Applying the 2011 St Gallen panel of prognostic markers on a large single hospital cohort of consecutively treated primary operable breast cancers

O. Brouckaert1*, A. Laenen2, J. Vanderhaegen1, H. Wildiers1, K. Leunen1, F. Amant1, P. Berteloot1, A. Smeets1, R. Paridaens1, M. R. Christiaens1, G. Floris1, P. Moerman1, E. Van Limbergen1, S. Peeters1, C. Weltens1, I. Vergote1 & P. Neven1

1Multidisciplinary Breast Centre, University Hospital Leuven, Leuven; 2Department of Electrical Engineering (ESAT-SISTA), Katholieke Universiteit Leuven, Leuven, Belgium

Received 21 December 2011; revised 30 January 2012; accepted 31 January 2012

Background: Many easily measurable and readily available factors are now established as being prognostic in primary operable breast cancer. We here applied the 2011 St Gallen surrogate definition for breast cancer subclassification using tumor grade instead of Ki67.

Patients and methods: Four thousand three hundred and eighteen consecutive patients who had surgery for primary operable breast cancer (1 January 2000 and 31 December 2009) in UZ Leuven excluding primary metastatic male breast cancers and those receiving neoadjuvant therapy. Five different surrogate phenotypes were created using the combined expression of estrogen receptor, progesterone receptor, human epidermal growth factor receptor-2 together with tumor grade. Disease-free interval (DFI), distant metastasis-free interval (DMFI), locoregional relapse-free interval (LRRFI), breast cancer-specific survival (BCSS) and overall survival (OS) were calculated.

Results: Surrogate phenotypes present with significant differences in DFI, DMFI, LRRFI, BCSS and OS. ‘Luminal A’ tumors presented with the best outcome parameters but the effect weakened at longer follow-up.

Conclusions: The four surrogate markers, agreed upon by the 2011 St Gallen consensus, defined five prognostic surrogate phenotypes in a large series of consecutively treated breast cancer patients. Their prognostic value changed with longer follow-up. The added value of gene expression profile over classical pathological assessment remains to be defined.

Key words: breast cancer, ER, HER-2, PR, subtypes

Introduction

Worldwide, breast cancer is the most frequently diagnosed malignancy in women; the lifetime probability of developing breast cancer is around one in eight [1–3]. Although breast cancer mortality has declined over the last decades in most developed countries, due to screening programs and improvements in adjuvant therapies, it remains the leading cause of cancer death [1, 3].

Prognostic heterogeneity in breast cancers has been appreciated since long but genetically and molecularly confirmed only a decade ago, with the discovery of five ‘intrinsic’ subtypes: luminal, human epidermal growth factor receptor-2 (HER-2) enriched, basal like, ‘normal breast-like’ and ‘claudin-low’ [4–21]. These subtypes exhibit different risk factors, pathological presentations, treatment response, clinical behavior and prognosis. In parallel, several prognostic gene expression profiles (GEP) have been developed, some of which are now commercially available [Oncotype DX (Genomic Health, Redwood City, CA); Mammaprint (Agenda BV, Amsterdam, The Netherlands); MapQuant DX (Ipsogen, Marseille, France)] [16, 18, 22–25]. Cost, complexity and lack of data supporting a predictive value for adjuvant therapy hampered introduction in daily practice. Although they were believed to replace the standard work-up consisting of immunohistochemistry (IHC) and pathological assessment, its now increasingly being recognized that both the methods largely overlap and are probably complementary only in a limited number of situations [18, 26, 27].

