EBioMedicine 5 (2016) 74–81

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.ebiomedicine.com

Research Paper

An Eighteen-Gene Classifier Predicts Locoregional Recurrence in Post-Mastectomy Breast Cancer Patients

Skye H. Cheng [a,*], Chen-Fang Horng [a], Tzu-Ting Huang [a], Erich S. Huang [b], Mei-Hua Tsou [a], Li-Sun Shi [c], Ben-Long Yu [d], Chii-Ming Chen [a], Andrew T. Huang [a,e]

[a] Koo Foundation Sun Yat-Sen Cancer Center, Taipei, Taiwan
[b] Duke University Medical Center, Durham, NC, USA
[c] Department of Laboratory and Pathology, Koo Foundation Sun Yat-Sen Cancer Center, Taipei, Taiwan
[d] Department of Surgery, Koo Foundation Sun Yat-Sen Cancer Center, Taipei, Taiwan
[e] Duke University, USA

ARTICLE INFO

Article history:
Received 23 December 2015
Received in revised form 26 January 2016
Accepted 11 February 2016
Available online 16 February 2016

Keywords:
Breast cancer
Mastectomy
Breast-conserving surgery
Radiotherapy
Locoregional recurrence
Gene-expression profiles
Prediction

ABSTRACT

We previously identified 34 genes of interest (GOI) in 2006 to aid the oncologists to determine whether post-mastectomy radiotherapy (PMRT) is indicated for certain patients with breast cancer. At this time, an independent cohort of 135 patients having DNA microarray study available from the primary tumor tissue samples was chosen. Inclusion criteria were 1) mastectomy as the first treatment, 2) pathology stages I-III, 3) any locoregional recurrence (LRR) and 4) no PMRT. After inter-platform data integration of Affymetrix U95 and U133 Plus 2.0 arrays and quantile normalization, in this paper we used 18 of 34 GOI to divide the mastectomy patients into high and low risk groups. The 5-year rate of freedom from LRR in the high-risk group was 30%. In contrast, in the low-risk group it was 99% ($p < 0.0001$). Multivariate analysis revealed that the 18-gene classifier independently predicts rates of LRR regardless of nodal status or cancer subtype.

1. Introduction

The conventional method in determining the indication for post-mastectomy radiotherapy (PMRT) is largely based on clinical variables, such as tumor size, axillary lymph node involvement, hormone receptor status, age at diagnosis, lymphovascular invasion (LVI), etc. These factors are known risk factors associated with locoregional recurrence (LRR). They are, however, imperfect in predicting recurrence (Cheng et al., 2006a; Taghian et al., 2004). More reliable biological markers are being sought. Sporadic reports have attempted to show that some isolated genes could link to LRR (van der Hage et al., 2004; Zellars et al., 2000).

Patients with four or more axillary lymph-node involvement (N2 or N3 disease) generally would be given PMRT (Recht et al., 2001). However, there is controversy concerning patients with 1–3 positive nodes (N1 disease), even though NCCN guidelines "strongly consider" giving PMRT based on large meta-analyses from many randomized control trials (Clarke et al., 2005; Early Breast Cancer Trialists' Collaborative, 2002; EBCTCG et al., 2014). In fact, the likelihood of LRR after 10 years in node negative patients is reported to be approximately 2–8%, N1 patients 20%, and N2 patients 32% (Clarke et al., 2005; EBCTCG et al., 2014). Therefore, if we rely only on clinical parameters, around 70–80% of the node-positive patients could potentially undergo overtreatment, whereas those at risk in node negative disease could potentially be undertreated.

Recent progress in genomic analyses for evaluating tumor biology show significant agreement in the outcome predictions for individual patients who are probably sharing a common set of biologic phenotypes. This opened a new possibility to improve risk stratification that led to more personalized prognostication for breast cancer patients (Fan et al., 2006; Sorlie et al., 2001). Studies from gene expression profiling have shown a greater capability of determining prognosis and predicting response to adjuvant chemotherapy in Tamoxifen-treated patients (Paik et al., 2004a, 2004b, 2006). In 2006, we reported 34 and 258 gene sets that could partition the LRR high risk patients from the LRR low risk patients after mastectomy. The low risk group determined by the gene expression profiling had a 3-year LRR rate of less than 3%, and the high risk group had an LRR rate of more than 50% (Cheng et al., 2006b). In this study, we evaluated 34 genes of interest (GOI) in the prediction of LRR using a completely different patient population

* Corresponding author at: Department of Radiation Oncology, Koo Foundation Sun Yat-Sen Cancer Center, No. 125, Lih-Der Road, Pei-Tou District, Taipei, Taiwan.
E-mail addresses: skye@kfsyscc.org (S.H. Cheng), Cfhorng@kfsyscc.org (C.-F. Horng), htting@kfsyscc.org (T.-T. Huang), erich@icloud.com (E.S. Huang), MHTSOU@kfsyscc.org (M.-H. Tsou), lsshih210@kfsyscc.org (L.-S. Shi), benlong@kfsyscc.org (B.-L. Yu), cmchen@kfsyscc.org (C.-M. Chen), athuang@kfsyscc.org (A.T. Huang).

http://dx.doi.org/10.1016/j.ebiom.2016.02.022

and aimed to identify individuals with low risk of recurrence for whom PMRT could be avoided.

2. Material and Methods

2.1. Patients

Patients with invasive breast cancer were collected in a study of "*Precision Medicine in Oncology for Taiwan*" at Koo Foundation Sun Yat-Sen Cancer Center in an attempt to develop a new taxonomy for breast cancer, which was approved by the Bio-bank Ethics Committee and Institutional Review Board. All 135 patients eligible for this study had had the approval of these two review bodies.

The present study focused on validating the gene expression profiles that are related to LRR following mastectomy. Patients are eligible if they satisfied the inclusion criteria set in our 2006 study: i.e. 1) mastectomy as first treatment, 2) frozen fresh tissue available, 3) pathology stages I-III disease, 4) any LRR, 5) no PMRT, 6) minimal follow up of 2 years, and 7) an informed consent (Fig. 1). One hundred thirty five (135) patients were enrolled in this study.

We hypothesize that if the 18-gene classifier were capable of identifying locally aggressiveness of tumor biology and predicting LRR for mastectomy patients, it would also be effective to predict LRR after BCS even though they had adjuvant radiotherapy. We then used the breast-conserving surgery (BCS) patients who were treated in the same study period (n = 87) as a confirmatory cohort.

2.2. Samples and Microarray Analysis

A total of 135 frozen tissue samples came from surgical specimens of the primary tumors taken from patients prior to any treatment between 2005 and 2012. Tumor RNA was extracted from primary tumor tissues with Trizol (Invitrogen, Carlsbad, CA) and purified with the RNeasy Mini Kit (Qiagen, Valencia, CA), and assessed for quality with an Agilent 2100 Bioanalyzer. According to the Affymetrix protocol, hybridization targets were prepared from the total RNA and hybridized to U133 Plus 2.0 arrays. The details of the study method have been reported previously by investigators of our institution (Kao et al., 2011).

