 0.5 are often obtained even when there is no discrimination between the groups. To distinguish a real discrimination between the groups from the case of no discrimination, a p-value was calculated. A small p-value makes it unlikely that the ROC area can be reconciled with the case of no discrimination. The p-value was calculated by a permutation test. The local recurrence labels were shuffled randomly 1,000 times and the ROC areas were found for the corresponding classifications. The *P* value was calculated as the fraction of the 1,000 permutations that had an ROC area larger than the real one. If the *P* value was zero, the random ROC areas were fitted to a normal distribution and the area under the tail above the real ROC area was used as the *P* value. The *P* value of the ROC area for the case of a fixed test set (that is, no cross-validation) was calculated by a permutation test of the labels in the test set. Odds ratios, confidence intervals of odds ratios, and *P* values of odds ratios were calculated with the R function Fisher test, which uses the conditional maximum likelihood estimator.

Gene Ontology

The Gene Ontology (GO) [33] OBO (open biomedical ontologies) file of 14 November 2006 was used. Gene annotation was performed using ACID (Array Clone Information Database), which is a publicly available web application that provides GO categories for genes [34]. A total of 6,841 GO

Table 2

Clinical and pathological characteristics of the 66 patients, not receiving radiotherapy, with or without the development of local recurrence

All	LR^+RT^-	LR^-RT^-	
	n = 22	n = 44	
Time to local recurrence, months			
Median	35	-	
Range	11–96	-	
Follow-up, months			
Median	-	84	
Range	-	21–166	
Tamoxifen	2	4	
Chemotherapy	1	1	
Tamoxifen and chemotherapy	1	0	
No adjuvant treatment	18	39	
Inclusion 1 and 2	Inclusion 1	Inclusion 1	Inclusion 2
Menopause			
Pre	9	3	4
Post	13	30	2
Not available	0	0	5
Age at operation, years			
Median	53	61	49
Range	44–73	45–70	40–62
Size, millimeters			
Median	15	13	16
Range	7–30	6–40	10–26
Not available	0	0	0
Grade			
1	4	13	5
2	10	9	3
3	5	8	3
Not available	3	3	0
Estrogen receptor			
Positive	14	27	11
Negative	8	6	0
Progesterone receptor			
Positive	15	17	11
Negative	7	14	0
Not available	0	2	0
Health care region			
South	9	21	11

Table 2 *(Continued)*

Clinical and pathological characteristics of the 66 patients, not receiving radiotherapy, with or without the development of local recurrence

West	13	10	0
South-East	0	0	0
Stockholm	0	2	0

LR^-RT^- = no local recurrence, no radiotherapy given; LR^+RT^- = local recurrence developed, no radiotherapy given.

categories belonging to 'Cellular component', 'Biological process', or 'Molecular function' had at least one gene in common with our data. The genes were ranked according to their Wilcoxon rank sum *P* value between LR^+RT^+ and LR^-RT^+/LR^-RT^- groups in the ER^+ group. A Wilcoxon rank sum test was performed for each GO category to test for over-representation of genes toward the top of the ranked gene list using Catmap [35].

Results

Patients with a capacity to develop local recurrence despite radiotherapy

To identify this group of patients, we compared LR^+RT^+ versus LR^-RT^+/LR^-RT^- groups. There was an association between the LR^+RT^+ group and ER^- status (odds ratio 6.8, 95% confidence interval 2.0 to 24; $P = 0.0007$) (Table 3). In this analysis, only the patients from the first inclusion were used, as the second inclusion was made on ER and local recurrence status. Age can also distinguish the LR^+RT^+ group. Histological grade was marginally able to separate the LR^+RT^+ group from the LR^-RT^+/LR^-RT^- group, whereas tumor size was not (Table 3).

After filtering the microarray data on low-quality spots and missing values, 26,824 reporters remained, representing 16,895 unique genes. From a Sammon map of the gene expression profiles of the 100 ER^+ patients from both inclusions, it was evident that the LR^+RT^+ and LR^-RT^+/LR^-RT^- groups were well separated even without gene selection (Figure 2). For supervised classification, an SVM was used. The areas under the receiver operating curve (ROC areas) were 0.91 ($P = 9 \times 10^{-6}$) within the ER^+ group, 0.74 ($P = 0.08$) within the ER^- group (Figure 3), and 0.83 ($P = 9 \times 10^{-5}$) within the combined ER^+/ER^- group (data not shown). The ER^+ group was by far the larger group, which could explain the superior performance of the ER^+ group compared with the ER^- one. Also, the classification performance was deteriorated by combining ER^+ and ER^- in one SVM; it was preferable to use distinct SVMs for the two subpopulations (Figure 3). For the ER^+ group, at 90% sensitivity (18 of 20 LR^+RT^+ correctly classified), the specificity was 87.5% (70 of 80 LR^-RT^+/LR^-RT^- correctly classified), and at 90% specificity (72 of 80 LR^-RT^+/LR^-RT^- correctly classified), the sensitivity was 80% (16 of 20 LR^+RT^+ correctly classified).

