Original Paper

Pathobiology

Pathobiology 2011;78:140–148
DOI: 10.1159/000323809

Received: July 20, 2010
Accepted after revision: December 22, 2010
Published online: May 26, 2011

PRO_10 – A New Tissue-Based Prognostic Multigene Marker in Patients with Early Estrogen Receptor-Positive Breast Cancer

A.E. Moor[a, c] C. Guevara[b] H.J. Altermatt[d] R. Warth[e] R. Jaggi[a] S. Aebi[c]

[a]Department of Clinical Research, and [b]Institute of Pathology, University of Bern, [c]Department of Medical Oncology, Bern University Hospital, [d]Pathology Länggasse, and [e]Foundation Biobank-Suisse, Bern, Switzerland

Key Words
PRO_10 · Estrogen receptor · Breast cancer

Abstract
Background/Aims: Clinicopathological and molecular factors determine the prognosis of breast cancer. PRO_10 is a prognostic score based on quantitative RT-PCR of 10 proliferation-associated genes obtained from formalin-fixed, paraffin-embedded breast cancer tissues. We revalidated PRO_10 in patients treated in a non-trial setting. ***Methods:*** The charts of 315 patients with postmenopausal estrogen receptor (ER)-positive breast cancer between 1996 and 2004 were reviewed. Forty-eight cases relapsed within 5 years of diagnosis; they were paired with controls by matching the N and T stage, histological grade, percent ER-positive cells, human epidermal growth factor receptor 2, age, adjuvant chemo- and endocrine therapy. The score was tested by conditional logistic regression. ***Results:*** Despite strict matching, PRO_10 remained prognostic for recurrence in the whole group (odds ratio, OR = 4.7, p = 0.005) and in subgroups of grade 2 (OR = 5.5, p = 0.009) and N0 cancers (OR = 15, p = 0.002). Five-year recurrence-free survival was 29% in patients with high and 67% in patients with low scores (p = 0.002). PRO_10 was prognostic for overall survival (5-year overall survival 71 vs. 91%). ***Conclusion:*** PRO_10 is an independent prognostic marker in postmenopausal ER-positive breast cancer. It is based on formalin-fixed, paraffin-embedded tissue and could be integrated easily into the routine diagnostic workflow.

Introduction

Estrogen receptor (ER)-positive breast cancer is a heterogenous disease in terms of clinical course, response to chemotherapy and endocrine treatments and underlying deregulated pathways. The anatomic stage (e.g., TNM stage) and histopathological appearance (e.g., histological grade) supplemented by immunohistochemical markers determine the recommendation of different treatment options [1]. The histological grade is usually assigned by using the Elston-Ellis grading system [2]. While this system reliably identifies patients with relatively favorable (grade 1) and relatively poor (grade 3) prognosis, about half the patients fall into the intermediate prognostic category of grade 2 cancers [3–5]. Recent advances in molecular diagnostics allow a more individualized approach to prognosis and therapeutic recommendations. cDNA microarray analyses of differential gene expression have led to the identification of 'intrinsic subtypes' of breast

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

1015–2008/11/0783–0140$38.00/0

Accessible online at:
www.karger.com/pat

Prof. S. Aebi
Division of Medical Oncology
Luzerner Kantonsspital
CH–6000 Luzern 16 (Switzerland)
Tel. +41 41 205 5860, E-Mail stefan.aebi@onkologie.ch

cancer [6] and to the development of prognostic multigene expression profiles [7–9]. Such profiles could become instrumental in identifying subsets of patients who will or will not benefit from different cancer treatment modalities and drugs. For instance, in addition to being prognostic for distant [7] and locoregional failure [10], the 21-gene-based recurrence score appears to identify patients who might not benefit from adjuvant chemotherapies [11, 12]. Such methods might also improve the cost-effectiveness of future therapies [13] by identifying the patients who are most likely to profit. Early expression profiling studies relied on microarrays and required fresh frozen cancer tissue [8]. Newer reliable scores based on formalin-fixed, paraffin-embedded tissues (FFPE) can be integrated easily into the routine pathological workflow [7, 14]. They are equally reliable as scores based on fresh-frozen cancer tissue [15]. The commercially available scoring systems Oncotype DX™ and Mammaprint™ have been validated retrospectively for different subsets of breast cancer patients [11, 12, 16–18] and are currently being evaluated in prospective randomized studies [19, 20].

The PRO_10 score [14] is based on proliferation-associated genes and proved to be an independent predictor of relapse in participants of the Breast International Group (BIG) 1-98 trial [21] in addition to conventional prognostic factors. Hence, the objective of the present study was to determine if these results for the PRO_10 score can be replicated in a retrospective study using matched pairs of patients with postmenopausal ER-positive breast cancer.

