High \textit{BRCA1} gene expression increases the risk of early onset distant metastasis in ER$^+$ breast cancers

Hui-Ju Chang$^1$, Ueng-Cheng Yang$^{1,2,*}$, Mei-Yu Lai$^3$, Chen-Hsin Chen$^{3,4}$, Yang-Cheng Fann$^5$

$^1$ Institute of Biomedical Informatics, National Yang-Ming University (No. 155, Sec. 2, Li-Nong street, Taipei, Taiwan 11221, R.O.C.)
$^2$ Center for Systems and Synthetic Biology, National Yang-Ming University (No. 155, Sec. 2, Li-Nong street, Taipei, Taiwan 11221, R.O.C.)
$^3$ Institute of Statistical Science, Academia Sinica, (128 Academia Road, Section 2, Taipei, Taiwan 115201, R.O.C.)
$^4$ Graduate Institute of Epidemiology and Prevention Medicine, National Taiwan University, (17 Hsu-Chow Road, Taipei, Taiwan 100025, R.O.C.)
$^5$ Division of Intramural Research, National Institute of Neurological Disorders and Stroke, National Institutes of Health (35 Convent Dr, Bethesda, MD 20892, USA)

* Corresponding author:
Ueng-Cheng Yang
yang@ym.edu.tw
Abstract

**Background:** More than 30% of ER\(^+\) breast cancer patients developed distant metastasis after adjuvant tamoxifen therapy. It has been shown that decreased BRCA1 protein function confers anti-estrogen “resistance.” Since the functional status of BRCA1 protein auto-regulates its own gene expression, *BRCA1* transcript level should possibly be used to assess its active protein levels. Thus, it is interesting to explore the potential links between *BRCA1* gene expression status and the prognosis of ER\(^+\) breast tumors.

**Methods:** Early-staged ER\(^+\) Breast cancer microarray samples were selected from the NCBI Gene Expression Omnibus (GEO) database. The aggressiveness of these tumors was evaluated based on molecular subtypes or patients’ distant metastasis-free survival (DMFS) time found in associated GEO datasets. In survival analysis, the optimal threshold to define high and low transcript levels was assessed by the logistic-accelerated failure time mixture regression model in a pooled data set, which simultaneously envisages the lifetime risk of distant metastasis and the distribution of the DMFS time from the surgery of a patient susceptible to distant metastasis.
Results: On the basis of molecular subtyping, the relatively aggressive Luminal B tumor cells expressed significantly more BRCA1 transcripts than the less aggressive Luminal A tumor cells. This observation was supported by DMFS time analysis. When the upper 10% of the high BRCA1 expression patients was compared with the rest of the patients, the result was quite drastic. In patients susceptible to distant metastasis, the median onset time of distant metastasis for high BRCA1 expression group (2.2 years) was about one-fifth of the BRCA1 low expression group (10.5 years).

Conclusions: High BRCA1 transcript level does not increase the incidence of distant metastasis, but increases the risk of early onset distant metastasis in ER+ breast cancers.

As new therapeutic drugs have been developed for tumors with BRCA function loss, our findings suggest that BRCA1 gene expression test not only is useful for predicting patients’ outcome but also can be used in treatment decision.

Key words: BRCA1, early breast cancer, distant metastasis-free survival, ER+, tamoxifen, cure model
Introduction

Breast cancer has been the most common and the leading cause of cancer deaths in women worldwide [1,2]. Most breast cancers are sporadic and have no family history.

Based on their molecular characteristics [3], breast cancer can be classified into different subtypes. These subtypes may help predicting patient outcome and serve as targets for adjuvant therapies. According to the American Cancer Society, two out of three breast cancers are positive for hormone receptors and most of them are estrogen receptor positive (ER⁺). Binding to ER, estrogen may activate estrogen-ER signaling pathway and promote breast tumor formation [4,5] or induce the proliferation of ER⁺ breast tumor cells [6-8]. Therefore, a standard adjuvant therapy to treat ER⁺ breast cancers is to suppress the estrogen-induced proliferation by anti-estrogen therapies, such as using tamoxifen. However, despite the clinical success of tamoxifen, about 22-52% of early ER⁺ breast tumors exhibit recurrence in long-term follow up studies after this anti-estrogen treatment [9]. What causes this resistance is not clear yet, but decreased

Breast cancer type 1 susceptibility protein (BRCA1) has been reported to confer
tamoxifen resistance [10].