The established pathological work-up is easier and cheaper and therefore remains our golden standard in estimating the risk for breast cancer relapse and guiding clinical decisions regarding efficacy from endocrine, chemo or targeted (i.e. trastuzumab) therapy [26]. Several surrogate panels of prognostic and predictive markers, defined in a variety of ways, have been challenging GEP [28–40]. The 2011 St Gallen
patients and methods

patient selection

Cases were prospectively entered in our institutional database containing all patients with newly diagnosed invasive breast cancer treated at the Multidisciplinary Breast Centre in UZ-Leuven between 1 January 2000 and 31 December 2009. We retrieved those patients with primary surgery for early invasive disease. Women who received preoperative systemic adjuvant therapy \( n = 407 \) or presented with metastases at diagnosis \( n = 228 \) were excluded. Male breast cancer cases \( n = 28 \) and patients who had primary surgery and pathology reporting elsewhere were excluded as well \( n = 530 \). This resulted in the inclusion of 4318 consecutive cases.

tumor characteristics

For a detailed description of IHC and pathological assessment of tumor characteristics [tumor size, grade, nodal status, Nottingham Prognostic Index (NPI), NPI subgroups, ER, PR, HER-2], we refer to our previous work [30]. Expression of ER, PR and HER-2 was routinely assessed with IHC, using NLC-ER-6F11, NCL-PR-312 and CB11 antibodies, respectively (Novacosta Laboratories, Newcastle-Tyne, UK). Since 2005, highly sensitive rabbit monoclonal SP1 and SP2 antibodies were used (Labvision, Fremont, CA). Stainings were scored using the semiquantitative Allred score considering both the proportion of stained tumor cell nuclei (scored on a 0–5 scale) and staining intensity (scored on a 0–3 scale). A total sum of more than two was considered as positive for ER and PR status. The semiquantitative H-score was used before 2003, calculated by summing the products of the percentage of cells stained (0%–100%) by staining intensity \( 0–3+ \). For steroid receptors, any nuclear staining in invasive tumor cells was considered positive, both using Allred- or H-score. HER-2 immunostaining was scored according to the standardized HercepTest scoring system \( 0–3+ \). In intermediate scoring \( 2+ \) cases, HER-2 gene amplification was investigated by fluorescence in situ hybridization or FISH (PathVision; Vysis, Downers Grove, IL). Only when FISH was positive in score \( 2+ \) cases, HER-2 status was considered positive, score \( 3+ \) cases were considered positive irrespective of missing FISH.

subgroup definitions

We defined, according to the St Gallen recommendations, five breast cancer phenotypes based on ER, PR, HER-2 and tumor grade as luminal A, luminal B1, luminal B2, HER-2 like and basal like (Table 1). We applied the REcommendations for tumor MARKer prognostic studies (REMARK).

<table>
<thead>
<tr>
<th>Grade</th>
<th>Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Luminal A</td>
</tr>
<tr>
<td>II</td>
<td>Luminal B2</td>
</tr>
<tr>
<td>III</td>
<td>HER-2 like</td>
</tr>
<tr>
<td>IV</td>
<td>Basal like</td>
</tr>
</tbody>
</table>

Table 1. Interaction between ER, PR, HER-2 and grade with the St Gallen 2011 surrogate molecular classification [5]

PPN: ‘Positive’ for ER, ‘Positive’ for PR, ‘Negative’ for HER-2 and so on.

ER was missing in 10, PR in 23 and HER-2 in 89 patients as routine HER-2 analysis was introduced during 2000. Tumor grade was missing in 9 cases, tumor size in 9 patients \( p(Tx) \) and nodal status in 46 patients \( p(Nx) \).

As a result, a breast cancer phenotype could not be determined in 94 cases.

treatment

Description of surgical and adjuvant treatment modalities has previously been described [30]. Information on adjuvant therapy was missing in 20, 16, 11 and 1, respectively, for chemo, radio, endocrine and trastuzumab therapy.

follow-up

Patients were postoperatively followed according to ASCO guidelines up to September 2011.

statistical analysis

We compared the five phenotypes for several prognostic outcomes. The disease-free interval (DFI) was calculated as the time between surgery and first relapse, further distinguishing between locoregional (breast or nodal, ipsi- or contralateral) [locoregional relapse-free interval (LRRFI)] and distant metastasis relapse-free interval (DMFI). Patients without relapse or who died from unknown cause were censored at last follow-up. Patients who died from a nonbreast cancer-related cause were censored at time of death. We also studied the differences between the phenotypes regarding breast cancer-specific survival (BCSS) (time between surgery and breast cancer-related death) and overall survival (OS) (time between surgery and death from any cause). Statistics on the follow-up time are based on the Kaplan–Meier estimate of potential follow-up.