Patients and Methods

Study Population

We conducted a case-control study within patients with early breast cancer treated at the Department of Medical Oncology, Inselspital Bern, Bern, Switzerland. Patients gave general consent to the use of their tissues for research. This study was approved according to Swiss law by the research ethics committee of the Canton Berne. To increase the median follow-up period, only patients diagnosed with breast cancer between 1996 and October 2004 were identified in the tumor registry. Eligible patients had histologically confirmed invasive breast cancer. They were postmenopausal at the time of diagnosis (aged ≥55 years with amenorrhea for more than 1 year at the time of diagnosis, or follicle-stimulating hormone levels indicating postmenopausal status at the time of diagnosis) and the tumors were endocrine responsive as evidenced by positive ER and/or progesterone receptor (PR) as measured by immunohistochemistry (at least 10% stained cells) or ligand binding assays (receptor content ≥10 fmol/mg cytosolic protein). All patients received adjuvant endocrine therapy. The pathology institutes involved participated in quality assurance programs (National External Quality Assurance Scheme). Exclusion criteria were as follows: male gender, received neoadjuvant therapy, multiple carcinomas and tissue blocks (not available in the collaborating institutes of pathology, i.e. Institute of Pathology, University of Bern, Pathology Länggasse, and Pathology Unilabs, Bern, Switzerland). The medical records were reviewed to obtain detailed information concerning diagnosis, therapy and outcome. Pathologic tumor information was gathered from the pathology reports and reviewed by a pathologist (C.G., H.J.A.). Eligibility for inclusion of the patients was verified through this review.

Case-Control Matching

We identified a total of 315 eligible patients. Cases were defined as patients with a relapse of disease (local recurrence or distant metastasis) within 60 months of diagnosis of primary tumor. The matching procedure was done according to a predefined matching score incorporating ER and PR, histologic grade, tumor size (American Joint Committee on Cancer), nodal status, age, adjuvant endocrine therapy and chemotherapy. For some variables, we defined subsets to facilitate the matching process (percentage of ER-positive cells 0, <10, <50 and ≤100%; percentage of PR-positive cells: 0, <10, <50 and ≤100%; tumor size: ≤20 and >20 mm; regional lymph node metastases: 0, ≤3 and >3). Age was considered matching if the difference between the patients' ages was <5 years. Endocrine therapy was classified as aromatase inhibitor or selective ER modulators. The chemotherapy variable describes any chemotherapy versus none. For the matching score, 2 points were awarded for matching histologic grade and ER subset, respectively, and 1 point was assigned for a match in the variables representing the PR subset, human epidermal growth factor receptor 2 (HER2) status, endocrine therapy, chemotherapy, age and tumor size. Cases and controls were considered as matched if the lymph node status was identical and if the sum of the matching score was ≥7 of a maximum of 10.

Sample Preparation

Hematoxylin and eosin (H&E) stained slides from the tissue blocks of the 100 matched patients were reviewed by board-certified pathologists (C.G., H.J.A.) to select the block with the highest tumor content. Depending on the available amount of tissue, 2 H&E slides and five 10-μm sections were cut from FFPE cancer tissue. The H&E slides were used to evaluate the percentage of tumor cells. Malignant cells constituted the majority of the epithelial tissue component in all samples. Total RNA was extracted from five 10-μm sections of FFPE cancer tissue. RNA was isolated and demodified as previously published [15]. Quality control of the extracted RNA was performed on an Agilent 2100 Bioanalyzer (Agilent Technologies, Inc., Santa Clara, Calif., USA). The RNA fragment length varied between 150 and 500 bases with one exception where the RNA was degraded more extensively (see below). This size distribution is similar to previous studies [14, 15, 22].

Genes

The PRO_10 RT-PCR assay consists of 10 genes that correlate with proliferation and 3 control genes. PRO_10 was built based on in silico gene selection and has been validated in an independent subset of participants of the randomized controlled clinical trial BIG 1-98 [14, 21]. Additionally, 2 further control genes and 11 new candidate genes were examined (Appendix 1). The candidate

genes were chosen through a review of recent literature focusing on genes that were reported to be operative in pathways leading to resistance to endocrine therapy. These candidate genes were ranked according to 4 criteria: (1) the published work was based on RNA expression data; (2) multiple reports exist of the same gene in this context; (3) the RNA expression level was predictive of tamoxifen response in postmenopausal breast cancer patients, and (4) well-characterized gene belonging to a molecular pathway of known relevance to carcinogenesis or drug resistance. The following 11 genes fulfilled these criteria and were included in this study based on the corresponding literature: EGFR [23, 24]; FGFR4 [25]; CCNE1 [26, 27]; TSC22D1 [28]; PSAP [28]; CCND1 [29, 30]; NCOA3 [31, 32]; CDKN1B [33, 34]; NCOA1 [35]; NCOR2 [24, 35]; PAX2 [36].