In addition to its well-known role in DNA damage repair [11], BRCA1 has many important biological functions. One of them is to regulate mammary cell growth and differentiation through interacting with estrogen-ER signaling pathway [12,13]. By forming a complex with ER, BRCA1 induces ER conformation change and regulates ER’s transcriptional activity [12]. Loss of BRCA1 function increases ER-coactivator association in tamoxifen treated cancer cells, which leads to tamoxifen resistance [10]. In fact, decreased BRCA1 protein staining is commonly observed in sporadic breast cancers [14,15] and this phenomenon is associated with high histologic grades [16-18] and short disease-free survival [19]. It is certain that such association is related to an E-ER signaling pathway.

If BRCA1 function does play a decisive role in breast cancer prognosis, this protein should be a good predictor for patients’ prognosis or even can serve as a target for treatment. However, there are still many problems to be solved before this idea can be applied clinically. One of them is how to develop a precise measuring tool to check the
active BRCA1 protein level in a tumor. Gene-expression based methods, such as PCR, could be good candidates, readily quantifiable, reliable, and cost-effective.

When the BRCA1 protein loses its function, a cell may undergo more mutations due to defective DNA repair and be transformed into a cancer cell. As BRCA1 protein auto-regulates its own gene transcription [20], this loss-of-function mutation may lead to a high $BRCA1$ gene expression. Therefore, high $BRCA1$ expression might be associated with aggressive behavior in breast tumors. This assumption was supported by previous study results showing that $BRCA1$ mRNA level was up-regulated in tumors [21-23] and that high $BRCA1$ gene expression was associated with fast metastasis [24]. However, some earlier studies presented opposite findings [25-28]. These controversial observations need to be resolved before exploring the possibility of using a gene-expression test to check active BRCA1 protein function.

This study used gene expression microarrays from the public domain to explore the association between $BRCA1$ mRNA level and the aggressiveness of breast tumors. Molecular subtypes and distant metastasis-free survival (DMFS) were used to assess the
aggressiveness of tumors, respectively. Among all different molecular subtypes of breast
cancers, only luminal A and B subtypes were selected to compare \textit{BRCA1} mRNA levels
because both subtypes are ER$^+$. In survival analysis, only early-stage ER$^+$ patients who
received standard tamoxifen treatment after surgery were included in the DMFS analysis
in attempt to attain homogeneity among study subjects. A “cure model” [29] was utilized
in this study to tackle the non-susceptibility of distant metastasis, which would be ignored
by the commonly-used Cox proportional hazards regression model [30] in survival
analysis.

\textbf{Materials and Methods}

\textbf{Study samples}

Microarray data used in this study were all from the Gene Expression Omnibus (GEO)
repository [31] located at the National Center for Biotechnology Information (NCBI). To
examine \textit{BRCA1} gene expressions in molecular subtypes of ER$^+$ breast tumors, we chose
dataset GSE45827 [32], which is composed of 130 primary invasive breast cancers (41
Triple Negative, 30 Her2, 29 Luminal A and 30 Luminal B tumors) as well as 11 normal
tissues and 14 cell lines using Affymetrix HG U133 plus 2.0 as the study platform. Because this study focused on ER+ tumors, only the BRCA1 expression levels of normal breast tissue, Luminal A, and Luminal B tumors from GSE45827 were compared.

Two criteria were considered in selecting the datasets for survival analysis: the first was to acquire datasets that tracked the survival time with related clinical data, and the second was to use datasets with an identical technology platform for microarray gene expression data. Ten breast cancer datasets with survival records, which used the U133A platform for expression analysis, were thus selected from the GEO database (see clinical information in Supplementary Table S1). However, because DMFS could be more directly associated with tumor behavior, the DMFS time was used as the endpoint of patients’ prognosis so that only those datasets with the DMFS time were selected for our study. To make our study samples homogenous, the following inclusion criteria were further used to select ER+ patients for early detection and treatment: 1) the tumor was localized at its original locus, 2) the lymph nodes were free of tumor, and 3) the tumor size was not larger than 5 cm. Study subjects must have an estrogen receptor status and adjuvant treatment
information. After thoroughly examining patients’ clinical information, only 359 patients in three GEO datasets (GSE12093 [33], GSE17705 [34] and GSE45255 [35]) were entered into our survival analysis (see Supplementary Table S2).