Differences between the five phenotypes regarding DFI and BCSS are graphically presented by the inverse of the cumulative incidence function, taking into account the presence of nonbreast cancer-related death as a competing risk. Calculations are based on the SAS macro %cuminc. Additionally, for DFI, a smoothed hazard function is given based on the cause-specific hazard for relapse, using the SAS macro %smooth. For OS, the differences between the five groups are visualized using the Kaplan–Meier curve.

In univariable analyses, differences between breast cancer phenotypes are tested using the log-rank test. When testing pairwise differences, the Bonferroni step-down correction for multiple testing was applied.

In multivariable analyses, we used the Cox proportional hazards model to compare luminal A with the four other phenotypes, correcting for a number of prognostic factors. The proportional hazards assumption was tested using an interaction model and the method of Lin, Wei and Ying (1993). When this assumption was violated, we extended the Cox model allowing non-proportional hazards by introducing an interaction term between phenotype and time.

Differences between the five phenotypes regarding clinical outcomes were analyzed, using the Kruskal–Wallis test in case of continuous outcomes and using the chi-square test in case of discrete outcomes. All
analyses have been performed using SAS software, version 9.2, of the SAS System for Windows. Copyright© 2002 SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC.

descriptive analysis

A surrogate molecular phenotype was successfully created for 4224 patients. Median and mean follow-up is 6.0 years. In total, 484 patients experienced a first relapse (366 distant metastasis and 166 locoregional relapses). When both distant and locoregional relapse appeared together, patients were allowed both in the DMFI and LRRFI analysis. We recorded 460 deaths, 202 breast cancer related, 165 nonbreast cancer related and 93 due to unknown cause. Significant differences between the five phenotypes in demographic characteristics, clinicopathological features and adjuvant therapy were summarized in Table 2.