2.3. Statistical Analysis to Identify Gene Expression Profiles of LRR

In this study, the microarray platform was shifted from Affymetrix U95 to U133 Plus 2.0 array (Fig. 1). There were 30 of 34 GOI identified in this new platform, where they were distributed in 84 probes. An

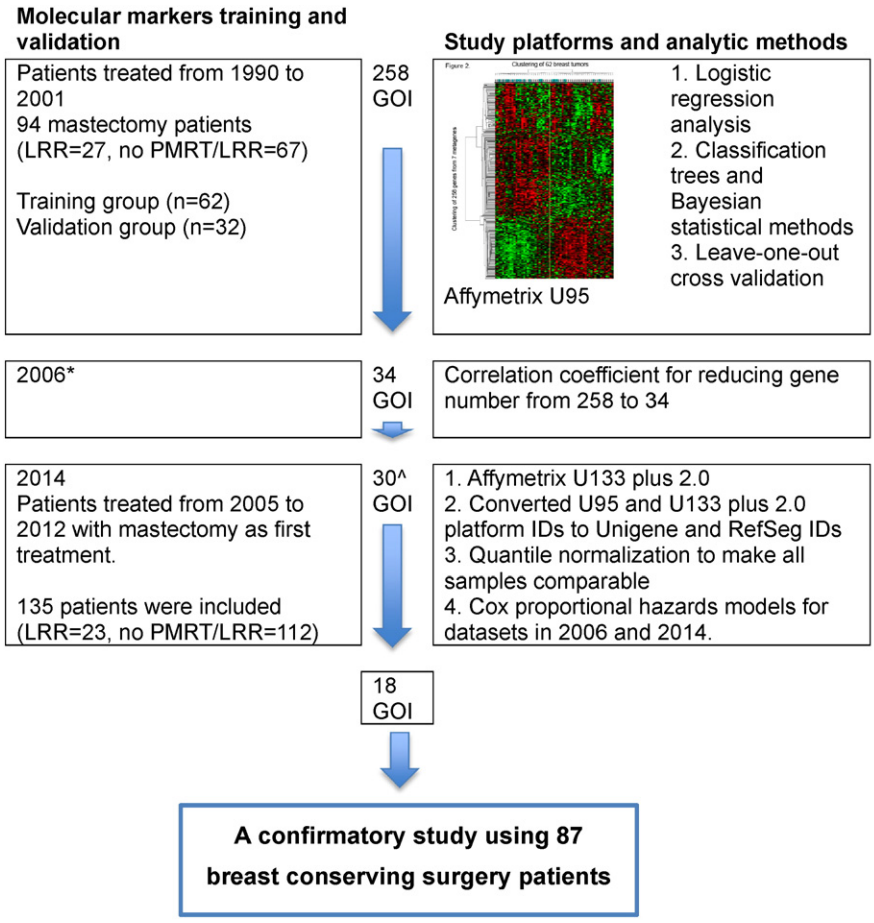

Fig. 1. Diagram to develop gene expression profiling that predicts locoregional recurrence in mastectomy patients. The genes of interest (GOI) reduced from 258 to 18 and maintained a similar, or greater ability to partition the low and high risk patients after mastectomy.

integration of inter-platform data was performed to convert individual platform IDs to Unigene and RefSeq IDs using identifier files provided by Affymetrix. A mean shift quantile normalization of 135 patients was done in the new platform to make sure all samples from both platforms were comparable (Fig. 1) (Sohal et al., 2008).

The statistical methods in our previous study were complicated: i.e. unsupervised clustering, logistic regression analysis, classification trees and Bayesian statistical methods, leave-one-out cross validation, and Pearson correlation coefficient. In brief, we used the concept of 'metagene' where each 'metagene' represented a key common pattern of expression of the genes in a cluster based on k-means clustering (Huang et al., 2003). Subsequently, we generated classification trees and used Bayesian statistical methods to explore multiple metagenes for optimal prediction. Finally 34 GOI were identified from 258 GOI using Pearson correlation coefficient (<0.3 or >0.3) (Cheng et al., 2006b).

In the current study, we selected the top 18 of 34 genes with highest correlation to partition mastectomy patients into the high and low risk groups and perform a cross validation of this 18-gene panel in our 2006 dataset (Table 2). We then optimize the 18-gene expression profiling in prediction of LRR and generate the 18-gene scoring algorithm according to the multivariate analysis (Fig. 2). The scoring algorithm is as follows:

$$\text{18-gene scores} = 4 \times \text{TRPV6} + 3 \times \text{DDX39} + 8 \times \text{BUB1B} + \text{CCR1} + \text{STIL} + 3 \times \text{BLM} + 11 \times \text{C16ORF7} + 4 \times \text{PIM1} + \text{TPX2} + 2 \times \text{PTI1} + 2 \times \text{TCF3} + \text{CCNB1} + \text{DTX2} + 2 \times \text{ENSA} + 5 \times \text{RCHY1} + 4 \times \text{NFATC2IP} + \text{OBSL1} + 2 \times \text{MMP15}.$$

2.4. Statistical Considerations

Cox proportional hazards regression models were used to assess the prognostic significance of the following risk factors: age at diagnosis, primary tumor size, number of involved axillary lymph nodes, nuclear grade, LVI, ER status, and the 18-gene score. Duration of locoregional control was calculated from the first day of treatment until the day of chest wall or regional nodal recurrence, or the last follow-up. LRR and distant metastasis that occurred simultaneously were counted as both. LRR that occurred after distant metastasis was censored. The LRR-free survival rates were calculated according to the Kaplan and Meier method (Kaplan and Meier, 1958). The log-rank test and chi-squared test were used to assess the statistical significance of the differences in LRR between patient subsets.

3. Results

The median follow-up of 135 patients was 65.1 months (ranging from 4.3–149.7 months). The details of patients' characteristics and treatment information are shown in Table 1. The majority of patients in this study were in the T2 or T3 disease groups (56%, n = 75), 68% (92) in N0 disease, and 77% (104) over the age of 40. Adjuvant hormonal therapy was given to 60% (81/135) of the patients, and chemotherapy to 79% (106) of the patients. There were 11 patients with LRR, 15 patients with distant recurrence and 10 patients with both locoregional and distant recurrences. Factors associated with any recurrence were the primary tumor stage and nodal involvement (Table 1). The study samples compared with the dataset in 2006 were earlier stage and less common of prominent LVI (Supplement Table 1).

Using the Cox regression model, 30 GOI with the new reference distribution of gene-expression levels, it was possible to distinguish mastectomy patients into the LRR low/high risk groups for the data set published in 2006 (Table 2) (Cheng et al., 2006b). This 30-gene panel could also distinguish the current data set (n = 135) into the low/high risk groups if each gene were given equal weighting for the scoring algorithm (Table 2). Patients classified as high risk had an LRR rate of 53%, whereas the low risk patients had an LRR rate of 3%. We used the top 18 genes with highest correlation to optimize our final prediction model. Again, if we assigned each gene with equal weighting of scoring, the 18 genes were capable of partitioning patients into the low and high risk groups among samples between 2006 and 2014 (Table 2, $p < 0.0001$).