**Microarray data analysis**

The microarray data were normalized by using the Bioconductor R “affy” package [36]. In order to minimize the false-negative and false-positive rates, the parameters were set up based on Choe’s [37] study. Namely, bgcorrect.method = ”mas”, pmcorrect.method = ”pmonly”, normalize.method = ”quantiles”, and summary.method = ”medianpolish” were set up and input into the “affy” package’s “expresso” function. All samples of GSE45827 (on HG-U133 plus 2.0 platform) were normalized. The 573 samples pooled from three chosen GEO datasets, GSE12093, GSE17705 and GSE45255 (on HG-U133A platform), were normalized together. After the normalization, eligible study samples satisfying the aforementioned inclusion criteria were selected for further analysis.

The R package “jetset” [38] was used to select the most representative probe set from
multiple probe sets of a gene in the Affymetrix microarray. This program computes “jetscores” and uses the probe set of the highest score for a gene designated for further analysis. In Affymetrix HG U133 series, the *BRCA1* gene has two corresponding probe sets, 204531_s_at and 211851_x_at with “jetscores” of 0.278 and 0.029, respectively. Therefore, 204531_s_at was substituted for *BRCA1* gene expression in our analysis.

151

**Statistical methods**

**The Kruskal-Wallis test**

To compare the differences in the *BRCA1* gene expression levels among several groups, the nonparametric Kruskal-Wallis test was used to test the null hypothesis that these independent samples came from the same population distribution.

**Logistic-accelerated failure time (AFT) mixture regression model**

We employed the logistic-AFT mixture regression model [29] to compare two
Kaplan-Meier curves [39] representing the DMFS time between high-level and low-level $BRCA1$ expressions, which was defined by a specified cut-off point of $BRCA1$ expression values, with only right censored data. A patient was either an event case with distant metastasis or right censored with a follow-up DMFS time. Fitting a patient’s indicator covariate representing the high-$BRCA1$ expression into the mixture model simultaneously estimates her lifetime risk probability of distant metastasis and the distribution of her DMFS time if she would be susceptible to distant metastasis. A brief introduction to the mixture model is as follows:

Let $T$ be the duration in years from the surgery to distant metastasis in breast cancer, and the survival distribution at $t$ years $Pr(T > t)$ indicates the probability that a patient’s cancer will not metastasized for $t$ years. The patient’s eventual susceptibility to distant metastasis is represented by the binary indicator $D$, $D = 1$ indicating the patient is susceptible, $D = 0$ indicating the patient is not susceptible. The probability that a patient with a risk factor $Z$, such as an indicator of high-$BRCA1$ expression, will not metastasized after $t$ years is expressed as:
Pr(T > t|Z) = Pr(D = 1|Z)Pr(T > t|D = 1, Z) + Pr(D = 0|Z),

where, for a given Z, Pr(D = 0|Z) is the lifetime DMFS probability and Pr(T > t|D = 1, Z) is the conditional survival function at t years for a susceptible patient.

The logistic-AFT mixture regression model [29] contains both a logistic regression submodel

\[ \ln \left( \frac{Pr(D = 1|Z)}{Pr(D = 0|Z)} \right) = \beta' Z, \]

and an AFT location-scale regression submodel

\[ \ln T(Z) = \gamma' Z + \exp(\alpha' Z)\epsilon. \]

Parameters $\beta$, $\gamma$, and $\alpha$ are in the logistic, location, and scale regression parts, respectively, and can be simultaneously estimated using the R package “MixtureRegLTIC”[40]. For a specific value of Z, the logistic submodel provides a parametric estimator for the lifetime DMFS probability $Pr(D = 0|Z)$, which can also be estimated nonparametrically by the tail of the Kaplan-Meier overall DMFS curve. In contrast, the AFT submodelformulates the parametric distribution of the DMFS time in the logarithm scale for a susceptible patient, while the Kaplan-Meier conditional DMFS curve is displayed as its nonparametric counterpart.
Since the log-logistic distribution was much better fitted to our pooled GSE breast cancer datasets for the AFT submodel than the other generalized gamma distribution available in the MixtureRegLTIC, we assumed that $T$ of a susceptible patient follows a log-logistic distribution in the AFT location-scale regression submodel; that is, $\varepsilon$ is a logistic error term with the density $f(\varepsilon) = e^\varepsilon/(1 + e^\varepsilon)^2$.