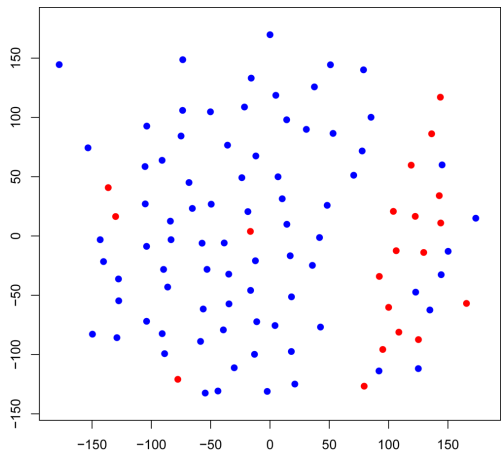
Table 3

A comparison between the LR^+RT^+ and LR^-RT^+/LR^-RT^- subgroups

Factor	LR^+RT^+	LR^-RT^+/LR^-RT^-	*P* value
ER status, number			
Negative	10	11	
Positive	9	69	0.0007[a]
Median age, years			
All	48	61	0.00004[b]
ER^- subgroup	49.5	53	0.12
ER^+ subgroup	46	61	0.002
Histological grade, number			
1	3	35	
2	8	22	
3	7	18	0.055[a]
Median tumor size, millimeters	15	15	0.95[b]

Only cases from inclusion 1 are included. [a]Fisher exact test. [b]Wilcoxon rank sum test. ER = estrogen receptor; LR^-RT^- = no local recurrence, no radiotherapy given; LR^-RT^+ = no local recurrence after radiotherapy; LR^+RT^+ = local recurrence developed after radiotherapy.

Figure 2



A Sammon map of the 100 estrogen receptor-positive patients within the LR^+RT^+ group (red dots, 20 patients) and the LR^-RT^+/LR^-RT^- group (blue dots, 80 patients). The Sammon map was performed with all 26,824 reporters. Euclidean distance in $\log_2$ expression values was used as the distance measure. LR^-RT^- = no local recurrence, no radiotherapy given; LR^-RT^+ = no local recurrence after radiotherapy; LR^+RT^+ = local recurrence developed after radiotherapy.

As age is a risk factor for local recurrence, we investigated whether the gene expression profiling has classification ability beyond that of age in the ER^+ subgroup. We constructed a training set from the 77 patients who were either older than 50 years and in the LR^-RT^+/LR^-RT^- group or younger than 50 years and in the LR^+RT^+ group. The test set consisted of the remaining 23 patients (for example, those who were either younger than 50 years and in the LR^-RT^+/LR^-RT^- group or older than 50 years and in the LR^+RT^+ group). The point is that the test set chosen contains patients who behave exactly opposite to the usual connection between age and local recurrence. Applying an SVM, we obtained an ROC area of 0.88 (P = 0.002). Furthermore, we checked the influence of health care regions by using the 68 samples from the South and South-East health care regions as a training set and the 32 samples from the West or Stockholm health care regions as a test set. The specific split into health care regions was done to get a reasonable amount of samples in LR^+RT^+ and LR^-RT^+/LR^-RT^- groups in both the training and the test set. No optimizations were performed in this regard. The ROC area of 0.87 (P = 0.002) shows that the classifier indeed works across health care regions.

The wound-response signature genes, also known as the core serum response genes [21], were shown to have the ability of partially predicting local recurrence [20]. We mapped the wound-response signature to our microarrays and performed an SVM classification using only this signature. The ROC areas were 0.75 (P = 0.007) within the ER^+ group, 0.75 (P = 0.08) within the ER^- group, and 0.61 (P = 0.10) within the combined ER^+/ER^- group.