The significance of 18 GOI related to LRR in the study patients using univariate analysis was shown in Supplement Table 2. Among them, TRPV6, C16ORF7, and RCHY1 were independent genes by multivariate analysis. The hazard ratios of 18 genes by multivariate analysis in the Cox model were composed of the prediction score (18-gene score).

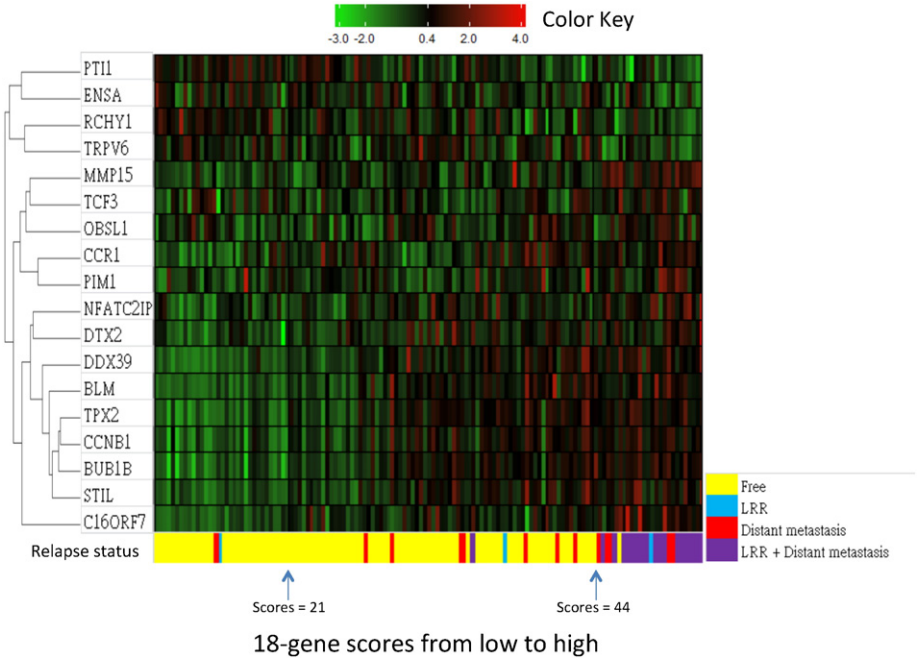

Fig. 2. An unsupervised cluster analysis of 18 genes and supervised clustering in 135 patients according to the 18-gene scores revealed distinct gene expression profiles in patients with and without recurrence. Patients with locoregional recurrence (LRR) are colored as blue (▮), both LRR and distant metastasis are colored as purple (▮), distant metastasis patients are colored as red (▮), and disease free patients are colored as yellow (▮) in the bottom of the heatmap.

Table 1
Patient characteristics and treatment information (N = 135).

Characteristics	No relapse (N = 99)	Any relapse[a] (N = 36)	p-Value
Median age (range)	50 (25–83)	49 (26–80)	0.0699
Age			
≤40	19 (61.3)	12 (38.7)	0.2183
41–50	31 (75.6)	10 (24.4)	
>50	49 (77.8)	14 (22.2)	
T stage			
T1	52 (86.7)	8 (13.3)	0.0017
T2–T3	47 (62.7)	28 (37.3)	
Axillary node positive			
N0	80 (87.0)	12 (13.0)	<0.0001
N1	18 (51.4)	17 (48.6)	
≥N2	1 (12.5)	7 (87.5)	
Nuclear grade			
I/II	53 (79.1)	14 (20.9)	0.1323
III	46 (67.7)	22 (32.4)	
Lymphovascular invasion			
Absent	45 (79.0)	12 (21.1)	0.1197
Focal	39 (75.0)	13 (25.0)	
Prominent	15 (57.7)	11 (42.3)	
Estrogen-receptor status			
Negative	32 (66.7)	16 (33.3)	0.1932
Positive	67 (77.0)	20 (23.0)	
Progesterone-receptor status			
Negative	50 (69.4)	22 (30.6)	0.2747
Positive	49 (77.8)	14 (22.2)	
HER2 overexpression			
Negative	63 (71.6)	25 (28.4)	0.5310
Positive	36 (77.0)	11 (23.0)	
Adjuvant hormonal therapy			
No	36 (66.7)	18 (33.3)	0.1527
Yes	63 (77.8)	18 (22.2)	
Adjuvant chemotherapy			
No	22 (75.9)	7 (24.1)	0.7282
Yes	77 (72.6)	29 (27.4)	

[a] 11 patients with locoregional recurrence, 15 patients with distant recurrence and 10 patients with both locoregional and distant recurrences.

Unsupervised clustering of 18 genes and later supervised clustering of 135 patients according to the 18-gene scores revealed gradually evolution of gene expression profiling and distinct gene expression patterns among patients with or without recurrence (locoregional or distant). There is existence of a gray zone in this heatmap (Fig. 2).

Clinical decisions are intended to be philosophically more conservative and tend toward over-treating patients. On that basis, the optimal cutoff score was 44 on the Receiver Operating Characteristic (ROC) curve. Patients with 18-gene scores of ≧44 are defined as high risk, and scores of <44 are defined as the low risk group. The overall accuracy of these predictions is 93%, with an estimated sensitivity of 87% and a specificity of 94% (Fig. 3a). The 5-year LRR-free survival rate in patients with scores of ≥44 and <44 is 30% and 99% ($p < 0.0001$), respectively (Fig. 3b). The time to recurrence (locoregional or distant) by the 18-gene score for all 135 patients is plotted in Fig. 3c. Patients with recurrences usually had scores of more than 40; patients with scores of <21 usually were disease-free (only one patient developed distant metastasis).

Table 2
Cross validation of the datasets between 2006 and 2014: The 30 and 18 GOI partitioning mastectomy patients into locoregional recurrence (LRR). Definition of low- and high- risk groups: each gene gives equal weighting for scoring algorithm.

Dataset	Risk group (scores)	Patient number	LRR# (%)	p Value
30 GOI				
2006 (N = 94)	High risk (≥21)	33	25(75.8)	<0.0001
	Low risk (<21)	61	2(3.3)	
2014 (N = 135)	High risk (≥21)	38	20(52.6)	<0.0001
	Low risk (<21)	97	3(3.1)	
18 GOI				
2006 (N = 94)	High risk (≥14)	23	16(69.6)	<0.0001
	Low risk (<14)	71	11(15.5)	
2014 (N = 135)	High risk (≥14)	43	20(46.5)	<0.0001
(Cross validation)	Low risk (<14)	92	3(3.3)	

3.1. Partitioning Patients According to Nodal Status and Breast Cancer Subtype

According to the lymph node status, the 5-year LRR-free survival rates in N0 and N1 patients with the scores of ≥44 and <44 were statistically different (51% versus 100%, and 27% versus 100%, $p < 0.0001$). N2 patients were too small to draw a conclusion (Table 3). Patients, who were defined by the 18-gene classifier as high risk, also had poor rates of 5-year distant metastasis-free survival, and overall survival, regardless of whether they were node-negative or node-positive. Node negative patients with scores of <44 were low risk of distant metastasis; the 5-year metastasis-free survival rate was excellent, at 95%. In contrast, the high risk N0 patients had a poor 5-year metastasis-free survival rate of 22% and overall survival rate of 44%. The detailed information is summarized in Table 3.