Quantitative RT-PCR

Each RNA was tested by a quantitative RT-PCR (qRT-PCR) with 5 control genes (GUSB, RPLP0, TFRC, GAPDH, UBB). For each sample, the mean of the 5 raw cycle threshold (Ct) values was determined. One sample with a mean raw Ct value >31 was considered of poor RNA quality and was therefore excluded from further analysis. The corresponding Bioanalyzer profile revealed strongly fragmented RNA (data not shown). The remaining 98 samples were used to measure the 30 genes specified in Appendix 1 by qRT-PCR on TaqMan Low Density Arrays (Applied Biosystems, Foster City, Calif., USA) using a one-step protocol (Invitrogen, Basel, Switzerland) on an Applied Biosystems 7900 HT instrument. The raw Ct values were inverted and normalized relative to 3 control genes (RPLP0, UBB, GUSB) [14] according to formulas 1 and 2 in Appendix 2. Three further samples had to be excluded from the study, as their PRO_10 scores were not calculable due to high Ct values that reached the prespecified cutoff values (Appendix 2). The remaining 17 probe slots on the TaqMan Low Density Arrays were used to measure expression levels of genes related to ER (10 genes), PR (5 genes) and HER2 (2 genes). Each gene was examined for its ability to differentiate between cases and controls by conditional logistic regression analysis (fig. 1). Four genes were combined with PRO_10 resulting in a new module score with potentially enhanced prognostic and/or predictive impact (for a detailed formula, see Appendix 2).

Statistical Analyses

To ensure comparability with previously published data, PRO_10 score values were determined using the published algorithm. Conditional logistic regression analysis was used to determine odds ratios (ORs) for the association of molecular scores and relapse of disease. ORs calculated by logistic regression treating the score values as a continuous variable are determined per one-unit increase on the score scales. The narrow range of the score values explains the markedly larger OR for a continuous score variable versus a dichotomized score variable; changes in the scale of the scores are reflected in the OR but do not influence the p values. Statistical significance was calculated by the likelihood ratio test and the Wald test. By using the maximum likelihood method, some subset conditional logistic analyses produced non-converging coefficients due to the small dataset. Hence, penalized conditional likelihood and Firths' bias correction method [37] were applied to determine ORs and 95% confidence intervals. Survival was estimated by the product limit method [38]. All statistical calculations were done in R (version 2.10.0).

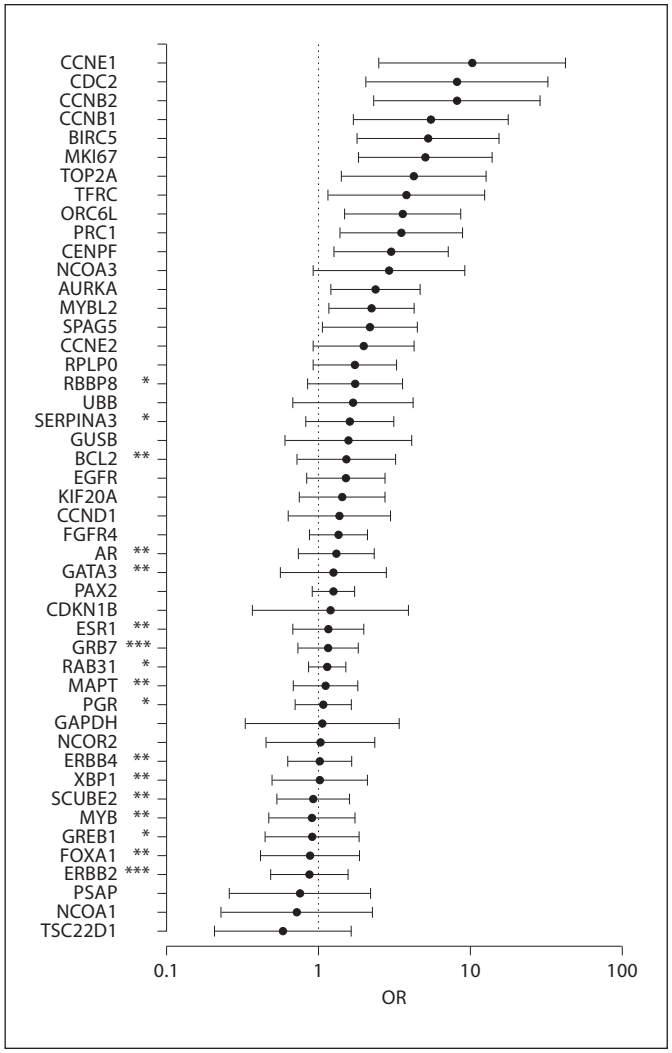

Fig. 1. Single-gene OR of relapse as determined by conditional logistic regression analysis (95% confidence intervals). Genes marked with asterisks were not used for score calculation in this study and belong to module scores from a previous study [15]. * PR module score; ** ER module score; *** HER2 module score.