Differentially expressed genes

A Wilcoxon rank sum test between LR^+RT^+ and LR^-RT^+/LR^-RT^- groups within the ER^+ subgroup was performed for all 26,824 reporters. A clear over-representation of genes with small P values was found; for example, there are 5,237 reporters with a P value below 0.001 corresponding to a Benjamini-Hochberg [36] false discovery rate of 0.005. A heatmap of the 81 genes with a known gene name, no missing values, and a Wilcoxon rank sum test P value between the LR^+RT^+ and LR^-RT^+/LR^-RT^- groups of below 10^{-6} is shown in Figure 4. A GO analysis was performed using Catmap [35]. A total of 6,841 GO categories belonging to 'Cellular component', 'Biological process', or 'Molecular function' were included. At a false discovery rate of 0.05, only the four categories of cytosolic ribosome (sensu Eukaryota), eukaryotic 43S preinitiation complex, eukaryotic 48S initiation complex, and cytosolic small ribosomal subunit (sensu Eukaryota) were significant.

Patients with no capacity to develop local recurrence

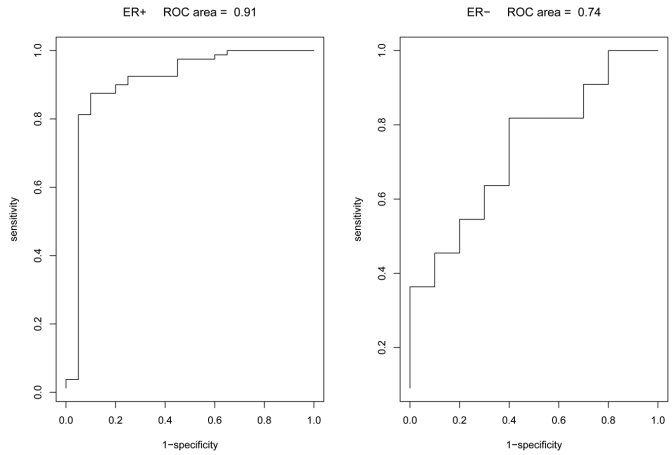
To identify this group of patients, we analyzed LR^+RT^- versus LR^-RT^- groups. ER^- status was weakly correlated with the LR^+RT^- group (odds ratio 2.5, 95% confidence interval 0.6 to 11; P = 0.21) (Table 4). Young age was correlated with local recurrence in the ER^+ group (Table 4). Neither histological grade nor tumor size had the power to separate the two groups.

An SVM gene expression classifier yielded an ROC area of 0.66 (P = 0.04) within the combined ER^+/ER^- group. The ER^- and ER^+ subgroups were too small to give a significant result on their own, even though there was a tendency of discriminative power within the ER+ subgroup. (ROC area = 0.62; P = 0.14). For the wound-response signature, the ROC areas were 0.64 (P = 0.10) in the ER^+ group, 0.69 (P = 0.27) in the ER^- group, and 0.68 (P = 0.03) in the combined group.

Discussion

We have found a highly significant gene expression profile associated with the development of local recurrence after breast-conservation surgery despite postoperative radiotherapy. If patients resistant to radiotherapy can be identified, they should be candidates for alternative treatment strategies such as mastectomy, other adjuvant treatments, and/or higher radiation doses as local recurrence implies an increased risk of both distant metastases and mortality [37-39]. So far, there are no markers useful in the clinic for the identification of radioresistant breast cancer. We found both age and ER status to be associated with local recurrence after radiotherapy. However, our gene expression signature provides substantially added value to these factors and also to histological grade and tumor size. A hybrid classifier of age and gene expression