Table 2. Descriptive analysis and outcome for the five phenotypes

<table>
<thead>
<tr>
<th>p, %</th>
<th>Luminal A</th>
<th>Luminal B1</th>
<th>Luminal B2</th>
<th>HER-2 like</th>
<th>Basal like</th>
<th>All</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative proportion</td>
<td>56.3%</td>
<td>22.4%</td>
<td>7.1%</td>
<td>4.2%</td>
<td>9.9%</td>
<td>100%</td>
<td>--</td>
</tr>
<tr>
<td>Age, mean/median</td>
<td>58.7/58.0</td>
<td>58.4/58.0</td>
<td>54.5/53.0</td>
<td>57.5/56.0</td>
<td>56.5/55.0</td>
<td>58.0/57.0</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Postmenopausal</td>
<td>1584 (66.6)</td>
<td>609 (64.3)</td>
<td>165 (54.8)</td>
<td>125 (70.6)</td>
<td>248 (59.2)</td>
<td>2788 (66.0)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>NPI, mean/median</td>
<td>3.7/3.4</td>
<td>5.2/5.0</td>
<td>4.9/4.6</td>
<td>5.1/4.9</td>
<td>4.9/4.6</td>
<td>4.3/4.3</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>NPI 1</td>
<td>1206 (50.7)</td>
<td>0 (0.0)</td>
<td>41 (13.6)</td>
<td>11 (6.2)</td>
<td>23 (5.5)</td>
<td>1316 (31.2)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>NPI 2</td>
<td>949 (39.9)</td>
<td>567 (59.9)</td>
<td>158 (52.5)</td>
<td>102 (57.6)</td>
<td>281 (67.1)</td>
<td>2097 (49.6)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Grade 1</td>
<td>624 (26.2)</td>
<td>0 (0.0)</td>
<td>6 (2.0)</td>
<td>1 (0.6)</td>
<td>5 (1.2)</td>
<td>652 (15.4)</td>
<td>--</td>
</tr>
<tr>
<td>Grade 2</td>
<td>1756 (73.8)</td>
<td>0 (0.0)</td>
<td>75 (24.9)</td>
<td>23 (13.0)</td>
<td>36 (8.6)</td>
<td>1938 (45.9)</td>
<td>--</td>
</tr>
<tr>
<td>Grade 3</td>
<td>0 (0.0)</td>
<td>947 (100)</td>
<td>218 (72.4)</td>
<td>153 (86.4)</td>
<td>377 (80.0)</td>
<td>1719 (40.7)</td>
<td>--</td>
</tr>
<tr>
<td>ER positive</td>
<td>2377 (99.9)</td>
<td>935 (98.7)</td>
<td>300 (99.7)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>3612 (85.5)</td>
<td>--</td>
</tr>
<tr>
<td>PR positive</td>
<td>2123 (89.2)</td>
<td>804 (84.9)</td>
<td>202 (67.1)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>3129 (74.1)</td>
<td>--</td>
</tr>
<tr>
<td>pN0</td>
<td>1523 (64.0)</td>
<td>490 (51.7)</td>
<td>164 (54.5)</td>
<td>91 (51.4)</td>
<td>276 (65.9)</td>
<td>2611 (61.8)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>pN1</td>
<td>629 (26.4)</td>
<td>306 (32.3)</td>
<td>87 (28.9)</td>
<td>52 (29.4)</td>
<td>107 (25.5)</td>
<td>1198 (28.4)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>pN2</td>
<td>134 (5.6)</td>
<td>92 (9.7)</td>
<td>32 (10.6)</td>
<td>22 (12.4)</td>
<td>25 (6.0)</td>
<td>309 (7.3)</td>
<td>--</td>
</tr>
<tr>
<td>pN3</td>
<td>67 (2.8)</td>
<td>49 (5.2)</td>
<td>15 (5.0)</td>
<td>11 (6.2)</td>
<td>10 (2.4)</td>
<td>152 (3.6)</td>
<td>--</td>
</tr>
<tr>
<td>pT1</td>
<td>1340 (56.3)</td>
<td>315 (33.3)</td>
<td>130 (43.2)</td>
<td>76 (42.9)</td>
<td>183 (43.7)</td>
<td>2107 (49.9)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>pT2</td>
<td>821 (34.5)</td>
<td>523 (55.2)</td>
<td>150 (49.8)</td>
<td>80 (45.2)</td>
<td>204 (48.7)</td>
<td>1802 (42.7)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>pT3</td>
<td>199 (8.4)</td>
<td>101 (10.7)</td>
<td>19 (6.3)</td>
<td>19 (10.7)</td>
<td>29 (6.9)</td>
<td>371 (8.8)</td>
<td>--</td>
</tr>
<tr>
<td>pT4</td>
<td>14 (0.6)</td>
<td>8 (0.8)</td>
<td>1 (0.3)</td>
<td>2 (1.1)</td>
<td>2 (0.5)</td>
<td>28 (0.7)</td>
<td>--</td>
</tr>
<tr>
<td>Histology: IDA</td>
<td>1873 (78.7)</td>
<td>873 (92.2)</td>
<td>285 (94.7)</td>
<td>173 (97.7)</td>
<td>412 (98.3)</td>
<td>3696 (87.5)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Histology: IIA</td>
<td>457 (19.2)</td>
<td>62 (6.5)</td>
<td>14 (4.7)</td>
<td>4 (2.3)</td>
<td>7 (1.7)</td>
<td>555 (13.1)</td>
<td>--</td>
</tr>
<tr>
<td>Histology: mixed</td>
<td>50 (2.1)</td>
<td>12 (1.3)</td>
<td>1 (0.3)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>66 (1.6)</td>
<td>--</td>
</tr>
<tr>
<td>Mastectomy</td>
<td>909 (38.2)</td>
<td>452 (47.7)</td>
<td>147 (48.8)</td>
<td>123 (69.5)</td>
<td>181 (43.2)</td>
<td>1812</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Axillary lymph node dissection</td>
<td>1710 (71.9)</td>
<td>767 (81.0)</td>
<td>235 (78.1)</td>
<td>153 (86.4)</td>
<td>325 (77.6)</td>
<td>3190</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Endocrine therapy</td>
<td>2240 (94.1)</td>
<td>881 (93.0)</td>
<td>281 (93.4)</td>
<td>4 (2.3)</td>
<td>9 (2.2)</td>
<td>3415</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Radiotherapy</td>
<td>2030 (85.3)</td>
<td>821 (86.7)</td>
<td>255 (84.7)</td>
<td>139 (78.5)</td>
<td>350 (83.5)</td>
<td>3595</td>
<td>&lt;0.0520</td>
</tr>
<tr>
<td>Chemotherapy</td>
<td>546 (22.9)</td>
<td>436 (46.0)</td>
<td>208 (69.1)</td>
<td>137 (77.4)</td>
<td>313 (74.7)</td>
<td>1640</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Endocrine + chemotherapy</td>
<td>530 (22.3)</td>
<td>390 (41.2)</td>
<td>194 (64.5)</td>
<td>3 (1.7)</td>
<td>6 (1.4)</td>
<td>1123</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Trastuzumab</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>120 (39.9)</td>
<td>63 (35.6)</td>
<td>0 (0.0)</td>
<td>183 (0.339)</td>
<td>--</td>
</tr>
</tbody>
</table>

