Gene Expression Profiling in Prediction of Distant Recurrence in HR-Positive and HER2-Negative Breast Cancer Patients

Nicolas Pennarun1, Yu-Hao Cheng, Tzu-Ting Huang, Cheng-Fang Horng and Skye Hung-Chun Cheng

1Department of Medical Research, Koo Foundation Sun Yat-Sen Cancer Center, Taiwan
2Department of Radiation Oncology, Koo Foundation Sun Yat-Sen Cancer Center, Taiwan

Abstract

Background: There had been several studies using gene-expression profiling in the prediction of distant recurrence in breast cancer. In this study, we developed an 18-gene classifier (18-GC) to predict distant recurrence of breast cancer and compared it in performance with the 21-gene panel (OncoType DX®, ODx) model.

Method: Included in the study were 224 patients treated for breast cancer with positive hormonal receptor (HR+) and negative human epidermal growth factor receptor 2 (HER2-). We compared the demographic, clinical, and survival information collected from the patients, and further compared the recurrence risk prediction obtained from microarray studies using the 18-GC with that obtained with the 21-gene panel (ODx). To have the best combined sensitivity and specificity, receiver operating characteristics (ROC) curve analysis was performed at the gene level to determine the cutoff values for several breakpoint scores.

Results: For the new 18-GC, a breakpoint score of 21 was adopted to produce a combined highest sensitivity (95 %) and specificity (39%) in detection of distant recurrence. At this breakpoint score, 164 of the 224 patients were classified by the 18-GC in the same risk level as by the 21-gene panel, giving a concordance rate of 73%. Along with patient age and tumor stage, this 18-GC was found to be independent significant prognostic factors of distant metastasis of breast cancer.

Conclusion: We created a new 18-GC for prediction of distant recurrence in HR+ and HER2- breast cancer patients. With a high concordance rate with ODx, the new panel we developed may serve as a good tool for individual breast cancer patients to make an informed decision on whether adjuvant chemotherapy should be performed post-surgery.

Keywords: Breast cancer; Distant recurrence; Microarray; Gene-expression profiling; Prediction; Chemotherapy

Introduction

While breast cancer in Asia is characterized by a lower incidence rate than in the United States and Europe [1,2], it is still one of the leading causes of cancer death in Asia, particularly in Taiwan [3-5]. This disease is featured by its complexity due to the genetic heterogeneity of breast carcinomas [6]. Since a few decades ago, many gene-expression profiling studies of breast cancer have revealed the existence of four major subtypes differing markedly in prognosis: luminal-A, luminal-B, HER2-amplified, and basal-like [6], prevalence of which varies by racial/ethnic groups [7]. Among the Asian populations, prevalence of the luminal-A, luminal-B, HER2+/HER2- basal-like, and unclassified subtype has been shown to be 55-65%, 10-20%, 10-15%, 10-15%, and 0-5%, respectively [8,9]. In the current study, we focused on luminal-like breast cancer comprising the luminal-A and luminal-B subtypes, which are defined by the presence of hormonal receptor and absence of HER2 on the plasma membrane of tumor cells (i.e., HR+/HER2-) by the immunohistochemistry [10]. Several studies have suggested that luminal-like cancers tend to have the most favorable prognosis and longer-term survival when compared with the other subtypes [11,12]. However, early-stage breast cancer patients with the luminal-like subtypes are commonly (up to 75%) over treated with adjuvant chemotherapy despite that recent studies have indicated that adjuvant chemotherapy may not provide significant benefit in reducing risk of recurrence [13,14]. To overcome this issue, several multigene panels,
such as Oncotype DX® (ODx), MammaPrint® and EndoPredict® assay kits, have been developed to help clinical decision-making regarding adjuvant chemotherapy for patients with early-stage breast carcinomas. ODx (Genomic Health Inc., Redwood, CA) is a prognostic and predictive assay kit for women with HR+ and HER2-breast cancer. It is a 21-gene RT-PCR assay for 16 cancer-related and five housekeeping control genes with an aim to aid physicians and patients to determine the best course of treatment by predicting the risk of distant recurrence of breast cancer. The ODx assay produces a numerical recurrence score and places patients into three categories: low-, intermediate- and high-risk [15]. Another test (MammaPrint by Agendia BV, Amsterdam, Netherlands) is a microarray-based gene-expression profiling assay that can classify the risk of distant recurrence into two categories, low and high, by analyzing 70 genes of HR+ and node-negative patients that had not received adjuvant systemic therapy [16]. The third test, EndoPredict (Myriad Genetics, Salt Lake City, UT), is a 11-gene RT-PCR test that provides prognostic information regarding the risk of distant recurrence of breast cancer to patients with HR+ and HER2- tumors [17]. The assay measures the expression of eight cancer-related and three control genes, and classifies patients under endocrine therapy into low- or high-risk of distant recurrence. Even though all such tools can help in making treatment decision on adjuvant chemotherapy for patients with early-stage breast carcinoma, none of them was originally developed for Asian patients even though Asian people may have different mechanisms in breast cancer due to multiple factors such as ethnic, environmental, and genetic variations [18-20]. To overcome such potential limitations, we developed an 18-gene classifier (18-GC) with tumor tissues obtained from Asian breast cancer patients and compared its performance with that of ODx in predicting distant metastasis in early-stage HR+/HER2- patients.