Figure 3



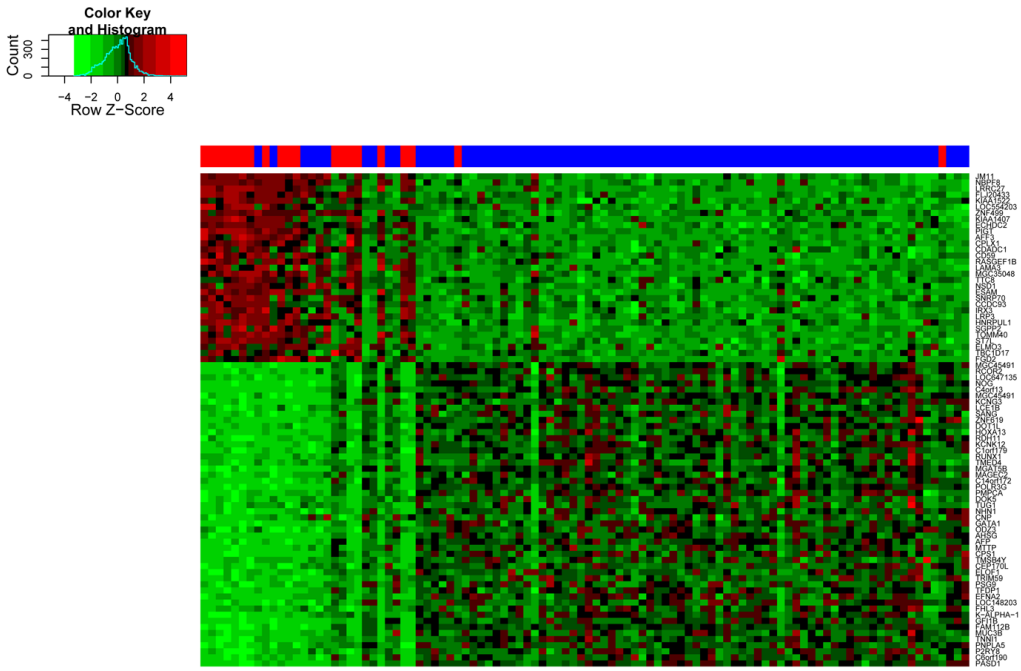
Receiver operating characteristic (ROC) curves for the support vector machine classification of LR^+RT^+ versus LR^-RT^+/LR^-RT^- groups within the estrogen receptor-positive (ER^+) group (left part) and estrogen receptor-negative (ER^-) group (right part). The specificity is defined as the fraction of the LR^-RT^+/LR^-RT^- patients correctly classified, and the sensitivity as the fraction of the LR^+RT^+ patients correctly classified. LR^-RT^- = no local recurrence, no radiotherapy given; LR^-RT^+ = no local recurrence after radiotherapy; LR^+RT^+ = local recurrence developed after radiotherapy.

should perform even better than age or gene expression alone. Due to the sample size in this study, we did not have the possibility to build such a hybrid classifier. We have focused on the question of whether gene expression analysis *per se* is useful for the identification of patients with different risks of developing local recurrences. A thorough and more specific evaluation of the gene list should be performed after a confirmative study in which not only the genes, but also the pathways in which they are involved, are considered. The high proportion of ribosomal-related genes is noteworthy but also needs to be confirmed. The samples were collected from four health care regions with different routines for handling fresh tumor tissue prior to freezing. However, we could clearly demonstrate that these differences did not influence the importance of the gene expression signature.

To our knowledge, only one previous study has reported a gene expression signature significantly associated with local recurrence after breast-conservation surgery [20], but only when using a predefined wound-response signature gene list. One reason for not finding a significant profile when using the entire gene set may be that their material, which included 17 local recurrences, was more heterogeneous than ours with regard to tumor-free margins, tumor size, lymph node status, and dose of radiotherapy. Furthermore, they did not separate the samples with regard to ER status. ER^+ and ER^- breast tumors are known to have distinct gene expression profiles and indeed we found a stronger gene expression profile when including only ER^+ tumors compared with when ER^- tumors were included (ROC area 0.91 compared with 0.83). This finding further strengthened the notion that ER^+ and ER^- breast cancer should be handled as two separate entities when evaluating gene expression data, as has previously been stated by authors in analyses of gene expression profiles associated with distant metastases [14,24,40]. In our material, the wound-response signature genes were able to predict local recurrence within both the ER^+ group and ER^- group with reasonable accuracy, whereas the prediction in the combined group was rather weak. For the ER^+ group and the combined group, the classification performance is inferior to the results obtained with all genes. This degradation of performance shows that the advantage of restricting the gene set used in the classifier to the focused set of wound-response signature genes, which are known to be relevant to cancer, is outweighed by the loss of information of discarding the majority of the genes. The reason that the SVM using all genes was so much better at classifying the combined group than the wound-response signature genes is probably that the ER signal is contained in the full gene set but is more or less absent in the wound-response signature genes. With respect to individual samples, it is seen that the samples that were misclassified with all genes were also misclassified with the wound-response signature genes but that many of the misclassified samples with the wound-response signature genes were correctly classified with all genes.