Results

Patient Characteristics

Among the patients with early breast cancer treated between 1996 and 2004 at the Department of Medical Oncology, Inselspital Bern, Bern, Switzerland, we identified 315 eligible patients. Among those patients, 50 cases (local or distant recurrence of disease within 60 months) could be matched to 50 controls. The FFPE blocks of 1 patient could not be retrieved by the proper institute of

Table 1. Baseline characteristics

Characteristic	Cases	Controls	All
Age at diagnosis, years			
Median	64.27	60.65	62.68
Range	48–82	46–78	46–82
Menopausal category			
Postmenopausal before chemotherapy	48 (100)	47 (100)	95 (100)
Premenopausal (ineligible)	0	0	0
Tumor size			
≤2 cm	18 (37)	17 (36)	35 (37)
>2 cm	30 (63)	30 (64)	60 (63)
Tumor grade			
Grade 1	1 (2)	3 (6)	4 (4)
Grade 2	29 (60)	31 (66)	60 (63)
Grade 3	18 (38)	13 (28)	33 (33)
Nodal status			
Negative	20 (42)	19 (40)	39 (41)
Positive (1–3)	15 (31)	15 (32)	30 (32)
Positive (>3)	13 (27)	13 (28)	26 (27)
ER and PR status			
ER and PR positive	40 (83)	39 (83)	79 (83)
ER positive, PR negative	6 (13)	7 (15)	13 (14)
ER negative, PR positive	2 (4)	1 (2)	3 (3)
ER and PR negative (ineligible)	0	0	0
Local therapy			
BCS and RT	29 (60)	35 (74)	64 (67)
BCS and no RT	4 (8)	0	4 (4)
Mastectomy and RT	3 (6)	3 (6)	6 (6)
Mastectomy and no RT	12 (25)	9 (19)	21 (22)
Adjuvant chemotherapy			
Yes	21 (45)	22 (47)	43 (45)
No	27 (55)	25 (53)	52 (55)
HER2 status			
Negative	25 (52)	23 (49)	48 (51)
Positive	11 (23)	11 (23)	22 (23)
Unknown	12 (25)	13 (28)	25 (26)

Figures in parentheses are percentages. BCS = Breast-conserving surgery; RT = radiation therapy.

pathology. After evaluation of the qRT-PCR results, 4 further patients had to be excluded due to insufficient RNA quality. Patient and tumor characteristics of the remaining 95 samples are summarized in table 1. Median follow-up estimated by inverse Kaplan-Meier analysis is 70.9 months [39].

PRO_10 Score

The risk of relapse within 60 months was strongly and positively associated with the PRO_10 score (table 2), whether analyzed as a continuous variable ($p < 0.001$), dichotomized into high and low scores at the median ($p < 0.001$), or categorized according to the prespecified cutoff value ($p = 0.005$) established in a prior validation study [14]. Lymph node-negative, node-positive and grade 2 tumors were analyzed as separate subsets and the PRO_10 score remained significant for node-negative and grade 2 subsets.

Five-year recurrence-free survival (RFS) was 29% in patients with high (above median) PRO_10 and in 67% in patients with low PRO_10 ($p = 0.002$) (fig. 2); the median RFS was 4.0 years with high scores, but was not reached with low scores. PRO_10 was also prognostic for overall survival (5-year overall survival 71 vs. 91%, median overall survival 8.1 years vs. not reached; $p = 0.0057$).

Table 2. ORs associated with the PRO_10 score

Score	Cases	Controls	OR	95% CI	p value (LR)
All patients					
Continuous	48 (100)	47 (100)	11.14	2.48–50.12	<0.001
Prespecified cutoff					
Low proliferation (<14.5)	26 (54)	39 (83)	1.0	reference	
High proliferation (≥14.5)	22 (46)	8 (17)	4.67	1.34–16.24	0.005
Median					
Below median	14 (29)	32 (68)	1.0	reference	
Above median	34 (71)	15 (32)	9.0	2.09–38.79	<0.001
Node-negative tumors					
Continuous	18 (100)	18 (100)	27.7	1.39–550.92	<0.001
Prespecified cutoff					
Low proliferation (<14.5)	8 (44)	15 (83)	1.0	reference	
High proliferation (≥14.5)	10 (56)	3 (17)	15.0	1.83–19.47[1]	0.002
Median					
Below median	4 (22)	14 (78)	1.0	reference	
Above median	14 (78)	4 (22)	21.0	2.71–27.01[1]	<0.001
Node-positive tumors					
Continuous	27 (100)	27 (100)	6.18	0.94–40.55	0.058[2]
Prespecified cutoff					
Low proliferation (<14.5)	18 (67)	22 (81)	1.0	reference	
High proliferation (≥14.5)	9 (33)	5 (19)	2.33	0.6–9.02	0.2
Median					
Below median	11 (41)	16 (59)	1.0	reference	
Above median	16 (59)	11 (41)	3.5	0.73–16.85	0.086
Grade 2 tumors					
Continuous	26 (100)	26 (100)	12.32	1.96–77.47	<0.001
Prespecified cutoff					
Low proliferation (<14.5)	4 (15)	22 (85)	1.0	reference	
High proliferation (≥14.5)	13 (50)	13 (50)	5.5	1.22–24.81	0.009
Median					
Below median	8 (31)	18 (69)	1.0	reference	
Above median	18 (69)	8 (31)	6.0	1.34–26.81	0.005

Figures in parentheses are percentages. CI = Confidence interval; LR = likelihood ratio. The conditional logistic regression model calculates ORs by taking into account the paired structure of cases and controls. Therefore, some patients had to be omitted for the subset analysis when the paired case and control did not belong to the same subset. Patients omitted: node-negative patients (n = 3), node-positive patients (n = 2), grade 2 tumors (n = 8).