As for the breast cancer subtype, the 18-gene classifier also demonstrated a similar capability of predicting LRR and distant metastasis regardless of the breast cancer subtype (Table 3). Luminal-like (hormonal receptor positive and HER2 negative) subtype predicted as the high risk group had the best LRR-free survival rate of 50% when compared with those rates of HER2 subtype (0%) and triple negative breast cancer (14%).

3.2. Cox Proportional Hazards Model in Mastectomy Patients

We also examined whether the 18-gene classifier is an independent prognostic factor that is related to LRR. Multivariate analyses revealed that the prognostic relevance of clinical factors is the extent of lymph node metastasis and ER status with respect to LRR (Table 3). We re-analyzed the 135 patient samples including these clinical factors together with the 18-gene scores. This analysis confirms the significance of the 18-gene classifier as an independent factor associated with LRR. With the incorporation of the 18-gene score in the proportional hazards analyses, the hazard ratio for LRR is 31.1 (95% confidence interval, 8.3–115.9) in patients with 18-gene score ≥ 44 (Table 4).

3.3. Performance of the 18-Gene Classifier in BCS Patients

We hypothesized that the gene expression profiling derived from mastectomy patients would also be effective for prediction of LRR in BCS patients. Eighty-seven (87) BCS patients with microarray information were available to prove this hypothesis; these patients (82/87, 94%) have had post-operative radiotherapy (Table 5). Clinical characteristics of BCS patients are shown in Table 5.

Univariate analysis revealed that BCS patients with the 18-gene scores of ≥44 (HR 5.8, 95% CI 1.3–26.0) and prominent LVI (HR 4.3, 95% CI 1.0–19.3) were at high risk for LRR. Multivariate analysis by stepwise selection identified 18-gene classifier was the only risk factor to predict LRR in BCS and radiotherapy patients (Table 5).

4. Discussion

We have demonstrated that the 18-gene classifier is capable of partitioning mastectomy patients into more homogeneous risk groups of LRR with a biologic difference (Table 2). The 18-gene classifier also independently predicts LRR regardless of lymph node status, breast cancer subtype and surgical type (Tables 2 and 5). For N0 and N1

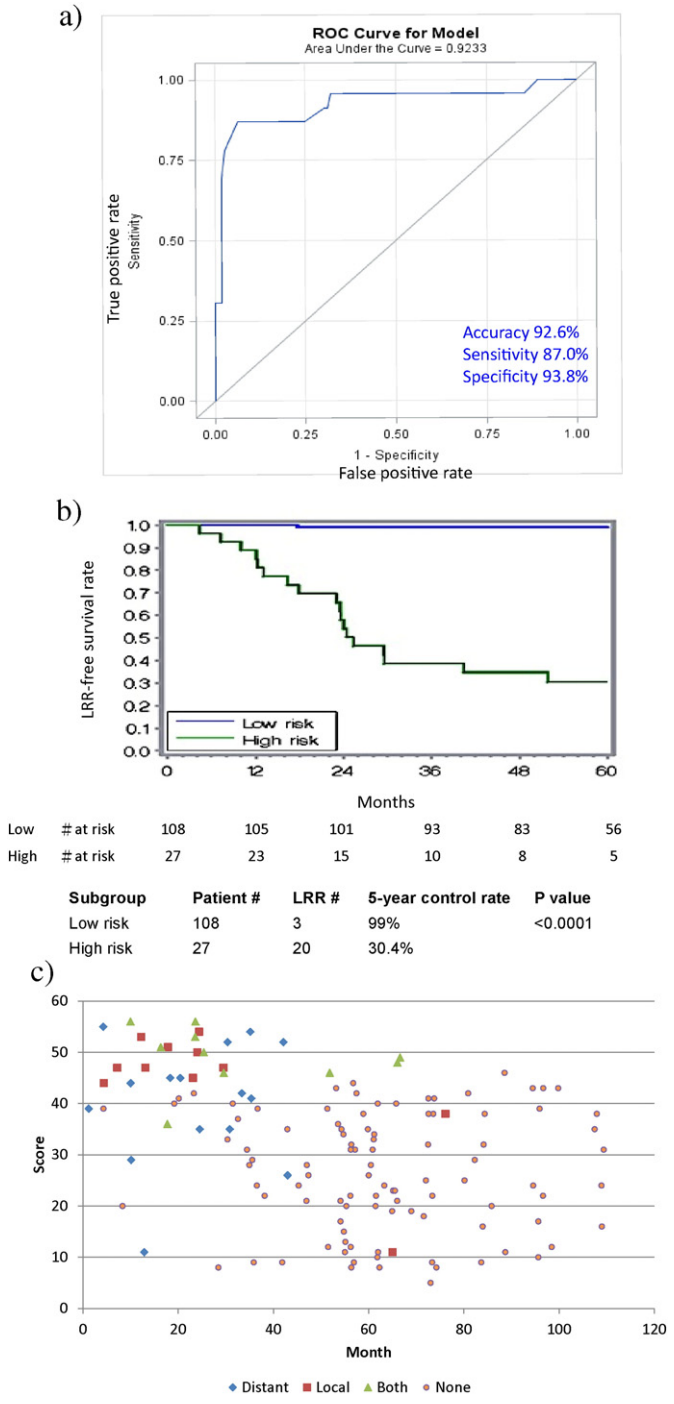

Fig. 3. (a) The receiver operating characteristic (ROC) curve constructed from 18-gene scoring algorithm as a predictor of locoregional recurrence in patients after mastectomy. The area under the receiver ROC curve: 0.752 ± 0.040 (95% CI: 0.672–0.831) ($p < 0.001$). (b) The 5-year locoregional recurrence (LRR) free survival rates in patients with an 18-gene score of <44 (low risk group) and ≥44 (high risk group). Among the low risk group, two LRRs occurred after 5 years of follow-up (see Fig. 3c). (c) The relation between the 18-gene score and time to recurrence. The X axis is the follow-up interval (month). The Y axis is the 18-gene score. The orange dot (●) represents disease free; the light green triangle (▲) represents LRR and distant recurrences simultaneously; the red square (■) represents LRR; and the light blue rhombus (◆) represents distant metastasis.

Table 3
The 18-gene classifier partitioning patients into different risk subgroups according to lymph node status and breast cancer subtype.