For the identification of patients with no capacity to develop local recurrence, we compared the LR^+RT^- and LR^-RT^- subgroups. The gene expression signal was weaker, but still significant (ROC area = 0.66). The reason for a weaker signal

Figure 4



The 81 most important genes. A heatmap of 81 genes (rows) and 100 patients (columns) with patients ordered according to their leave-one-out support vector machine classification value: estrogen receptor-positive patients within LR^+RT^+ (red, n = 20) and LR^-RT^+/LR^-RT^- (blue, n = 80) groups. The gene selection and ordering are described in the text. Expression values are logarithmically transformed, centered around zero, and normalized to unit standard deviation. LR^-RT^- = no local recurrence, no radiotherapy given; LR^-RT^+ = no local recurrence after radiotherapy; LR^+RT^+ = local recurrence developed after radiotherapy.

may be the small number of patients (n = 66). For the wound-response genes, there is a tendency for correct predictions, but the results are too weak to draw any conclusions. More samples would be needed to test it further.

Today, the vast majority of breast cancer patients are treated with radiotherapy after breast-conservation surgery to lower the risk of local recurrence. However, a large proportion of these patients have a very low risk for local recurrence, and the positive effects must be weighed against costs and side effects of the treatment. The identification of patients with a very low risk to develop local recurrence, and consequently not in need of postoperative radiotherapy, is of great clinical importance.

No patient included in our study had tumor-involved margins, a risk factor for local recurrence. The local recurrence rate is influenced by the amount of uninvolved breast parenchyma surrounding the tumor as higher recurrence rates have been reported after lumpectomy [41] than after sector resection [25] or quadrantectomy [42]. With smaller resection margins, microscopic residual tumor is more likely to be present in the breast and the administered radiotherapy dose may be too low to give complete protection for local recurrence. Thus, it cannot be excluded that the gene expression profile can be influenced by resection margins.

Our gene expression profile was associated with local recurrence after radiotherapy at commonly used doses. Higher doses of radiation, with a boost of 16 Gy, have been shown to significantly reduce the local recurrence rate, particularly in patients below 50 years of age [43]. Unfortunately, higher radiation doses cause a less satisfactory cosmetic result [44]. Whether gene profiling also can be used for identification of patients in need of a boost is unclear at present.

One might speculate why the gene expression profile method performs significantly better for the prediction of local recurrence than for the prediction of distant metastasis. It is possible that the gene profile associated with the development of

Table 4

A comparison between the LR^+RT^- and LR^-RT^- subgroups

Factor	LR^+RT^-	LR^-RT^-	*P* value
ER status, number			
Negative	8	6	
Positive	14	27	0.21[a]
Median age, years			
All	53	61	0.02[b]
ER^- subgroup	59.5	54	0.90
ER^+ subgroup	50.5	62	0.02
Histological grade, number			
1	4	13	
2	10	9	
3	5	8	0.19[a]
Median tumor size, millimeters	15	14	0.67[b]

Only cases from inclusion 1 are included. [a]Fisher exact test. [b]Wilcoxon rank sum test. ER, = estrogen receptor; LR^-RT^- = no local recurrence, no radiotherapy given; LR^+RT^- = local recurrence developed, no radiotherapy given.

local recurrence despite radiotherapy (indicating radio-resistance) is more homogeneous than the one associated with distant recurrences. It is believed that the development of metastases is a more complicated process and that different groups of genes may be of variable importance in distinguished subgroups of breast cancer.

Conclusion

We have found a very promising gene expression profile for predicting local recurrence despite radiotherapy – a profile that might be associated with radio-resistance. The signature provides substantially added value to the conventional factors used to predict risk of local recurrence. If confirmed in further studies, this profile might be a most important tool in the decision making for type of surgery and adjuvant therapy.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

EN-M participated in conceiving the design of the study, collecting the patient material and information of basic patient and tumor characteristics and clinical follow-up, performing gene expression and statistical analyses, interpreting data, and writing the paper. MK and CP participated in conceiving the design of the study, performing gene expression and statistical analyses, interpreting data, and writing the paper. EN-M and MK contributed equally to this manuscript. PM participated in conceiving the design of the study, collecting the patient material and information of basic patient and tumor characteristics and clinical follow-up, interpreting data, and writing the paper. CS participated in performing gene expression and statistical analyses, interpreting data, and writing the paper. IF, PK, BN, and OS participated in collecting the patient material and information of basic patient and tumor characteristics and clinical follow-up. GÖ re-evaluated the histopathological parameters. MF participated in conceiving the design of the study, interpreting data, and writing the paper. All authors read and approved the final manuscript.