[1] ORs and confidence intervals were obtained by conditional logistic regression using Firth's bias reduction [39].

[2] p value calculated by the Wald test; p = 0.032, calculated by the likelihood ratio test.

Patients with a high PRO_10 score (above median) who were treated with adjuvant chemotherapy had a median RFS of 4.3 versus 3.2 years without adjuvant chemotherapy (p = 0.185). The corresponding RFS curves for patients with low PRO_10 scores were superimposable (p = 0.896).

Discussion

Molecular scores based on gene expression data can be predictors of the clinical course of disease in patients with ER-positive breast cancer independent of established clinical markers [7, 8, 14, 40]. PRO_10 was designed in silico based on a meta-analysis of gene expres-

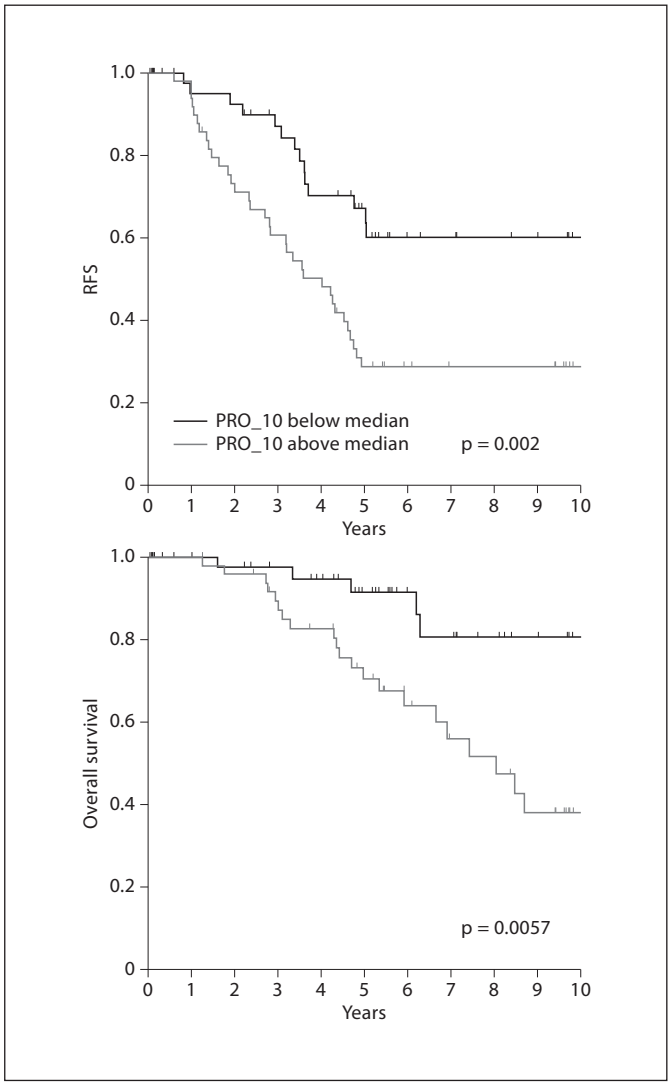

Fig. 2. Kaplan-Meier plots for overall survival and RFS. Patients were dichotomized into low and high PRO_10 scores at the median value. The p value corresponds to the log-rank test.

sion profiles [41] and validated on a subset of 342 participants of the BIG 1-98 trial [14]. The aim of this case-control study was to further validate the previously published proliferation score PRO_10. To avoid that the results of molecular profiling are confounded by clinical prognostic markers, a strict matching procedure was applied including the most important clinical and pathological determinants of breast cancer prognosis. This procedure precludes any statistical analysis of the matching variables (given their artificially even distribution among cases and controls) but it allows evaluating the scores under investigation independently of a potential impact of these confounding variables. The study population is not representative of a general population of breast cancer patients; the paired structure of cases and controls intentionally increases the rate of recurrence in the study population. Nevertheless, we observed a highly significant association ($p < 0.001$) of the PRO_10 score with the risk of local or distant recurrence even after the strict matching procedure.

From a clinical perspective, the greatest benefit of molecular profiling is anticipated when it aids in the decision making process for adjuvant therapy. The prognosis is most uncertain in patients with grade 2 node-negative breast cancer. Thus, we separately analyzed patient subsets with intermediate risk of recurrence and an uncertain likelihood to benefit from adjuvant treatment (grade 2 tumors, node-negative patients). The PRO_10 score was still significantly associated with the risk of recurrence in both patient subsets. When applied to the patients in this study and treated as a binary variable (cutoff median), the PRO_10 score resulted in a sensitivity of 71% and a specificity of 68% for correct prognosis of recurrence (receiver operating characteristic: area under the curve = 0.734). This result documents the prognostic power of PRO_10, and at the same time, it is evident that there is still a considerable potential for improvement of the score.

It is likely that PRO_10 might be improved by adding additional expression measures of gene correlating with prognosis. As an example, we have constructed an 'explorative score' comprising 14 genes drawn from Appendix 1 as outlined in Appendix 2. Although this score appears to be more accurate than PRO_10 (Appendix 3), it must be considered explorative and needs to be validated in an independent set of breast cancer samples.