**Materials and Methods**

**Patient selection**

Retrospectively we included in this study a total of 224 luminal-like (HR+/HER2- ) and T1-3N0-1 breast cancer patients treated at Koo Foundation Sun Yat-Sen Cancer Center (KFSYSCC) in Taipei, Taiwan between 2005 and 2012, for evaluation of the 18-GC developed in our institute [21]. The institutional review board of KFSYSCC reviewed and approved the protocol and informed consent documents for the study. Eligible patients had invasive breast cancer; surgery as first treatment (mastectomy or breast-conserving surgery); a positive test result for estrogen or progesterone receptors (HR+); a

![Figure 1: ROC curve analyses of the 18-GC and ODx.](image-url)
negative test result for HER2 (HER2-); a few positive lymph nodes between 0 and 3. Patients with an N2, N3 or M1 stage and treated with pre-operative chemotherapy were excluded.

The 18-gene classifier

Developed based on 135 breast cancer patients treated in KFSYSCC, the new classifier includes a panel of 18 genes, all of which are significantly related to loco regional recurrence after mastectomy [21]. With a range of risk scores between zero and 56, the breakpoint value of 21 was used to separate the low- from the high-risk category [21]. With a range of risk scores between zero and 56, the breakpoint value of 21 was used to separate the low- from the high-risk category [21]. Unlike ODx, we did not include an intermediate risk group because it is usually binary in clinic decision-making. By adopting the same statistical predictive model used by Paik et al. [22].

The raw recurrence score ($X_i$) is first calculated by using the following expression:

$$X_i = 0.47\times\text{GRB7 group score} - 0.34\times\text{ER group score} + 1.04\times\text{proliferation group score} + 0.10\times\text{invasion group score} + 0.05\times\text{CD68} - 0.08\times\text{GSTM1} - 0.07\times\text{BAG}.$$

The final recurrence score ($Y_i$) was then calculated by transforming $X_i$ using the following expression:

$$Y_i = (X_i - 5.1031) + 100 \times 1 / (10.7148 - 5.103)$$

**Table 2: Genes selected for gene-expression profiling analysis.**

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Gene title</th>
<th>GenBank accession number</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRPV6</td>
<td>Transient receptor potential cation channel, subfamily V, member 6</td>
<td>NM_018646</td>
</tr>
<tr>
<td>DDX39</td>
<td>DEAD (Asp-Glu-Ala-Asp) box polypeptide 39</td>
<td>NM_005804</td>
</tr>
<tr>
<td>BUB1B</td>
<td>Building uninhibited by benzimidazoles 1 homolog beta (yeast)</td>
<td>NM_001211</td>
</tr>
<tr>
<td>CCR1</td>
<td>Chemokine (C-C motif) receptor 1</td>
<td>NM_001295</td>
</tr>
<tr>
<td>STIL</td>
<td>SCL/TAL1 interrupting locus</td>
<td>NM_003035</td>
</tr>
<tr>
<td>BLM</td>
<td>Bloom syndrome</td>
<td>NM_00057</td>
</tr>
<tr>
<td>C16orf7</td>
<td>Chromosome 16 open reading frame 7</td>
<td>NM_004913</td>
</tr>
<tr>
<td>PIM1</td>
<td>Pim-1 oncogene</td>
<td>NM_002648</td>
</tr>
<tr>
<td>TPX2</td>
<td>TPX2, microtubule associated</td>
<td>NM_012112</td>
</tr>
<tr>
<td>PTI1</td>
<td>Homo sapiens elongation factor 1-alpha 1</td>
<td>NM_001402</td>
</tr>
<tr>
<td>TCF3</td>
<td>Transcription factor 3 (E2A immunoglobulin enhancer binding factors E12/E47)</td>
<td>NM_003200</td>
</tr>
<tr>
<td>CCNB1</td>
<td>Cyclin B1</td>
<td>NM_031966</td>
</tr>
<tr>
<td>DTX2</td>
<td>Deltex 2, E3 Ubiquitin Ligase</td>
<td>NM_020892</td>
</tr>
<tr>
<td>ENSA</td>
<td>Endosulfine alpha</td>
<td>NM_004436</td>
</tr>
<tr>
<td>RCHY1</td>
<td>Ring Finger And CHY Zinc Finger Domain Containing 1, E3 Ubiquitin Protein Ligase</td>
<td>NM_015436</td>
</tr>
<tr>
<td>NFATC2IP</td>
<td>Nuclear Factor Of Activated T-Cells, Cytoplasmic, Calcineurin-Dependent 2 Interacting Protein</td>
<td>NM_032815</td>
</tr>
<tr>
<td>OBSL1</td>
<td>Obscurin-like 1</td>
<td>NM_015311</td>
</tr>
<tr>
<td>MMP15</td>
<td>Matrix Metallopeptidase 15 (Membrane-Inserted)</td>
<td>NM_002428</td>
</tr>
</tbody>
</table>