BCSS, breast cancer-specific survival; DFI, disease-free interval; DMFI, distant metastasis-free interval; ER, estrogen receptor; IDA, invasive ductal adenocarcinoma; IIA, invasive lobular adenocarcinoma; LRRFI, locoregional relapse-free interval; NPI, Nottingham Prognostic Index; OS, overall survival; PR, progesterone receptor.

DFI, LRRFI and DMFI by surrogate phenotype

Luminal A presented with a significantly better DFI when compared with all four other phenotypes (Table 2); luminal B2 had a better DFI when compared with basal like (P = 0.0190) and HER-2 like (P = 0.0456), but significance for the latter two tests disappeared when correcting for multiple testing. Basal like and HER-2 like patients relapsed earlier (both locoregional and distant) compared with the three other phenotypes (Figure 1A and B). The assumption of proportional hazards was rejected, meaning that ratios in relapse risk between phenotypes varied over time.
Supplemental Figure S1 (available at *Annals of Oncology* online) presents the smoothed hazard functions and also illustrates the higher short-term risk for HER-2 like and basal like tumors, while long-term risk is low, relative to the other phenotypes. The relapse risk for luminal A cases was nearly constant over time up to 7 years but increased thereafter. Luminal B2 presented with high short-term, low mid-term and an increased late-term risk. A multivariable Cox model, correcting for the following possible confounders, is presented in Table 3: surrogate phenotypes, age at diagnosis, tumor size, nodal status, adjuvant use of chemotherapy, endocrine therapy and radiotherapy, type of breast and axillary surgery. To account for the above-mentioned time-dependent effect, an extended Cox model was used, allowing that phenotype effects vary over time. Supplemental Table S1 (available at *Annals of Oncology* online) presents hazard ratios with confidence intervals comparing luminal A with the other phenotypes. The first column presents the overall effect (representing average differences between the phenotypes over the entire follow-up period) and shows that surrogate phenotype is an independent significant predictor for relapse. The second and third columns present differences at specific time points (at 3 and 7 years of follow-up) and learns that the differences are mainly manifested at the first years after surgery and disappear after a longer follow-up, except for luminal A versus luminal B1.

At 5 years, local recurrences occurred less frequently in luminal A tumors (Table 2; Figure 1B). HER-2 like and basal like phenotypes have the highest incidence of locoregional relapse. Luminal B1 and luminal B2 tumors behave intermediately. A multivariable Cox model did not reveal an independent role of the any of the five surrogate phenotypes in LRRFI (Table 3 and supplemental Table S1, available at *Annals of Oncology* online). In contrast to DFI, the differences between phenotype did not change over time.