In summary, we have confirmed the prognostic value of PRO_10 on a set of 'real life' patients treated in the context of a university hospital after tight matching of clinical parameters in cases and controls. The results show that the quantification of the expression of proliferation-related genes in tumor tissue allows to classify patients into groups with favorable and poor prognosis. The investigation and validation of prognostic and predictive scores like PRO_10 may contribute to the development of improved personalized diagnosis and treatment of patients with breast cancer.

Appendix 1

Gene Identifications, Categories and Score Affiliations

Gene	Category	Accession No.	Description	AS, bp	PRO_10
GAPDH	control	NM_002046.3	glyceraldehyde-3-phosphate dehydrogenase	74	
GUSB	control	NM_000181.2	β-glucuronidase	81	
RPLP0	control	NM_053275.3	ribosomal protein, large, P0	105	
TFRC	control	NM_003234.2	transferrin receptor (p90, CD71)	79	
UBB	control	NM_018955.2	ubiquitin B	120	
MKI67	proliferation	NM_002417.3	antigen identified by monoclonal antibody Ki-67	131	x
AURKA	proliferation	NM_003600.2	aurora kinase A	85	
BIRC5	proliferation	NM_001012270.1	baculoviral inhibitor of apoptosis repeat-containing 5	93	
CCNB1	proliferation	NM_031966.2	cyclin B1	104	
MYBL2	proliferation	NM_002466.2	v-myb myeloblastosis viral oncogene homolog (avian)-like 2	81	
CCNB2	proliferation	NM_004701.2	cyclin B2	73	x
CCNE2	proliferation	NM_057749.1	cyclin E2	70	x
CDC2	proliferation	NM_033379.2	cell division cycle 2, G_1 to S and G_2 to M	92	x
CENPF	proliferation	NM_016343.3	centromere protein F, 350/400ka (mitosin)	99	x
KIF20A	proliferation	NM_005733.2	kinesin family member 20A	130	x
ORC6L	proliferation	NM_014321.2	origin recognition complex, subunit 6 like (yeast)	78	x
PRC1	proliferation	NM_199413.1	protein regulator of cytokinesis 1	66	x
SPAG5	proliferation	NM_006461.3	sperm-associated antigen 5	114	x
TOP2A	proliferation	NM_001067.2	topoisomerase (DNA) II α 170 kDa	125	x
EGFR	new	NM_201282.1	epidermal growth factor receptor [erythroblastic leukemia viral (v-erb-b) oncogene homolog, avian]	86	
FGFR4	new	NM_022963.2	fibroblast growth factor receptor 4	73	
CCNE1	new	NM_001238.1	cyclin E1	64	
TSC22D1	new	NM_006022.2	TSC22 domain family, member 1	67	
PSAP	new	NM_001042465.1	prosaposin	74	
CCND1	new	NM_053056.2	cyclin D1	57	
NCOA3	new	NM_181659.1	nuclear receptor coactivator 3	59	
CDKN1B	new	NM_004064.3	cyclin-dependent kinase inhibitor 1B (p27, Kip1)	71	
NCOA1	new	NM_147223.2	nuclear receptor coactivator 1	59	
NCOR2	new	NM_001077261.1	nuclear receptor corepressor 2	75	
PAX2	new	NM_000278.3	paired box 2	57	

AS = Amplicon size.

Appendix 2

$$gene\_cutoff = if\,(gene\_Ct > 37),\ then\ (0),\ else\ (37 - gene\_Ct) \quad (1)$$

$$gene\_normalized = \frac{33 \times \left(gene\_cutoff - mean\ \left(control\ genes\_cutoff\right) + 37\right)}{74} \quad (2)$$

$$PRO\_14\_unscaled = 5.33 \times birc5 + 5.53 \times ccnb1 + 8.24 \times ccnb2 + 8.25 \times cdc2 + 3.04 \times cenpf + 5.09 \times mki67 + 2.25 \times mybl2 + 3.62 \times orc6l + 3.56 \times prc1 + 2.19 \times spag5 + 4.28 \times top2a + 10.38 \times ccne1 + 2.4 \times aurka + 2 \times ccne2 \quad (3)$$

$$PRO\_14 = \frac{PRO\_14\_unscaled}{66.16} \mid 66.16 = coef_1 + coef_2 + \ldots + coef_n \quad (4)$$

Equation 1 is used to invert raw Ct values; a cutoff point is established at 37 cycles.

Equation 2 normalizes the inverted expression values relative to UBB, GUSB and RPLP0. The coefficients ensure a positive theoretical scale from 0 to 33.

Equation 3 describes the algorithm for PRO_14; normalized expression values are utilized.

Equation 4 scales PRO_14 to achieve comparability with PRO_10.

Appendix 3

The normalized expression levels of each investigated gene (Appendix 1) were examined for their ability to differentiate between cases and controls by conditional logistic regression analysis (fig. 1). The ORs of 14 genes were used as coefficients and multiplied by the corresponding expression levels (equation 3 in Appendix 2) to create the expolrative score PRO_14.