Acknowledgements

We are indebted to participating departments of the South, West, South-East Sweden, and Stockholm Breast Cancer Group for providing samples and clinical follow-up. We especially thank Dorthe Grabau for re-evaluation of histological grade and Dick Killander for fruitful discussions at early stages of the study. The study was supported by funds from the Swedish Cancer Society, the Swedish Research Council, the Swedish Foundation for Strategic Research, the Gunnar, Arvid, and Elisabeth Nilsson Foundation, the Mrs Berta Kamprad Foundation, the University Hospital of Lund Research Foundation, the Knut Alice Wallenberg Foundation through the Swegene consortium, the Strategic Science Foundation CREATE Health Centre, Skane County Council's Research and Development Foundation, and Governmental Funding of Clinical Research within the Nation Health Service.

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:**1687-1717.
2. Voogd AC, Nielsen M, Peterse JL, Blichert-Toft M, Bartelink H, Overgaard M, van Tienhoven G, Andersen KW, Sylvester RJ, van Dongen JA: **Differences in risk factors for local and distant recurrence after breast-conserving therapy or mastectomy for stage I and II breast cancer: pooled results of two large European randomized trials.** *J Clin Oncol* 2001, **19:**1688-1697.
3. Fredriksson I, Liljegren G, Palm-Sjövall M, Arnesson LG, Emdin SO, Fornander T, Lindgren A, Nordgren H, Idvall I, Holmqvist M, Holmberg L, Frisell J: **Risk factors for local recurrence after breast-conserving surgery.** *Br J Surg* 2003, **90:**1093-1102.
4. Leopold KA, Recht A, Schnitt SJ, Connolly JL, Rose MA, Silver B, Harris JR: **Results of conservative surgery and radiation therapy for multiple synchronous cancers of one breast.** *Int J Radiat Oncol Biol Phys* 1989, **16:**11-16.
5. Kurtz JM, Jacquemier J, Amalric R, Brandone H, Ayme Y, Hans D, Bressac C, Spitalier JM: **Breast-conserving therapy for macroscopically multiple cancers.** *Ann Surg* 1990, **212:**38-44.
6. Jacquemier J, Kurtz JM, Amalric R, Brandone H, Ayme Y, Spitalier JM: **An assessment of extensive intraductal component as a risk factor for local recurrence after breast-conserving therapy.** *Br J Cancer* 1990, **61:**873-876.
7. Jobsen JJ, Palen J van der, Meerwaldt JH: **The impact of age on local control in women with pT1 breast cancer treated with conservative surgery and radiation therapy.** *Eur J Cancer* 2001, **37:**1820-1827.
8. Magee B, Swindell R, Harris M, Banerjee SS: **Prognostic factors for breast recurrence after conservative breast surgery and radiotherapy: results from a randomised trial.** *Radiother Oncol* 1996, **39:**223-227.
9. Borger J, Kemperman H, Hart A, Peterse H, van Dongen J, Bartelink H: **Risk factors in breast-conservation therapy.** *J Clin Oncol* 1994, **12:**653-660.
10. Cheng SH, Horng CF, Clarke JL, Tsou MH, Tsai SY, Chen CM, Jian JJ, Liu MC, West M, Huang AT, Prosnitz LR: **Prognostic index score and clinical prediction model of local regional recurrence after mastectomy in breast cancer patients.** *Int J Radiat Oncol Biol Phys* 2006, **64:**1401-1409.
11. Liljegren G, Lindgren A, Bergh J, Nordgren H, Tabar L, Holmberg L: **Risk factors for local recurrence after conservative treat-**