According to the minimum Akaike information criterion (AIC), we determined the best model selection of the covariate to be included in the suitable regression parts of the logistic-AFT location-scale mixture regression model. To check the adequacy of the fitted model for high and low BRCA1 expressions, we plotted the overall survival curves $\Pr(T > t | Z)$ and the conditional curves $\Pr(T > t | D = 1, Z)$ by superimposing on the corresponding Kaplan-Meier overall and conditional curves to visualize the goodness-of-fit for the best-fitted mixture model. The resultant p-value of the likelihood ratio test in the best-fitted mixture model is used for evaluating the significance of the difference between the Kaplan-Meier overall DMFS curves.
Results

High BRCA1 gene expression is associated with aggressive behavior in ER$^+$ breast tumors

BRCA1 is known to regulate mammary cell proliferation through interacting with ER. Thus, we focused on ER$^+$ tumors (molecular subtypes luminal A and B) to make our samples homogeneous. The GEO dataset GSE45827 contains Affymetrix HG-U133 plus 2.0 microarrays of normal breast tissues and primary breast tumors with different molecular subtypes. It allows us to compare the BRCA1 expression levels in different molecular subtypes of ER$^+$ tumors. In Figure 1, BRCA1 expression levels in luminal B tumors were significantly higher than those in normal breast tissues or luminal A tumors. In contrast, BRCA1 expression levels were not different in normal tissue and luminal A tumors. Luminal A tumors are defined as tumors which are positive for ER but low in either HER2 or proliferation marker Ki-67 levels. These tumors grow slowly and have a good prognosis. Luminal B tumors are also ER$^+$ but with a high level of proliferation
marker Ki-67. In other words, luminal B tumors grow faster than luminal A tumors. Our observation reveals that high \textit{BRCA1} expression is associated with the proliferation of ER$^+$ cells.

**Figure 1. Boxplots of \textit{BRCA1} expression in normal breast tissue, Luminal A and Luminal B tumors of GSE45827.** Normal breast tissues (red), luminal A tumors (green) and luminal B tumors (blue) are displayed from left to right. Each dot represents the \textit{BRCA1} expression level of a certain sample. Pairwise p-values between different subtypes were calculated using the Kruskal-Wallis test.

High \textit{BRCA1} gene expression is associated with fast distant metastasis in ER$^+$ early-stage breast cancers

Molecular subtyping clearly demonstrated the association between \textit{BRCA1} gene
expression level and tumor aggressiveness. It is expected that this association can be reflected in the patients’ survival pattern. Thus, a pooled dataset with both microarray data and the survival data were collected from the GEO repository. Since a patient may not have a long-term survival once she develops a distant metastasis, the DMFS is a good endpoint for the aggressiveness of tumors.

The BRCA1 expression level is a continuous variable and the histogram, which is composed of the BRCA1 expression levels of the pooled samples from the three GEO datasets, is unimodal. To compare the differences in the DMFS time of high- or low-BRCA1 expression groups, a threshold for BRCA1 expression is needed. Different deciles in the histogram of BRCA1 expression distribution were used to categorize the patients into high- or low-expression groups, respectively.

Results from fitting the logistic-AFT mixture model were not significant for each cut-off point ranging from the 1st to 4th deciles (results not shown). Results falling into the 5th to 8th deciles were reported in Figure 2A-D and Table 1, and those for the cutoff-point at the 9th decile were displayed in Figure 3 and Table 2. In each of these figures, the left panel
indicates the overall DMFS curves and the right panel the conditional DMFS curves. The former displayed the DMFS pattern for all patients, and the latter showed the DMFS distribution of the patients who would be susceptible to distant metastasis. The time-axis of the conditional DMFS curve was plotted in a logarithmic scale and could be used to estimate the median onset time of distant metastasis for susceptible patients.