ORs Associated with the Explorative Score PRO_14

Score	Cases	Controls	OR	95% CI	p value (LR)
All patients					
Continuous	48 (100)	47 (100)	16.73	3.16–88.46	<0.001
Median					
Below median (<14.02)	13 (27)	35 (74)	1.0	reference	
Above median (≥14.02)	35 (73)	12 (26)	11.0	2.59–46.78	<0.001
Node-negative patients					
Continuous	18 (100)	18 (100)	22.59	1.29–396	<0.001
Median					
Below median (<14.16)	4 (22)	14 (78)	1.0	reference	
Above median (≥14.16)	14 (78)	4 (22)	21.0	2.71–27.01[1]	<0.001
Node-positive patients					
Continuous	27 (100)	27 (100)	13.76	1.65–114.58	0.003
Median					
Below median (<13.93)	10 (37)	16 (59)	1.0	reference	
Above median (≥13.93)	17 (63)	11 (41)	4.0	0.85–18.84	0.08[2]
Grade 2 tumors					
Continuous	26 (100)	26 (100)	15.85	2.15–116.77	<0.001
Median					
Below median (<13.9)	8 (31)	18 (69)	1.0	reference	
Above median (≥13.9)	18 (69)	8 (31)	6.0	1.34–26.81	0.00

Figures in parentheses are percentages. CI = Confidence interval; LR = likelihood ratio. The conditional logistic regression model calculates ORs by taking into account the paired structure of cases and controls. Therefore, some patients had to be omitted for the subset analysis when the paired case and control did not belong to the same subset; patients omitted: node-negative patients (n = 3), node-positive patients (n = 2), grade 2 tumors (n = 8).

[1] ORs and confidence intervals were obtained by conditional logistic regression using Firth's bias reduction [39]. [2] p value calculated by the Wald test.

Acknowledgements

We thank Andrea Oberli and Isabelle Minder for providing excellent technical assistance. We thank Dr. Urs Wagner, MD, Pathology Unilabs Berne, for providing archival tissue. We thank Dr. Georg Heinze, Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, for providing assistance in the statistical application of Firth's bias reduction. This work was supported by Cancer Research Switzerland/Oncosuisse (grant OCS-02132-08-2007 to R.J., S.A. and H.J.A.), the Foundation Biobank-Suisse (to R.J.), and the Scientific Fund of the Department of Medical Oncology, Berne University Hospital, Inselspital, Bern (to S.A.). Senior authorship is shared equally between R.J. and S.A.

References

1 Goldhirsch A, Ingle JN, Gelber RD, et al: Thresholds for therapies: highlights of the St Gallen International Expert Consensus on the primary therapy of early breast cancer 2009. Ann Oncol 2009;20:1319–1329.

2 Elston CW, Ellis IO: Pathological prognostic factors in breast cancer. 1. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. Histopathology 1991;19:403–410.

3 Page DL, Gray R, Allred DC, et al: Prediction of node-negative breast cancer outcome by histologic grading and S-phase analysis by flow cytometry: an Eastern Cooperative Oncology Group Study (2192). Am J Clin Oncol 2001;24:10–18.