18-gene score	Patient #	Five-year LRR-free survival rate	Five-year metastasis-free survival rate	Five-year overall survival rate
N0 patients				
Low risk	83	100.0%	95.1%	95.6%
High risk	9	50.8%	22.2%	44.4%
P value		<0.0001	<0.0001	<0.0001
N1 patients				
Low risk	24	95.2%	76.6%	77.6%
High risk	11	27.3%	22.7%	26.7%
P value		<0.0001	0.0014	0.0272
≥N2 patients	8	Too small to be analyzed		
Luminal-like subtype				
Low risk	55	100%	90.4%	90%
High risk	12	50%	31.3%	57.1%
P value		<0.0001	<0.0001	<0.0001
HER2 subtype				
Low risk	38	97.4%	94.7%	97.4%
High risk	8	0%	0%	14.6%
P value		<0.0001	<0.0001	<0.0001
Triple negative subtype				
Low risk	13	100%	92.3%	84.6%
High risk	7	14.3%	14.3%	14.3%
P value		<0.0001	0.0007	0.0050

mastectomy patients with the 18-gene scores of <44, the 5-year LRR-free survival rates are 95–100%; those with scores of 44 or greater, the LRR-free survival rates drop to only 27–51% (Table 3). This observation gives new insight that our gene panel may help patients and clinicians to make better decisions about PMRT.

Gene expression profiles that predict distant metastases in breast cancer have been reported since 2002 (Glinsky et al., 2004; van 't Veer et al., 2002; van de Vijver et al., 2002). The current study gives another example that gene expression profiling is capable of estimating the risk of LRR. This is crucial in clinical practice because high risk LRR requires post-operative radiotherapy not only to reduce local recurrence, but also prevent distant metastasis (EBCTCG et al., 2014). The Danish Breast Cancer Cooperative Group (DBCG) reported a 7-gene profile, which could divide mastectomy patients into the "high LRR risk" and "low LRR risk" groups (Tramm et al., 2014). The LRR rate in the "low risk" group is about 7–8% and the "high risk" group ranges from 50–60%. Their LRR genes are totally different from ours. The LRR rate in their "low risk" patients is higher than ours (our data shows <2% at 5 years, Fig. 3B). This may be related to different treatment years and ethnicity (Curtis et al., 2008). Others have reported a 12-gene expression assay, which is useful to predict LRR in ductal carcinoma in situ patients after a wide excision (Solin et al., 2013). Our work here observes similar

Table 4
Multivariate analysis by Cox proportional hazards model for locoregional recurrence in all mastectomy patients (N = 135).

Variable		Hazards ratio (95% confidence interval)	p Value
Analysis without 18-gene classifier			
ER status	Positive	1.0	
	Negative	3.6 (1.4, 9.4)	0.0089
N stage	N0	1.0	
	N1	14.2 (4.4, 45.3)	<0.0001
	N2	31.2 (8.8, 110.3)	<0.0001
Analysis with 18-gene classifier			
ER status	Positive	1.0	
	Negative	2.4 (1.0, 5.8)	0.0597
N stage	N0	1.0	
	N1	5.0 (1.5, 16.4)	0.0088
	N2	4.3 (1.1, 16.8)	0.0345
18-gene score	<44	1.0	
	≥44	31.1 (8.3, 115.9)	<0.0001

Table 5
Univariate and multivariate analysis of locoregional recurrence (LRR) for breast-conserving surgery patients (N = 87).

Risk factor	Patient#	LRR#	LRR rate	Univariate analysis (95% confidence interval)	Multivariate analysis (95% confidence interval)[a]
Age					
≤40	19	2	10.5	1.4 (0.3–7.1)	
>40	68	5	7.4	1.0	
T stage					
T1	54	2	3.7	1.0	
T2	33	5	15.2	4.3 (0.8–22.2)	
N stage					
N0	56	4	7.1	1.0	
N1	21	2	9.5	1.3 (0.2–7.2)	
≥N2	10	1	10.0	1.6 (0.2–14.7)	
Nuclear grade					
III	44	4	9.1	1.4 (0.3–6.4)	
I/II	43	3	7.0	1.0	
Lymphovascular invasion					
Prominent	14	3	21.4	**4.3 (1.0–19.3)***	
Absent/focal	73	4	5.5	1.0	
ER status					
Negative	25	3	12.0	2.0 (0.5–9.1)	
Positive	62	4	6.5	1.0	
PR status					
Negative	32	3	9.4	1.3 (0.3–5.9)	
Positive	55	4	7.3	1.0	
HER2 overexpression					
Positive	25	2	8.0	1.2 (0.2–6.2)	
Negative	62	5	8.1	1.0	
Adjuvant H/T					
No	26	4	15.4	3.5 (0.8–15.7)	
Yes	61	3	4.9	1.0	
Adjuvant C/T					
No	11	0	0.0	NA	
Yes	76	7	9.2		
18-gene score					
≥44	18	4	22.2	**5.8 (1.3–26.0)***	**5.8 (1.3–26.0)***
<44	69	3	4.4	1.0	1.0

Bold values represents p value < 0.05.
* $p < 0.05$.
[a] Stepwise selection for multivariate Cox model.

LRR predictive ability by the 18-gene classifier in invasive cancer, which can distinguish the LRR low risk group from the LRR high risk after breast cancer surgery.

The 18 genes are associated with the oncogenic process, proliferation invasion, inflammation, cell–cell interaction, apoptosis and metabolism (Ellis et al., 1995; Irie et al., 1998; Jin et al., 1997; Kamps et al., 1990; Nomura et al., 1993; Pines and Hunter, 1989; Semba et al., 1986). The expression of one or more of BLM, TCF3, PIM1, DDX39, BUB1B, STIL, TPX2, CCNB1, MMP15, CCR1, NFATC2IP, OBSL1, C16ORF7, and DTX2 indicates an increased likelihood of breast cancer LRR. On the contrary, the expression of one or more of RCHY1, PTI1, ENSA, and TRPV6 indicates a decreased likelihood of breast cancer LRR (Fig. 2). Compare to the Oncotype Dx, our gene set has one overlapping gene (CCNB1) and another gene in the same family (MMP11 Oncotype Dx and MMP15 for ours) (Soonmyung Paik et al., 2004). CCNB1 is also an important gene in PAM50 gene set (Dowsett et al., 2013). The details of 18-gene function are listed in Table S4.

The 18-gene signatures are mainly for N0 and N1 mastectomy patients. In literature, the risk of LRR in N0 patients is about 2–8%, whereas our gene classifier identifies about 10% of N0 patients to be high risk (Clarke et al., 2005; EBCTCG et al., 2014). Among them, 49% had LRR and 78% developed distant metastases (Table 2). This is extremely important because by current practice guidelines these patients would not be given PMRT; in fact, they are high risk of both locoregional and distant recurrences. Similarly, the risk of LRR in N1 patients is about 20% in literature, whereas our gene classifier identifies 31% of N1 patients to be at high risk, whose risk of LRR is 73% (Table 2) (EBCTCG et al., 2014). Although our patient number is relatively small, it provides good opportunity for better cancer care. As mentioned in our previous study, optimal sensitivity and specificity of the gene expression profiles are desirable in order to avoid having "truly" high risk patients undergo suboptimal treatment and "truly" low risk patients undergo over-treatment (Cheng et al., 2006b). Achieving such a goal appears possible according to the current study (Fig. 3).