The difference between the Kaplan-Meier overall DMFS curves (step functions) for these two expression groups defined by a specified decile was then examined. It was noted that divided by higher deciles (such as 70% to 90%), the corresponding DMFS curves of the high-\textit{BRCA1} expression groups (the red step functions in the left panels of Figures 2C, 2D and 3) flatten off, respectively. In the high \textit{BRCA1} expression group, some patients developed distant metastasis early on around the first seven years after surgery. In contrast, the rest (more than 60%) of the patients rarely metastasized in the 15-year follow-up. These results implied that the hazard ratio of the high-expression group to the low one changed over the follow-up course, so using the logistic-AFT mixture regression model [29] to fit the pooled GEO data with \textit{BRAC1}, instead of the Cox proportional hazards model [30],
was quite adequate.

The mixture regression model enabled us to predict the DMFS accurately so that the resultant overall and conditional DMFS (smooth) curves defined at higher deciles derived from the model fitting were much closer to the corresponding empirical Kaplan-Meier overall and conditional DMFS curves (step functions). In addition to the significant p-values resulting from the likelihood ratio test (LRT) for the mixture models (cut-off from 50% to 90%), a cutoff at 60 to 90% (Tables 1 and 2) seems to emerge as a trend. The Kaplan-Meier conditional DMFS curves for patients susceptible to distant metastasis (right panels in Figure 2A to 2D and in Figure 3) found in the high- and low- BRCA1 expression groups were also separated more distinctly with increasing cut-off values. Our model fitting obviously confirmed the BRCA1-gene expression dependency with increasing BRCA1-expression levels. This trend was revealed with the statistical significance of the high-BRCA1 expression in the “Location Regression Part” in Tables 1 and 2, together with the well-separated low- and high-BRCA1 expression (smooth) curves in Figures 2 and 3. The separation due to a negative estimate of the location parameter resulted in a
smaller median onset time of distant metastasis in the high-\textit{BRCA1} expression group than in the lower one. On the other hand, each of the “Scale Regression Part” in Tables 1 and 2 did not include the high-\textit{BRCA1} expression, so the shapes of the low- and high-\textit{BRCA1} expression conditional DMFS curves were similar.

Table 1. Analysis of the log-logistic-AFT mixture regression model stratified by BRCA1-level with different cut-off points.

<table>
<thead>
<tr>
<th>Cut-off Decile</th>
<th>Covariate</th>
<th>Logistic Regression Submodel</th>
<th>AFT Submodel (Log-logistic Event Time Distribution)</th>
<th>Mixture Model p-value$^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Location Regression Part</td>
<td>Scale Regression Part</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Estimate 95% CI</td>
<td>Estimate 95% CI</td>
<td>p-value</td>
</tr>
<tr>
<td>50%</td>
<td>Intercept</td>
<td>-0.166, -1.773, 1.441</td>
<td>2.826, 1.764, 3.888</td>
<td>-0.390, -0.739, -0.041</td>
</tr>
<tr>
<td></td>
<td>BRCA1-H</td>
<td>-0.653, -1.163, -0.143</td>
<td>0.012</td>
<td></td>
</tr>
<tr>
<td>60%</td>
<td>Intercept</td>
<td>-0.330, -1.780, 1.120</td>
<td>2.605, 1.590, 3.620</td>
<td>-0.407, -0.764, -0.050</td>
</tr>
<tr>
<td></td>
<td>BRCA1-H</td>
<td>-0.557, -1.088, -0.026</td>
<td>0.040</td>
<td></td>
</tr>
<tr>
<td>70%</td>
<td>Intercept</td>
<td>-0.631, -1.505, 0.243</td>
<td>2.365, 1.681, 3.049</td>
<td>-0.494, -0.823, -0.165</td>
</tr>
<tr>
<td></td>
<td>BRCA1-H</td>
<td>-0.752, -1.332, -0.172</td>
<td>0.011</td>
<td></td>
</tr>
<tr>
<td>80%</td>
<td>Intercept</td>
<td>-0.799, -1.414, -0.184</td>
<td>2.183, 1.677, 2.689</td>
<td>-0.570, -0.866, -0.274</td>
</tr>
<tr>
<td></td>
<td>BRCA1-H</td>
<td>-1.016, -1.659, -0.373</td>
<td>0.002</td>
<td></td>
</tr>
</tbody>
</table>
Figure 2. Survival curves fitted by the “Logistic-AFT mixture regression model” at the 5th-8th deciles. Kaplan-Meier (step function) and the mixture regression (smooth curve) estimators of overall and conditional DMFS curves for the distant metastasis onset time after the surgery stratified by BRCA1-level defined at the cut-off point (A) 50th, (B) 60th, (C) 70th, and (D) 80th percentile, and corresponding p values of likelihood ratio test (PLRT) were 0.011, 0.038, 0.014 and 0.006, respectively. The BRCA1 high- and low-expression groups were shown in red and in blue, respectively, for each cut-off point.