4 Pinder SE, Murray S, Ellis IO, et al: The importance of the histologic grade of invasive breast carcinoma and response to chemotherapy. Cancer 1998;83:1529–1539.
5 Simpson JF, Gray R, Dressler LG, et al: Prognostic value of histologic grade and proliferative activity in axillary node-positive breast cancer: results from the Eastern Cooperative Oncology Group Companion Study, EST 4189. J Clin Oncol 2000;18:2059–2069.
6 Sorlie T, Perou CM, Tibshirani R, et al: Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci USA 2001;98:10869–10874.
7 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;351:2817–2826.
8 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.
9 van't Veer LJ, Dai H, van de Vijver MJ, et al: Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002;415:530–536.
10 Mamounas EP, Tang G, Fisher B, et al: Association between the 21-gene recurrence score assay and risk of locoregional recurrence in node-negative, estrogen receptor-positive breast cancer: results from NSABP B-14 and NSABP B-20. J Clin Oncol 2010;28:1677–1683.
11 Albain KS, Barlow WE, Shak S, et al: Prognostic and predictive value of the 21-gene recurrence score assay in postmenopausal women with node-positive, oestrogen-receptor-positive breast cancer on chemotherapy: a retrospective analysis of a randomised trial. Lancet Oncol 2010;11:55–65.
12 Paik S, Tang G, Shak S, et al: Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. J Clin Oncol 2006;24:3726–3734.
13 Bast RC Jr, Hortobagyi GN: Individualized care for patients with cancer – a work in progress. N Engl J Med 2004;351:2865–2867.
14 Antonov J, Popovici V, Delorenzi M, et al: Molecular risk assessment of BIG 1-98 participants by expression profiling using RNA from archival tissue. BMC Cancer 2010;10:37.
15 Oberli A, Popovici V, Delorenzi M, et al: Expression profiling with RNA from formalin-fixed, paraffin-embedded material. BMC Med Genomics 2008;1:9.
16 Habel LA, Shak S, Jacobs MK, et al: A population-based study of tumor gene expression and risk of breast cancer death among lymph node-negative patients. Breast Cancer Res 2006;8:R25.
17 Glas AM, Floore A, Delahaye LJ, et al: Converting a breast cancer microarray signature into a high-throughput diagnostic test. BMC Genomics 2006;7:278.
18 Buyse M, Loi S, van't Veer L, et al: Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. J Natl Cancer Inst 2006;98:1183–1192.
19 Sparano JA, Paik S: Development of the 21-gene assay and its application in clinical practice and clinical trials. J Clin Oncol 2008;26:721–728.
20 Cardoso F, van't Veer L, Rutgers E, et al: Clinical application of the 70-gene profile: the MINDACT trial. J Clin Oncol 2008;26:729–735.
21 Mouridsen H, Giobbie-Hurder A, Goldhirsch A, et al: Letrozole therapy alone or in sequence with tamoxifen in women with breast cancer. N Engl J Med 2009;361:766–776.
22 Antonov J, Goldstein DR, Oberli A, et al: Reliable gene expression measurements from degraded RNA by quantitative real-time PCR depend on short amplicons and a proper normalization. Lab Invest 2005;85:1040–1050.
23 Giltnane JM, Ryden L, Cregger M, et al: Quantitative measurement of epidermal growth factor receptor is a negative predictive factor for tamoxifen response in hormone receptor positive premenopausal breast cancer. J Clin Oncol 2007;25:3007–3014.
24 van Agthoven T, Sieuwerts AM, Meijer-van Gelder ME, et al: Relevance of breast cancer antiestrogen resistance genes in human breast cancer progression and tamoxifen resistance. J Clin Oncol 2009;27:542–549.
25 Meijer D, Sieuwerts AM, Look MP, et al: Fibroblast growth factor receptor 4 predicts failure on tamoxifen therapy in patients with recurrent breast cancer. Endocr Relat Cancer 2008;15:101–111.
26 Butt AJ, McNeil CM, Musgrove EA, Sutherland RL: Downstream targets of growth factor and oestrogen signalling and endocrine resistance: the potential roles of c-Myc, cyclin D1 and cyclin E. Endocr Relat Cancer 2005;12(suppl 1):S47–S59.
27 Span PN, Tjan-Heijnen VC, Manders P, et al: Cyclin E is a strong predictor of endocrine therapy failure in human breast cancer. Oncogene 2003;22:4898–4904.
28 Meijer D, Jansen MP, Look MP, et al: TSC22D1 and PSAP predict clinical outcome of tamoxifen treatment in patients with recurrent breast cancer. Breast Cancer Res Treat 2009;113:253–260.
29 Bostner J, Ahnstrom Waltersson M, Fornander T, et al: Amplification of CCND1 and PAK1 as predictors of recurrence and tamoxifen resistance in postmenopausal breast cancer. Oncogene 2007;26:6997–7005.
30 Jirstrom K, Stendahl M, Ryden L, et al: Adverse effect of adjuvant tamoxifen in premenopausal breast cancer with cyclin D1 gene amplification. Cancer Res 2005;65:8009–8016.
31 Osborne CK, Bardou V, Hopp TA, et al: Role of the estrogen receptor coactivator AIB1 (SRC-3) and HER-2/neu in tamoxifen resistance in breast cancer. J Natl Cancer Inst 2003;95:353–361.
32 Yamashita H, Takahashi S, Ito Y, et al: Predictors of response to exemestane as primary endocrine therapy in estrogen receptor-positive breast cancer. Cancer Sci 2009;100:2028–2033.
33 Filipits M, Rudas M, Heinzl H, et al: Low p27 expression predicts early relapse and death in postmenopausal hormone receptor-positive breast cancer patients receiving adjuvant tamoxifen therapy. Clin Cancer Res 2009;15:5888–5894.
34 Pohl G, Rudas M, Dietze O, et al: High p27Kip1 expression predicts superior relapse-free and overall survival for premenopausal women with early-stage breast cancer receiving adjuvant treatment with tamoxifen plus goserelin. J Clin Oncol 2003;21:3594–3600.
35 Green AR, Burney C, Granger CJ, et al: The prognostic significance of steroid receptor co-regulators in breast cancer: co-repressor NCOR2/SMRT is an independent indicator of poor outcome. Breast Cancer Res Treat 2008;110:427–437.
36 Hurtado A, Holmes KA, Geistlinger TR, et al: Regulation of ERBB2 by oestrogen receptor-PAX2 determines response to tamoxifen. Nature 2008;456:663–666.
37 Heinze G, Puhr R: Bias-reduced and separation-proof conditional logistic regression with small or sparse data sets. Stat Med 2010;29:770–777.
38 Kaplan E, Meier P: Nonparametric estimation from incomplete observations. J Am Stat Assoc 1958;53:457–481.
39 Schemper M, Smith TL: A note on quantifying follow-up in studies of failure time. Control Clin Trials 1996;17:343–346.
40 Sotiriou C, Wirapati P, Loi S, et al: Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. J Natl Cancer Inst 2006;98:262–272.
41 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:R65.