ment in stage I breast cancer. Definition of a subgroup not requiring radiotherapy. *Ann Oncol* 1997, **8:**235-241.
12. van't Veer LJ, Dai H, Vijver MJ van de, He YD, Hart AA, Mao M, Peterse HL, Kooy K van der, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH: **Gene expression profiling predicts clinical outcome of breast cancer.** *Nature* 2002, **415:**530-536.
13. Vijver MJ van de, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, Parrish M, Atsma D, Witteveen A, Glas A, Delahaye L, Velde T van der, Bartelink H, Rodenhuis S, Rutgers ET, Friend SH, Bernards R: **A gene-expression signature as a predictor of survival in breast cancer.** *N Engl J Med* 2002, **347:**1999-2009.
14. Wang Y, Klijn JG, Zhang Y, Sieuwerts AM, Look MP, Yang F, Talantov D, Timmermans M, Meijer-van Gelder ME, Yu J, Jatkoe T, Berns EM, Atkins D, Foekens JA: **Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer.** *Lancet* 2005, **365:**671-679.
15. Perou CM, Sørlie T, Eisen MB, Rijn M van de, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lønning PE, Børresen-Dale AL, Brown PO, Botstein D: **Molecular portraits of human breast tumours.** *Nature* 2000, **406:**747-752.
16. Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, Rijn M van de, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Eystein Lønning P, Børresen-Dale AL: **Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications.** *Proc Natl Acad Sci USA* 2001, **98:**10869-10874.
17. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, Demeter J, Perou CM, Lønning PE, Brown PO, Børresen-Dale AL, Botstein D: **Repeated observation of breast tumor subtypes in independent gene expression data sets.** *Proc Natl Acad Sci USA* 2003, **100:**8418-8423.
18. Cheng SH, Horng CF, West M, Huang E, Pittman J, Tsou MH, Dressman H, Chen CM, Tsai SY, Jian JJ, Liu MC, Nevins JR, Huang AT: **Genomic prediction of locoregional recurrence after mastectomy in breast cancer.** *J Clin Oncol* 2006, **24:**4594-4602.
19. Kreike B, Halfwerk H, Kristel P, Glas A, Peterse H, Bartelink H, Vijver MJ van de: **Gene expression profiles of primary breast carcinomas from patients at high risk for local recurrence after breast-conserving therapy.** *Clin Cancer Res* 2006, **12:**5705-5712.
20. Nuyten DS, Kreike B, Hart AA, Chi JT, Sneddon JB, Wessels LF, Peterse HJ, Bartelink H, Brown PO, Chang HY, Vijver MJ van de: **Predicting a local recurrence after breast-conserving therapy by gene expression profiling.** *Breast Cancer Res* 2006, **8:**R62.
21. Chang HY, Sneddon JB, Alizadeh AA, Sood R, West RB, Montgomery K, Chi JT, Rijn M van de, Botstein D, Brown PO: **Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds.** *PLoS biology* 2004, **2:**E7.
22. Gruvberger S, Ringner M, Chen Y, Panavally S, Saal LH, Borg A, Ferno M, Peterson C, Meltzer PS: **Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns.** *Cancer Res* 2001, **61:**5979-5984.
23. Sotiriou C, Neo SY, McShane LM, Korn EL, Long PM, Jazaeri A, Martiat P, Fox SB, Harris AL, Liu ET: **Breast cancer classification and prognosis based on gene expression profiles from a population-based study.** *Proc Natl Acad Sci USA* 2003, **100:**10393-10398.
24. Teschendorff AE, Naderi A, Barbosa-Morais NL, Pinder SE, Ellis IO, Aparicio S, Brenton JD, Caldas C: **A consensus prognostic gene expression classifier for ER positive breast cancer.** *Genome Biol* 2006, **7:**R101.
25. Malmström P, Holmberg L, Anderson H, Mattsson J, Jönsson PE, Tennvall-Nittby L, Balldin G, Lovén L, Svensson JH, Ingvar C, Möller T, Holmberg E, Wallgren A, Swedisj Breast Cancer Group: **Breast conservation surgery, with and without radiotherapy, in women with lymph node-negative breast cancer: a randomised clinical trial in a population with access to public mammography screening.** *Eur J Cancer* 2003, **39:**1690-1697.
26. Fredriksson I, Liljegren G, Arnesson LG, Emdin SO, Palm-Sjovall M, Fornander T, Frisell J, Holmberg L: **Time trends in the results of breast conservation in 4694 women.** *Eur J Cancer* 2001, **37:**1537-1544.
27. Elston C, Ellis I: **Assessment of histological grade.** In *Systemic Pathology Volume 13*. 3rd edition. London: Churchill Livingstone; 1998.