Consequently, in Table 2, the p value of the LRT corresponding to the 9th decile was <0.001 (still significant even after adjustment for multiple comparison). It was the smallest among
the LRT p-values for all deciles; it suggested that the best-fitted model was based on the 90th percentile (of expression level 7.04) as the cut-off in the histogram for the pooled BRAC1 gene expression values. As shown in Figure 3, this cut-off point sharply contrasted the difference for the two Kaplan-Meier DMFS curves depicting the high and low levels of BRCA1 expression. The decreasing trend of the AIC values (results not shown) among increasing cut-off deciles implied that the possibility of over-fitting the model by using this optimal cut-off point rather than other points was low.

The logistic regression submodel (Table 2) did not include the high-BRCA1 expression variable. The estimated intercept (-0.520) indicated that the estimated common lifetime probability of the DMFS in both expression levels was 0.627. That is, 62.7% of the early-stage ER+ tamoxifen-treated breast cancer patients were free of distant metastasis during the 15 years of follow-ups, no matter what the BRCA1 expression level was. It also readily explained why the Kaplan-Meier overall DMFS curves depicting the BRCA1 high- (red step function) and low- (blue step function) expression groups in the left panel of Figure 3 crossed at similar values of the DMFS proportion right before the 15th year of
follow-up. The location parameter in the AFT location-scale regression submodel, however, included a high-\textit{BRCA1} expression variable with a negative parameter estimate -1.571 (p < 0.001 of the Wald test, Table 2). This indicated that \textit{BRCA1} had an impact on the distant metastasis after surgery time such that median onset time for susceptible patients in the high-\textit{BRCA1} expression group was significantly earlier than that of the low expression group. The left panel of Figure 3 revealed that the Kaplan-Meier curve illustrating the high \textit{BRCA1} expression group (red step function) dropped sharply and then flattened off afterwards. This was because distant metastasis in this group all occurred in the first 7 years. Therefore, the rest of the patients appeared to be “cured” towards the end of the 15-year follow-up period. Nevertheless, this phenomenon was not observed in the low \textit{BRCA1} expression group because patients with low \textit{BRCA1} expression gradually developed distant metastasis, \textit{i.e.}, their relapses occurred in a steady rate during the 15 years. The median onset time of distant metastasis in the high- and low-expression groups was estimated as 2.2 and 10.5 years, respectively. Because the median onset time of distant metastasis for susceptible patients in the high \textit{BRCA1} expression group was about one-fifth
of that in the low expression group, these patients seemed to have extremely aggressive tumors. Therefore, \textit{BRCA1} expression level may assist in identifying these high-risk patients. Patients in the high \textit{BRCA1} expression group should be treated differently from other patients.

\textbf{Figure 3. Survival curves fitted by the “Logistic-AFT mixture regression model” at the 9th decile.} Kaplan-Meier (step function) and the mixture regression (smooth curve) estimators of overall and conditional DMFS curves for the distant metastasis onset time after the surgery stratified by \textit{BRCA1}-level defined at the cut-off point 90th percentile. The \textit{BRCA1} high- and low-expression groups were shown in red and in blue, respectively.
Table 2. Analysis of the logistic-AFT mixture regression model stratified by BRCA1-level cut-off at the 90th percentile.