HER-2 like and basal like tumors metastasize in 15.2% and 16.3%, respectively, after 5 years of follow-up, compared with only 3.8% in luminal A tumors (Table 2; Figure 1A). Luminal B1 and luminal B2 have quite comparable relapse rates, with marginally more metastatic relapse in the luminal B1 phenotype. A multivariable extended Cox model confirmed the independent role of the phenotypes in a pattern similar to what we observed for DFI (Table 3 and supplemental Table S1, available at *Annals of Oncology* online). This confirms that distant relapse explains the time effects which we observed for DFI. This is also confirmed by formal tests for violations of the proportional hazards assumption.

**BCSS and OS**

The 5-year BCSS and OS in our cohort were 96.0% and 91.2%, respectively. Luminal A had significantly better BCSS compared with all other phenotypes (Table 2; Figure 1C and D). Basal like tumors have the worst BCSS. A multivariable model for BCSS retained surrogate phenotype as significant
parameter, together with nodal status, use of adjuvant chemotherapy and use of adjuvant endocrine therapy. Again, a time-dependent effect is present. Luminal A had a better BCSS compared with the other four phenotypes over the entire study period. However, at 7-year follow-up and in contrast with DFI and DMFI data, luminal B2 now joins luminal B1 in having significant worse outcome compared with luminal A.

After correction for multiple testing, luminal A remained to have a significant better OS compared with basal like (P < 0.0001), luminal B1 (P < 0.0001) and HER-2 like (P = 0.0492). However, when compared with luminal B2, statistical significance disappeared (P = 0.0937).

**Discussion**

In order to avoid the further use of heterogeneous surrogate panel definitions, we applied the 2011 St Gallen recommendation (Table 1) using the tumor grade instead of Ki67. This implies that (i) while the use of ER, PR, HER-2 and Ki67 has been validated against GEP [35], this alternative approach has not and (ii) while the recommended cut-off for Ki67 (14%) is well defined, there is no such recommendation for grade. In our study, surrogate markers for being classified as luminal A were grade 1 or 2 lesions (Table 1), considering (A) that luminal A is more prevalent than luminal B [35, 37], (B) that in one study when using a 10% cut-off for Ki67, 49.8% and 50.2% of grade 2 breast cancers were respectively lower and higher risk. Although cohorts and methodology differ, using Cheang’s 14% cut-off for Ki67 would probably yield more than half of grade 2 tumors to be ‘low risk’ or luminal A and (C) previously, two other studies also preferred to allow luminal A to be grade 1 or 2 in their study [37].

We confirm Cheang’s ability to prognostically separate luminal A from luminal B breast cancer using the 2011 St Gallen defined four-marker panel. Luminal A also differed significantly from HER-2 like and basal like breast cancers. Multivariate analysis confirmed the independent prognostic value of our surrogate panel (Table 3). Although some GEP perform better than surrogate panels, multivariate models in GEP studies are often limited in the number of clinicopathological features incorporated, and therefore, prognostic abilities of GEP may be overestimated [26, 36, 38]. Furthermore, we also described an important time-dependent effect. While over the entire study frame and at 3 years of follow-up, luminal A has significantly better outcome (DFI, DMFI, BCSS) compared with the four other phenotypes; this is however not true at 7 years of follow-up (supplemental Table S1, available at Annals of Oncology online).

Our results are not easily comparable with other surrogate panel studies for a number of reasons. For example, while Dawood et al. [37] preceded the St Gallen recommendation with an almost identical surrogate panel, differences are not neglectable. They used markers not recommended by the St. Gallen guidelines, defined HER-2-positive tumors differently, have around 5% non-classifiable breast cancers, their cohort dates from a different era, important differences in adjuvant treatment apply, and methodology differed (markers retrospectively carried out on archived tissue) [35, 37]. Our study might have shorter follow-up, it is one of the largest single centre-based cohorts reported to date, inherently more representative for 21st century international treatment guidelines. For all of the above reasons, and since, to our knowledge, other studies applying an identical surrogate panel are lacking, we decided it is not appropriate to stringently compare outcome data with other studies.