Decisions regarding whether to assign N1 mastectomy patients to adjuvant radiotherapy, which are made based on clinical parameters, results in the over-treatment of 80% of patients (EBCTCG et al., 2014). The prevalence of PMRT for N1 patients has increased gradually from 32% in 2007 to 46% in 2012 by SEER registry data since the recommendation of NCCN guidelines (Frasier et al., 2015). The lack of biological markers may partially explain the low guideline adherence rate. However, PMRT applied to patients reduces not only LRR, but also improves their overall survival odds by preventing distant metastases (Poortmans et al., 2015; Ragaz et al., 2005). It is essential to identify truly "high risk" patients for the prevention of LRR and distant metastasis. The present study reveals that N0 and N1 patients can be sorted into more homogeneous subgroups by the 18-gene classifier (Table 2). Both MA20 and EORTC 22922 studies have shown marginal effects on overall survival by regional node irradiation (Poortmans et al., 2015; Whelan et al., 2015). The 18-gene panel is potentially useful in identification of the truly "high risk" patients who would benefit most from PMRT/regional node irradiation, and it would omit radiotherapy in the low risk patients.

The 18-gene classifier is in many ways similar to the Oncotype Dx 21-gene panel, except in supporting the decision of adjuvant radiotherapy. The cost-effectiveness analysis suggests that the 21-gene panel is cost saving in comparison with conventional decision-making processes (Holt et al., 2013; Kondo et al., 2011; Vataire et al., 2012). It is possible that our test will provide an opportunity to optimize treatment prescription by avoiding unnecessary radiotherapy and by prescribing radiotherapy to women who would not have received it based on the standard decision criteria (Table 3).

PMRT is given to the patient one fraction per day and 5 fractions per week. It usually takes 5 to 6 weeks to complete a course of treatment. The treatment places heavy burdens on not just the patients, but also their families and the society; inconveniences that include the daily commutes for treatment, absence from their workplaces, the rearrangements of manpower due to the employee having to take leave, additional financial expenses… etc. Our test may facilitate the right treatment for the right patient and would consequently alleviate the burdens on the patients, their families, and society.

The 18-gene classifier could distinguish the truly "high risk" breast cancer patients for post-operative radiotherapy. It contributes to the current call for precision medicine in oncology. Our gene expression profiling is an example of such effort for radiation oncologists to move toward greater precision in their practice of medicine. Our 18-gene classifier is useful for determining whether radiotherapy should or should not be considered for N0-N1 patients with breast cancer (Table 4).

Our present study is an initial exploration for genomic risk factors and therefore did not reach the level of randomization. The potential weakness lies in the fact that the proportion of N1 patients in this 135-patient series is fewer due to the tendency for N1 patients to be given PMRT between 2005 and now due to the encouragement by the EORTCG to irradiate N1 patients (Clarke et al., 2005). Our series thus contains larger N0 population and may not have included large enough N1 population.

Also, this study lacks certain parameters required of a validation study for biomarkers. In principle, the requirement should include: (1) Technical validity: the measurement is reproducible; (2) Clinical

validity: the assay identifies a biologic difference; and (3) Clinical utility: the assay leads to a new clinical decision (Novelli et al., 2008; Teutsch et al., 2009). According to this definition, this study has confirmed that the multigene panel developed since 2006 could identify biologically different subgroups and lead to a better clinical decision about post-mastectomy radiotherapy. This study has not totally met the first criteria because of different study platforms (Affymetrix U95 versus Affymetrix U133 Plus 2.0). A validation study using the dataset from the phase III study, such as MA.20 with pre-defined criteria and assay is necessary to confirm the true value of 18-gene classifier in the decision of adjuvant radiotherapy.

In summary, the 18-gene classifier independently estimates the likelihood of LRR in breast cancer patients after mastectomy. As we have mentioned the 18-gene classifier is capable of partitioning N0 and N1 patients into more homogeneous subgroups. A larger validation study of these patients by using formalin-fixed paraffin embedded tissues or fresh frozen tissues is underway to confirm the broader value of this genomic predictor and its value in improving health care via more individualized prediction of treatment outcomes.

Competing Interests

The authors are currently applying for a patent relating to the content of this manuscript (Patent No. 104115832 (Taiwan)). None of the authors have any conflicts of interests in this research, either financial or non-financial.

Authors' Contributions

SHC conceived of, and designed the study, participated in the molecular genetic analysis, interpretation of data, and drafted the manuscript. **CFH and TTH** acquired the data, carried out statistical analysis and drafted the work. **ESH** participated in the initial conception of this research and gave important intellectual comments on the draft. **MHT and LSS** participated in the tumor tissue preparation, specimen quality control, and helped revise the manuscript. **BLY and CMC** made substantial contributions to acquisition and interpretation of data, and revised and provided important comments on the draft. **ATH** coordinated this study and revised the manuscript critically for important intellectual content. All authors read and approved the final manuscript.

Funding Support

Grant number: NSC 102-3114-Y-368-001, the Ministry of Science and Technology, Taiwan

Protocol Approval

The Bio-bank Ethical Committee and the Institutional Review Board of the Koo Foundation Sun Yat-Sen Cancer Center have approved this study (Approved number: 20131001A).

Acknowledgments

The authors would like to give special thanks to Mike West, who is “The Art & Science Professor of Statistical & Decision Sciences” in the Department of Statistical Science at Duke University for his valuable comments about inter-platform comparison. We also wish to thank all the members of the multidisciplinary breast cancer team at the Koo Foundation Sun Yat-Sen Cancer Center (KF-SYSCC): (1) Patient care and sample collections: Drs. Tsung-Yen Cheng and Tzu-Jung, Tsai in the Department of Surgery; Drs. Wan-Chen, Tsai and Christopher KJ Lin in the Department of Radiology; Drs. James J. Jian and Yu-Chen, Tsai in the Department of Radiation Oncology; (2) Data management and quality control: Yen-Chun Lin; (3) Tumor Bank: Theresa M. Tsai; and (4) Microarray Laboratory: Lih-Chian Wu.

We thank Mr. Rick McDonald who has edited this manuscript. The following is his information: Rick McDonald: B.A., M.A. Literature and Writing, University of Northern Iowa; International Teacher of the International Baccalaureate, Advanced Placement, and Higher Level Literature and Writing Courses — Retired; Department Chair of Secondary English Departments at the International School Bangkok, International School of Kenya, and Tripoli Community School; Director of Writing at the American School of Riyadh in Saudi Arabia; Director of Forensics and Debate at the International School Bangkok.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ebiom.2016.02.022.