<table>
<thead>
<tr>
<th>Covariates</th>
<th>Estimate</th>
<th>95% CI</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Event Probability</td>
<td>Intercept</td>
<td>-0.520</td>
<td>-1.218, 0.178</td>
</tr>
<tr>
<td>Location</td>
<td>Intercept</td>
<td>2.351</td>
<td>1.836, 2.866</td>
</tr>
<tr>
<td>Location</td>
<td>BRCA1-H</td>
<td>-1.571</td>
<td>-2.263, -0.879</td>
</tr>
<tr>
<td>Scale</td>
<td>Intercept</td>
<td>-0.567</td>
<td>-0.841, -0.293</td>
</tr>
<tr>
<td>LRT mixture model</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: CI, confidence interval; LRT, likelihood ratio test; BRCA1-H, high BRCA1.

Discussion

Tamoxifen is usually used to treat the ER+ patients. This drug is a selective estrogen receptor modulator (SERM), competes with estrogen for ER-binding and inhibits the proliferation stimulated by estrogen. It was reported that decreased BRCA1 expression reduced the cellular response to tamoxifen [10] through increased ER-coactivator association. As BRCA1 protein can auto-regulate its own gene transcription [20], a loss of BRCA1 protein function may activate gene transcription and lead to high-BRCA1 gene levels.

In previous studies of the association between BRCA1 gene expression level and tumor
behavior, conflicting observations came to light. Our results show that high BRCA1 gene expression is associated with aggressive tumor behaviors and with fast distant metastasis in early stage ER\(^+\) breast cancers. Our findings may be contrasted with the survival study results reported by Hennigs et al. [41] from primary non-metastatic invasive breast cancer patients of different molecular subtypes, but we utilized survival data derived from a longer follow-up with additional BRCA1 gene expression information. Our study still has some limitations. We had expected to acquire a sufficiently large sample size of study subjects from the public domain; however, constrained by different study designs, varied or unclear clinical information items, and inconsistent follow-up periods found in the GSE datasets, it was necessary to adopt the pooled-data approach instead of meta-analysis in this study. Furthermore, the decile value cut at 90 % indicated the best result, but the sample size of this high-BRCA1 expression group was small. These factors might alter the statistical results, but the DMFS analysis illustrated by various cut-off points for the BRCA1 expression groups provided with coherent results showing a clear trend.
The high BRCA1 expression patients in this study displayed a biphasic survival curve, which implied that these patients could be divided into two subgroups. The subgroup susceptible to distant metastasis is likely to lose the BRCA1 protein function so that the disseminated cells proliferated fast. In contrast, the non-susceptible subgroup may retain a normal BRCA1 protein level. As a tumor suppressor, normal BRCA1 protein function would lead to DMFS. In other words, these two types of high BRCA1 expression patients may have different outcomes.

It is also noted that about 60% of patients in the pooled cohort were “cured” regardless of the presence of high or low expression of the BRCA1 gene. This observation indicated that BRCA1 expression status did not change the lifetime risk of DMFS. This cure proportion may be attributable to the effect of breast surgery combined with adjuvant tamoxifen therapy. Nevertheless, the time to distant metastasis was BRCA1 expression dependent. When the cutoffs were raised from 50% to 90%, the median onset times of distant metastasis for the susceptible patients of high-BRCA1 expression groups were gradually shortened, while those of lower-expression groups consistently retained (Figure...
This observation implied that the onset time might be affected by the proliferation rate of the disseminated tumors. Presumably, the higher the proliferation rate, the earlier the onset time is.

In fact, 20% of sporadic breast tumors [14] did not show BRCA1 staining in either the nucleus or cytoplasm, which demonstrated a defect in protein production. In some cases, breast tumor cells showed cytoplasmic BRCA1 retention along with a loss of nuclear staining [15], which implied a defect in nuclear transportation [42,43]. These BRCA1 protein-loss tumor cells were associated with poor prognosis and might induce high BRCA1 gene expression via auto-regulation.