Statistical analysis

The demographic, clinical, and survival information were collected and analyzed among patients with and without metastasis. Crude and adjusted Cox analyses were used to compare patients in the low- and high-risk groups assigned by the 18-GC. ROC curve analyses were then performed to identify the optimal breakpoint [23,9]. We then evaluated sensitivity, specificity, accuracy, Negative Predictive Value (NPV), Positive Predictive Value (PPV), and the Area under the Curve (AUC) to determine how well the new 18-GC prediction model performs as compared with the ODx assay. All the statistical analyses (p < 0.05) were performed using SAS Software, version 9.4.
Table 3: Concordance between the 18-GC and ODx.

<table>
<thead>
<tr>
<th>Risk group</th>
<th>18-gene classifier</th>
<th>Oncotype Low (&lt;18)</th>
<th>High (≧21)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low (&lt;21)</td>
<td>30 (75.0%)</td>
<td>10 (25.0%)</td>
<td>40 (17.9%)</td>
<td></td>
</tr>
<tr>
<td>High (≧21)</td>
<td>50 (27.2%)</td>
<td>134 (72.8%)</td>
<td>184 (82.1%)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>80 (35.7%)</td>
<td>144 (64.3%)</td>
<td>224 (100%)</td>
<td></td>
</tr>
</tbody>
</table>

In addition, we calculated the PPV and NPV for patients who was not treated by chemotherapy (n=39). Results showed that the PPV is at 10% and the NPV at 100%, meaning that the 18 GC is well precise for determining the risk if a patient will have distant metastasis or not.

Results

Patient characteristics

Two-hundred twenty-four HR+ and HER2- breast cancer patients were included in the study (Table 1) with 202 (90.2%) of the patients diagnosed with no metastasis during the development of their breast cancer and 185 (82.6%) treated with adjuvant chemotherapy. Among those receiving chemotherapy, 165 (89.2%) of the patients did not develop metastasis during their follow-up. We found that two characteristics were significantly associated with the presence of metastasis in unadjusted analysis: a) an age at 40 or younger; b) a stage of T2-T3 (Table 1). Among the patients with no metastasis, 81.2% were over 40 years old, whereas only 54.5% of the patients with metastasis were categorized in the same age group (p=0.011). On the other hand, probability for patients with a T2-T3 tumor stage to develop metastasis is significantly higher than that for patients with a T1 tumor stage (81.8% with metastasis vs. 47.5% without metastasis for T2-T3 compared with 18.2% with metastasis vs. 52.5% without metastasis for T1, p=0.003).

Determination of the breakpoint score

For the breakpoint determination in the original panel of the 18 genes selected previously (Table 2), we compared the ROC curves with different cutoff points that stratify patients into two groups: low- or high-risk of recurrence. We selected a breakpoint score of 21 (Figure 1) since this score minimized the distance on the ROC curve to the left top edge of the diagram and produced a combined greatest sensibility (95.4%) and specificity (39.1%). By using this breakpoint, 80 (35.7%) of the 224 patients were classified as having low risk and 144 (64.3%) high risk. Patients classified in the intermediate-group by ODx were considered high risk when in comparison with the 18-gene panel. By using a breakpoint score of 21 in our gene panel assay, a total of 164 patients was classified in the same risk level as the ODx assay (73.2% concordance), indicating a significant agreement in the outcome predictions for individual patients (Table 3).

Recurrence-free survival by distant metastasis

To evaluate the prognostic power of the 18-GC, we compared the status predicted by the 18-GC and the actual distant metastasis status (Table 4). Even though the calculated PPV at 14.6 is relatively low, the calculated NPV is relatively high at 98.8%, indicating that the 18-GC is relatively accurate in identifying patients that would not have distant metastasis in the end of clinical monitoring.

Funding Support

This research was supported by the research fund of the Ministry of Science and Technology (NSC 102-3114-Y-368-001) of the Republic of China (Taiwan).

Competing Interests

The author (SHC) owns a patent relating to the content of this manuscript (Taiwan patent number: 104115832). None of the authors has any conflicts of interests in this research, either financial or non-financial.

References