28. Strand C, Enell J, Hedenfalk I, Ferno M: **RNA quality in frozen breast cancer samples and the influence on gene expression analysis – a comparison of three evaluation methods using microcapillary electrophoresis traces.** *BMC Mol Biol* 2007, **8:**38.
29. Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M, Lightfoot S, Menzel W, Granzow M, Ragg T: **The RIN: an RNA integrity number for assigning integrity values to RNA measurements.** *BMC Mol Biol* 2006, **7:**3.
30. Jönsson G, Naylor TL, Vallon-Christersson J, Staaf J, Huang J, Ward MR, Greshock JD, Luts L, Olsson H, Rahman N, Stratton M, Ringnér M, Borg A, Weber BL: **Distinct genomic profiles in hereditary breast tumors identified by array-based comparative genomic hybridization.** *Cancer Res* 2005, **65:**7612-7621.
31. Cristianini N, Shawe-Taylor J: *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods* Cambridge, UK: Cambridge University Press; 2000.
32. Ihaka R, Gentleman R: **A language for data analysis and graphics.** *J Comp Graph Stat* 1996, **5:**299-314.
33. **The Gene Ontology homepage** [http://www.geneontology.org]
34. Ringner M, Veerla S, Andersson S, Staaf J, Hakkinen J: **ACID: a database for microarray clone information.** *Bioinformatics* 2004, **20:**2305-2306.
35. Breslin T, Eden P, Krogh M: **Comparing functional annotation analyses with Catmap.** *BMC Bioinformatics* 2004, **5:**193.
36. Benjamini Y, Hochberg Y: **Controlling the false discovery rate: a practical and powerful approach to multiple testing.** *J R Stat Soc Ser B* 1995, **57:**289-300.
37. Touboul E, Buffat L, Belkacémi Y, Lefranc JP, Uzan S, Lhuillier P, Faivre C, Huart J, Lotz JP, Antoine M, Pène F, Blondon J, Izrael V, Laugier A, Schlienger M, Housset M: **Local recurrences and distant metastases after breast-conserving surgery and radiation therapy for early breast cancer.** *Int J Radiat Oncol Biol Phys* 1999, **43:**25-38.
38. Fortin A, Larochelle M, Laverdiere J, Lavertu S, Tremblay D: **Local failure is responsible for the decrease in survival for patients with breast cancer treated with conservative surgery and postoperative radiotherapy.** *J Clin Oncol* 1999, **17:**101-109.
39. Clarke M, Collins R, Darby S, Davies C, Elphinstone P, Evans E, Godwin J, Gray R, Hicks C, James S, MacKinnon E, McGale P, McHugh T, Peto R, Taylor C, Wang Y, Early Breast Cancer Trialists' Collaborative Group (EBCTCG): **Effects of radiotherapy and of differences in the extent of surgery for early breast cancer on local recurrence and 15-year survival: an overview of the randomised trials.** *Lancet* 2005, **366:**2087-2106.
40. Eden P, Ritz C, Rose C, Ferno M, Peterson C: **'Good Old' clinical markers have similar power in breast cancer prognosis as microarray gene expression profilers.** *Eur J Cancer* 2004, **40:**1837-1841.
41. Fisher B, Anderson S, Redmond CK, Wolmark N, Wickerham DL, Cronin WM: **Reanalysis and results after 12 years of follow-up in a randomized clinical trial comparing total mastectomy with lumpectomy with or without irradiation in the treatment of breast cancer.** *N Engl J Med* 1995, **333:**1456-1461.
42. Veronesi U, Salvadori B, Luini A, Greco M, Saccozzi R, del Vecchio M, Mariani L, Zurrida S, Rilke F: **Breast conservation is a safe method in patients with small cancer of the breast. Long-term results of three randomised trials on 1,973 patients.** *Eur J Cancer* 1995, **31A:**1574-1579.
43. Bartelink H, Horiot JC, Poortmans P, Struikmans H, Bogaert W Van den, Barillot I, Fourquet A, Borger J, Jager J, Hoogenraad W, Collette L, Pierart M, European Organization for Research and Treatment of Cancer Radiotherapy and Breast Cancer Groups: **Recurrence rates after treatment of breast cancer with standard radiotherapy with or without additional radiation.** *N Engl J Med* 2001, **345:**1378-1387.
44. Bartelink H, Horiot JC, Poortmans PM, Struikmans H, Bogaert W Van den, Fourquet A, Jager JJ, Hoogenraad WJ, Oei SB, Wárlám-Rodenhuis CC, Pierart M, Collette L: **Impact of a higher radiation dose on local control and survival in breast-conserving therapy of early breast cancer: 10-year results of the randomized boost versus no boost EORTC 22881-10882 trial.** *J Clin Oncol* 2007, **25:**3259-3265.

Paper IV