References

Cheng, S.H., Horng, C.F., Clarke, J.L., Tsou, M.H., Tsai, S.Y., Chen, C.M., ... Prosnitz, L.R., 2006a. Prognostic index score and clinical prediction model of local regional recurrence after mastectomy in breast cancer patients. Int. J. Radiat. Oncol. Biol. Phys. 64 (5), 1401–1409 http://dx.doi.org/10.1016/j.ijrobp.2005.11.015 (doi:S0360-3016(05)02968-8 [pii]).

Cheng, S.H., Horng, C.F., West, M., Huang, E., Pittman, J., Tsou, M.H., ... Huang, A.T., 2006b. Genomic prediction of locoregional recurrence after mastectomy in breast cancer. J. Clin. Oncol. 24 (28), 4594–4602 http://dx.doi.org/10.1200/JCO.2005.02.5676 (doi: 24/28/4594 [pii]).

Clarke, M., Collins, R., Darby, S., Davies, C., Elphinstone, P., Evans, E., Early Breast Cancer Trialists' Collaborative, G., 2005. Effects of radiotherapy and of differences in the extent of surgery for early breast cancer on local recurrence and 15-year survival: an overview of the randomised trials. Lancet 366 (9503), 2087–2106 http://dx.doi.org/10.1016/S0140-6736(05)67887-7.

Curtis, E., Quale, C., Haggstrom, D., Smith-Bindman, R., 2008. Racial and ethnic differences in breast cancer survival: how much is explained by screening, tumor severity, biology, treatment, comorbidities, and demographics? Cancer 112 (1), 171–180 http://dx.doi.org/10.1002/cncr.23131.

Dowsett, M., Sestak, I., Lopez-Knowles, E., Sidhu, K., Dunbier, A.K., Cowens, J.W., ... Cuzick, J., 2013. Comparison of PAM50 risk of recurrence score with Oncotype DX and IHC4 for predicting risk of distant recurrence after endocrine therapy. J. Clin. Oncol. 31 (22), 2783–2790 http://dx.doi.org/10.1200/JCO.2012.46.1558.

Early Breast Cancer Trialists' Collaborative, G., 2002. Radiotherapy for early breast cancer. Cochrane Database Syst. Rev. 2, CD003647 http://dx.doi.org/10.1002/14651858.CD003647.

EBCTCG, McGale, P., Taylor, C., Correa, C., Cutter, D., Duane, F., ... Darby, S., 2014. Effect of radiotherapy after mastectomy and axillary surgery on 10-year recurrence and 20-year breast cancer mortality: meta-analysis of individual patient data for 8135 women in 22 randomised trials. Lancet 383 (9935), 2127–2135 http://dx.doi.org/10.1016/S0140-6736(14)60488-8.

Ellis, N.A., Groden, J., Ye, T.Z., Straughen, J., Lennon, D.J., Ciocci, S., ... German, J., 1995. The Bloom's syndrome gene product is homologous to RecQ helicases. Cell 83 (4), 655–666 (Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=7585968).

Fan, C., Oh, D.S., Wessels, L., Weigelt, B., Nuyten, D.S., Nobel, A.B., ... Perou, C.M., 2006. Concordance among gene-expression-based predictors for breast cancer. N. Engl. J. Med. 355 (6), 560–569 http://dx.doi.org/10.1056/NEJMoa052933.

Frasier, L.L., Holden, S., Holden, T., Schumacher, J.R., Leverson, G., Anderson, B., ... Neuman, H.B., 2015. Temporal trends in postmastectomy radiation therapy and breast reconstruction associated with changes in National Comprehensive Cancer Network Guidelines. JAMA Oncol. 1-7 http://dx.doi.org/10.1001/jamaoncol.2015.3717.

Glinsky, G.V., Higashiyama, T., Glinskii, A.B., 2004. Classification of human breast cancer using gene expression profiling as a component of the survival predictor algorithm. Clin. Cancer Res. 10 (7), 2272–2283 (Retrieved from file://localhost/Users/skyehcheng/Documents/My%20Library%20Data/GeneExprCancer-Converted.Data/2272.pdf).

Holt, S., Bertelli, G., Humphreys, I., Valentine, W., Durrani, S., Pudney, D., ... Phillips, C.J., 2013. A decision impact, decision conflict and economic assessment of routine Oncotype DX testing of 146 women with node-negative or pNlmi, ER-positive breast cancer in the U.K. Br. J. Cancer 108 (11), 2250–2258 http://dx.doi.org/10.1038/bjc.2013.207.

Huang, E., West, M., Nevins, J.R., 2003. Gene expression profiling for prediction of clinical characteristics of breast cancer. Recent Prog. Horm. Res. 58, 55–73 (Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12795414).

Irie, A., Koyama, S., Kozutsumi, Y., Kawasaki, T., Suzuki, A., 1998. The molecular basis for the absence of N-glycolylneuraminic acid in humans. J. Biol. Chem. 273 (25), 15866–15871 (Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9624188).

Jin, D.Y., Teramoto, H., Giam, C.Z., Chun, R.F., Gutkind, J.S., Jeang, K.T., 1997. A human suppressor of c-Jun N-terminal kinase 1 activation by tumor necrosis factor alpha. J. Biol. Chem. 272 (41), 25816–25823 (Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9325311).

Kamps, M.P., Murre, C., Sun, X.H., Baltimore, D., 1990. A new homeobox gene contributes the DNA binding domain of the t(1;19) translocation protein in pre-B ALL. Cell 60 (4), 547–555 (Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=1967983).

Kao, K.J., Chang, K.M., Hsu, H.C., Huang, A.T., 2011. Correlation of microarray-based breast cancer molecular subtypes and clinical outcomes: implications for treatment optimization. BMC Cancer 11, 143 http://dx.doi.org/10.1186/1471-2407-11-143 (1471-2407-11-143 [pii]).

Kaplan, E.L., Meier, P., 1958. Nonparametric estimation from incomplete observation. J. Am. Stat. Assoc. 53, 457–481.

Kondo, M., Hoshi, S.L., Yamanaka, T., Ishiguro, H., Toi, M., 2011. Economic evaluation of the 21-gene signature (Oncotype DX) in lymph node-negative/positive, hormone receptor-positive early-stage breast cancer based on Japanese validation study (JBCRG-TR03). Breast Cancer Res. Treat. 127 (3), 739–749 http://dx.doi.org/10.1007/s10549-010-1243-y.

Nomura, H., Nielsen, B.W., Matsushima, K., 1993. Molecular cloning of cDNAs encoding a LD78 receptor and putative leukocyte chemotactic peptide receptors. Int. Immunol. 5 (10), 1239–1249 (Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=7505609).

Novelli, G., Ciccacci, C., Borgiani, P., Papaluca Amati, M., Abadie, E., 2008. Genetic tests and genomic biomarkers: regulation, qualification and validation. Clin. Cases Miner. Bone Metab. 5 (2), 149–154 (Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/22460999).