Some tamoxifen resistant tumors retain responsiveness to estrogen and can be treated by other types of anti-estrogens, such as aromatase inhibitors (AIs) [44]. However, AIs have limitations because they can be only used for post-menopausal patients; therefore, for recurrent tumors with BRCA1 function loss, poly (ADP-ribose) polymerase (PARP) inhibitors might be a good alternative therapy. The PARP inhibitor was developed for BRCA-mutated [45] tumors, but now researchers seek to expand its use to high-risk
patients with similar molecular characteristics [46-48]. BRCA1 protein is involved in the homologous recombination repair pathway to repair DNA double strand breaks. With BRCA1 defects, tumor cells are forced to use alternative DNA-repair pathways [49], such as simple end joining, to fix DNA damage. By inhibiting alternative repair pathways, PARP inhibitors lead tumor cells to die, because all DNA double strand break repair capabilities for tumor cells were blocked. However, normal cells still retain a functional homologous recombination repair pathway, so they can survive quite well upon receiving PARP inhibitor treatment.

To develop a companion diagnostic for prescribing PARP inhibitors, the commonly used immunohistochemistry (IHC) staining may not be the best choice since the IHC is cumbersome and is not easy to standardize and automate. Furthermore, the specificity and sensitivity of some IHC antibodies are questionable [50]. In contrast, gene expression-based methods, such as polymerase chain reaction, are specific, sensitive, and quantitative. The latter has the potential to identify which recurrent patients are appropriate for PARP inhibitor therapy.
In summary, our study results can explain the conflicting results produced by previous studies. The high \textit{BRCA1} gene expression may either favor or not favor the patients’ survival. If the high \textit{BRCA1} gene expression leads to a high expression of normal \textit{BRCA1} protein, the patients appear to be cured. If, however, the high \textit{BRCA1} gene expression is caused by the loss of \textit{BRCA1} protein function, the patients are susceptible to fast distant metastasis.

**References**


estrogen signaling pathways collaborates with loss of Brca1 to promote development
of ERα-negative and ERα-positive mammary preneoplasia and cancer. Oncogene.
2008;27(6),794-802.

5. Wang L, Di LJ. BRCA1 and estrogen/estrogen receptor in breast cancer: Where they

6. Ciocca DR, Fanelli MA. Estrogen receptors and cell proliferation in breast cancer.

mediates proliferation of breast cancer MCF-7 cells via a p21/PCNA/E2F1-dependent

8. Tian JM, Ran B, Zhang CL, Yan DM, Li XH. Estrogen and progesterone promote
breast cancer cell proliferation by inducing cyclin G1 expression. Brazilian J Med

Breast-Cancer Recurrence after Stopping Endocrine Therapy at 5 Years. N Engl J


and TP53 marker genes expression in breast cancer patients. Int J Pharm Res. 2018;
10(2):147-54.


37. Choe SE, Boutros M, Michelson AM, Church GM, Halfon MS. Preferred analysis
methods for Affymetrix GeneChips revealed by a wholly defined control dataset.


42. Yang ES, Xia F. BRCA1 16 years later: DNA damage-induced BRCA1 shuttling.


48. McCann KE, Hurvitz SA. Advances in the use of PARP inhibitor therapy for breast


Declarations

Ethics approval and consent to participate

This work is a pooled analysis of public data from gene expression omnibus (GEO) at the National Center for Biotechnology Information. There is no need to get ethics approval.

Consent for publication

All authors agreed for this submission.

Availability of data and material

This work only analyzed public data; no new data were generated.

Competing interests
All authors declared no competing interests.

**Funding**

This work is partially supported by funding from 3 sources: 1) grant 103-2911-I-010-514 from Ministry of Science and Technology, Taiwan; 2) grant from 109-2321-B-002-038 from Ministry of Science and Technology, Taiwan; 3) Intramural Research Program of National Institute of Neurological Disorders and Stroke, National Institutes of Health, USA.

**Authors' contributions**

Hui-Ju Chang: conceived the idea, carried out the analysis, wrote the manuscript

Ueng-Cheng Yang: conceived the idea, revised the manuscript

Mei-Yu Lai: performed the cure model analysis

Chen-Hsin Chen: developed the logistic AFT cure model, revised the manuscript

Yang-Cheng Fann: helped with the programming, revised manuscript

**Acknowledgements**

We thank Mr. Wei-Hung Chuang for helpful discussion.