Although our study might have some strong points, it is not without caveats. First, we used different antibodies and scoring systems for ER and PR throughout the study period [30]. We considered ‘any nuclear staining’ as positive, although the most recent agreement is to use a 1% cut-off for the proportion of cells that stain. Although quantitative hormone receptor expression has not yet been validated against GEP (see below), misclassification of borderline positive ER/PR tumors may therefore have occurred. Second, we deliberately split luminal...
B1 from luminal B2. HER-2 overexpression in the latter warrants trastuzumab in most cases today, but trastuzumab was only routinely administrated since 2005 in our hospital. This leads to an important confounder but not in the luminal B1 phenotype. Exploratory analysis revealed that trastuzumab resulted in a nonsignificant trend toward better outcome in luminal B2 and in HER-2 like phenotypes (data not shown).

Both IHC classification and GEP have shortcomings. IHC markers may have a large inter- and intra-observer variability and GEP lack reproducibility, use different platforms, handle different statistical approaches for hierarchical clustering and use different methodologies [11, 14, 15]. GEPs are prognostic in and developed from untreated or endocrine only-treated patients and correlate well with routinely used markers such as age at diagnosis, size, histological grade and quantitative IHC PR expression. As for predictive purposes, classical pathological markers remain the golden standard. For a subgroup of patients, GEPs may prove to predict benefit from adjuvant chemotherapy in currently ongoing trials such as Tailor-X, Mindact and a recently started SWOG-trial. However, these trials mainly study ER-positive HER-2-negative patients with a different risk profile comparing classical pathological markers such as grade and nodal status and risk defined by GEP [21, 22]. For now, the classical work-up will remain the golden standard for patients not represented in these studies, but since the classical work-up may be improved for a number of reasons, it may persist to challenge GEP in any patient [21, 26]. First of all, Cheang et al. [34] previously showed that ‘basal like’ breast cancer defined by a five-marker panel has superior prognostic value than the ‘triple-negative’ phenotype. Second, loss of PR in ER-positive PR-positive HER-2-negative breast cancer significantly affects outcome [30]. PR has previously been shown to refine the prognostic value of ER but PR is not validated for subclassification of luminal tumors. Own data (not presented) show that PR-negative cases currently assigned to luminal A have a better outcome than those assigned to luminal B1. Therefore, they are well classified. However, within luminal B1 but not luminal A tumors, PR status seems to remain prognostic (8.8% absolute difference in development of metastasis over the entire study period) (subject to further study). Third, the introduction of quantitative rather than qualitative ER/PR information in surrogate panels is currently in study [21]. Fourth, nodal status could add invaluable prognostic information to a surrogate panel for at least three reasons. It has repeatedly been shown to be the single most important prognostic factor in operable breast cancer [21]. It carries time-dependent information that is missed in GEP as well in currently used surrogate panels [21]. Nodal status also remains a significant parameter in the multivariate Cox models for DMFI and BCSS (Table 3). Fifth, the same applies to tumor size, another important and time-dependent prognostic factor in breast cancer [21]. Sixth, new IHC markers are likely to be discovered and may further increase the performance of surrogate panels.

**conclusion**

Our results illustrated that a surrogate panel as suggested at the 2011 St Gallen meeting, using grade instead of Ki67, predicts relapse in women with a primary operable breast cancer, treated according to 21st century international treatment guidelines. We found no independent value of using surrogate panels for predicting local relapses. It should be appreciated that classical pathological assessment of breast cancer not only reflects the underlying genotype but also potentially reflects environmental and host influences as well as time-dependent effects, invaluable information currently not available in GEP models. Classical assessment and GEP are therefore considered complementary, each having their own strengths and weaknesses. Further refinement of surrogate panels and validation against GEP, or vice versa, are mandatory.

**disclosure**

The authors have declared no conflicts of interest.

**references**