Paik, S., Shak, S., Tang, G., Kim, C., Baker, J., Cronin, M., ... Park, T., 2004a. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N. Engl. J. Med. 351 (27), 2817–2826 (Retrieved from http://www.nejm.org/doi/pdf/10.1056/NEJMoa041588).

Paik, S., Shak, S., Tang, G., Kim, C., Baker, J., Cronin, M., ... Wolmark, N., 2004b. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N. Engl. J. Med. 351 (27), 2817–2826 http://dx.doi.org/10.1056/NEJMoa041588.

Paik, S., Tang, G., Shak, S., Kim, C., Baker, J., Kim, W., ... Wolmark, N., 2006. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. J. Clin. Oncol. 24 (23), 3726–3734 http://dx.doi.org/10.1200/JCO.2005.04.7985.

Pines, J., Hunter, T., 1989. Isolation of a human cyclin cDNA: evidence for cyclin mRNA and protein regulation in the cell cycle and for interaction with p34cdc2. Cell 58 (5), 833–846 (Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=2570636).

Poortmans, P.M., Collette, S., Kirkove, C., Van Limbergen, E., Budach, V., Struikmans, H., ... Breast Cancer, G., 2015. Internal mammary and medial supraclavicular irradiation in breast cancer. N. Engl. J. Med. 373 (4), 317–327 http://dx.doi.org/10.1056/NEJMoa1415369.

Ragaz, J., Olivotto, I.A., Spinelli, J.J., Phillips, N., Jackson, S.M., Wilson, K.S., ... Manji, M., 2005. Locoregional radiation therapy in patients with high-risk breast cancer receiving adjuvant chemotherapy: 20-year results of the British Columbia randomized trial. J. Natl. Cancer Inst. 97 (2), 116–126 (Retrieved from file://localhost/Users/skyehcheng/Documents/My%20Library%20Data/PMRT-Converted.Data/Ragaz%20jnci.pdf).

Recht, A., Edge, S.B., Solin, L.J., Robinson, D.S., Estabrook, A., Fine, R.E., ... Pfister, D.G., 2001. Postmastectomy radiotherapy: clinical practice guidelines of the American Society of Clinical Oncology. J. Clin. Oncol. 19 (5), 1539–1569 (Retrieved from file://localhost/Users/skyehcheng/Documents/My%20Library%20Data/PMRT-Converted.Data/1539.pdf).

Semba, K., Nishizawa, M., Miyajima, N., Yoshida, M.C., Sukegawa, J., Yamanashi, Y., ... Toyoshima, K., 1986. Yes-related protooncogene, syn, belongs to the protein-tyrosine kinase family. Proc. Natl. Acad. Sci. U. S. A. 83 (15), 5459–5463 (Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=3526330).

Sohal, D., Yeatts, A., Ye, K., Pellagatti, A., Zhou, L., Pahanish, P., ... Verma, A., 2008. Meta-analysis of microarray studies reveals a novel hematopoietic progenitor cell signature and demonstrates feasibility of inter-platform data integration. PLoS One 3 (8), e2965 http://dx.doi.org/10.1371/journal.pone.0002965.

Solin, L.J., Gray, R., Baehner, F.L., Butler, S.M., Hughes, L.L., Yoshizawa, C., ... Badve, S., 2013. A multigene expression assay to predict local recurrence risk for ductal carcinoma in situ of the breast. J. Natl. Cancer Inst. 105 (10), 701–710 http://dx.doi.org/10.1093/jnci/djt067.

Sorlie, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., ... Borresen-Dale, A.L., 2001. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc. Natl. Acad. Sci. U. S. A. 98 (19), 10869–10874 http://dx.doi.org/10.1073/pnas.191367098 (98/19/10869 [pii]).

Taghian, A., Jeong, J.H., Mamounas, E., Anderson, S., Bryant, J., Deutsch, M., Wolmark, N., 2004. Patterns of locoregional failure in patients with operable breast cancer treated by mastectomy and adjuvant chemotherapy with or without tamoxifen and without radiotherapy: results from five National Surgical Adjuvant Breast and Bowel Project randomized clinical trials. J. Clin. Oncol. 22 (21), 4247–4254 (Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15452182).

Teutsch, S.M., Bradley, L.A., Palomaki, G.E., Haddow, J.E., Piper, M., Calonge, N., ... Group, E.W., 2009. The evaluation of genomic applications in practice and prevention (EGAPP) initiative: methods of the EGAPP working group. Genet. Med. 11 (1), 3–14 http://dx.doi.org/10.1097/GIM.0b013e318184137c.

Tramm, T., Mohammed, H., Myhre, S., Kyndi, M., Alsner, J., Borresen-Dale, A.L., ... Overgaard, J., 2014. Development and validation of a Gene Profile Predicting benefit of postmastectomy radiotherapy in patients with high-risk Breast cancer: a study of gene expression in the DBCG82bc cohort. Clin. Cancer Res. 20 (20), 5272–5280 http://dx.doi.org/10.1158/1078-0432.CCR-14-0458.

van 't Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A., Mao, M., Friend, S.H., 2002. Gene expression profiling predicts clinical outcome of breast cancer. Nature 415 (6871), 530–536 (Retrieved from file://localhost/Users/skyehcheng/Documents/My%20Library%20Data/GeneExprCancer-Converted.Data/%20GEP%20of%20Br%20Ca%20in%20Nature.pdf).

van de Vijver, M.J., He, Y.D., van't Veer, L.J., Dai, H., Hart, A.A., Voskuil, D.W., ... Bernards, R., 2002. A gene-expression signature as a predictor of survival in breast cancer. N. Engl. J. Med. 347 (25), 1999–2009 (Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12490681).

van der Hage, J.A., van den Broek, L.J., Legrand, C., Clahsen, P.C., Bosch, C.J., Robanus-Maandag, E.C., ... van de Vijver, M.J., 2004. Overexpression of P70 S6 kinase protein is associated with increased risk of locoregional recurrence in node-negative premenopausal early breast cancer patients. Br. J. Cancer 90 (8), 1543–1550 http://dx.doi.org/10.1038/sj.bjc.6601741.

Vataire, A.L., Laas, E., Aballea, S., Gligorov, J., Rouzier, R., Chereau, E., 2012. Cost-effectiveness of a chemotherapy predictive test. Bull. Cancer 99 (10), 907–914 http://dx.doi.org/10.1684/bdc.2012.1652.

Whelan, T.J., Olivotto, I.A., Parulekar, W.R., Ackerman, I., Chua, B.H., Nabid, A., ... Investigators, M.A.S., 2015. Regional nodal irradiation in early-stage breast cancer. N. Engl. J. Med. 373 (4), 307–316 http://dx.doi.org/10.1056/NEJMoa1415340.

Zellars, R.C., Hilsenbeck, S.G., Clark, G.M., Allred, D.C., Herman, T.S., Chamness, G.C., Elledge, R.M., 2000. Prognostic value of p53 for local failure in mastectomy-treated breast cancer patients. J. Clin. Oncol. 18 (9), 1906–1913 (Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10784631).