Biomarker Insights

ORIGINAL RESEARCH

Prognosis of Treatment Response (Pathological Complete Response) in Breast Cancer

Jason B. Nikas, Walter C. Low and Paul A. Burgio

Applied Informatic Solutions, St. Paul, MN, USA. Corresponding author email: jbnikas@ouraisolutions.com

Abstract: Pertaining to the female population in the USA, breast cancer is the leading cancer in terms of annual incidence rate and, in terms of mortality, the second most lethal cancer. There are currently no biomarkers available that can predict which breast cancer patients will respond to chemotherapy with both sensitivity and specificity $>$ 80%, as mandated by the latest FDA requirements. In this study, we have developed a prognostic biomarker model (complex mathematical function) that—based on global gene expression analysis of tumor tissue collected during biopsy and prior to the commencement of chemotherapy—can identify with a high accuracy those patients with breast cancer (clinical stages I–III) who will respond to the paclitaxel-fluorouracil-doxorubicin-cyclophosphamide chemotherapy and will experience pathological complete response (Responders), as well as those breast cancer patients (clinical stages I–III) who will not do so (Non-Responders). Most importantly, both the application and the accuracy of our breast cancer prognostic biomarker model are independent of the status of the hormone receptors ER, PR, and HER2, as well as of the ethnicity and age of the subjects. We developed our prognostic biomarker model with 50 subjects [10 responders (R) and 40 non-responders (NR)], and we validated it with 43 unknown (new and different) subjects [10 responders (R) and 33 non-responders (NR)]. All 93 subjects were recruited at five different clinical centers around the world. The overall sensitivity and specificity of our prognostic biomarker model were 90.0% and 91.8%, respectively. The nine most significant genes identified, which comprise the input variables to the mathematical function, are involved in regulation of transcription; cell proliferation, invasion, and migration; oncogenesis; suppression of immune response; and drug resistance and cancer recurrence.

Keywords: breast cancer, biomarkers, prognostic biomarker models, treatment response, global gene expression analysis, systems biology

Biomarker Insights 2012:7 59–70

doi: 10.4137/BMI.S9387

This article is available from http://www.la-press.com.

Introduction

In the USA, breast cancer is the leading cancer in terms of annual incidence rate (~207,090 new cases/year) and, in terms of mortality, the second most lethal cancer (~39,840 deaths/year).[1,2] Treatment may entail lumpectomy or mastectomy and removal of some of the axillary lymph nodes, and it may involve chemotherapy (with taxol or other chemotherapeutic agents), before or after surgery, hormone therapy, or radiation.[1] There are currently no biomarkers available that can predict which breast cancer patients will respond to chemotherapy with both sensitivity and specificity $> 80\%$, as mandated by the latest FDA requirements. It follows, therefore, that a prognostic test that could identify with a high accuracy those breast cancer patients who will respond to chemotherapy, as well as those patients who will not do so, would constitute significant progress against this disease. More specifically, such a prognostic test would be invaluable in: (1) providing the physicians with the ability to identify responders from non-responders at the outset (at the time of the biopsy and prior to the commencement of chemotherapy), (2) providing alternative therapies to non-responders of chemotherapy, and (3) helping pharmaceutical companies to test and develop new analogs of chemotherapeutic agents that may be more effective for the non-responders.

In this study, by analyzing the global gene expression data of the tumor tissue obtained during biopsy (and, therefore, prior to the administration of chemotherapy) from 93 patients with breast cancer (clinical stages I–III), we developed a prognostic biomarker model that was able to identify with a high accuracy (overall sensitivity: 90.0% and overall specificity: 91.8%) both the responders (R) and the non-responders (NR) to the paclitaxel and fluorouracil-doxorubicin-cyclophosphamide (T/FAC) chemotherapy, regardless of the status of the estrogen (ER), progesterone (PR), and HER2 hormone receptors of the subjects. We developed our prognostic biomarker model using 50 patients [10 responders (R) and 40 non-responders (NR)], and we validated it with 43 unknown patients [10 responders (R) and 33 non-responders (NR)] that were new and different from those 50 used in the development of the model.

Our prognostic biomarker model (F_1) is a complex mathematical function of nine genes. Those nine genes are therefore deemed highly significant in the process of the response to the T/FAC treatment on the part of breast cancer patients. Of those nine genes, one has been known to induce tumorigenesis by altering cell-cycle progression; two genes are involved in cell proliferation, invasion, and migration; one gene is involved in immune response; two genes are known to interact directly with BRCA1 and BRCA2; one gene is known to interact directly with c-MYC and another with HRAS; and two genes are involved in lipid metabolism.

Materials and Methods

Data acquisition and clinical sample information

We used the raw intensity microarray data (CEL files) for 20 subjects that responded to chemotherapy and for 73 subjects that did not respond to chemotherapy as posted by Tabchy et al[3] at the GEO (Gene Expression Omnibus) of the NCBI (National Center for Biotechnology Information) [ID: GSE20271].

Briefly, according to Tabchy et al,[3] patients with breast cancer (clinical stages I–III) were recruited in five different clinical centers around the world [M. D. Anderson Cancer Center, Houston, TX, USA; Lyndon B. Johnson General Hospital, Houston, TX, USA; Instituto Nacional de Enfermedades Neoplasicas, Lima, Peru; Centro Medico Nacionalde Occidente, Guadalajara, Mexico; and Grupo Espanol de Investigacion en Cancer de Mama, Spain]. All subjects first underwent biopsy, and tumor tissue obtained thus was analyzed for global gene expression using the GeneChip array U133A by Affymetrix. Histological diagnosis of invasive cancer and status of the ER, PR, and HER2 receptors were also determined from tissue obtained from the biopsy. Following biopsy, all subjects were treated with chemotherapy comprising the following drugs and dosage protocol: weekly paclitaxel (80 mg/m^2/wk) $\times$ 12 courses followed by 5-fluorouracil (500 mg/m^2), doxorubicin (50 mg/m^2), and cyclophosphamide (500 mg/m^2) all on day 1 repeated in 21-day cycles $\times$ 4 courses. Following the completion of the aforementioned chemotherapy (T/FAC), all subjects underwent surgery (modified radical mastectomy or lumpectomy and sentinel lymph node biopsy or axillary node dissection) in order to determine whether a subject experienced pathological complete response to chemotherapy or

whether residual invasive cancer was still present. Pathological complete response to chemotherapy was defined as the absence of any residual invasive cancer at the breast site and at the nearest axillary lymph node site. Ninety three subjects were able to complete the aforementioned treatment protocol in terms of dosage and frequency (number of administered courses) of taxol and the other drugs of the T/FAC treatment. Of those 93 subjects, 20 responded to the T/FAC treatment and had no residual invasive cancer at the end of the six-month course, whereas the remaining 73 did not do so. For more demographic and clinical details, please see the study by Tabchy et al.[3]

Discovery and validation studies

Of the total 93 subjects, we randomly selected 50 of them [10 responders (R) and 40 non-responders (NR)] for the development and training of our prognostic biomarker model. The remaining 43 subjects [10 responders (R) and 33 non-responders (NR)] constituted the unknown subjects with which our prognostic biomarker model was tested. This validation method provided us with the means to test our prognostic biomarker model with 43 new and real unknowns that were different from the subjects used for—and, therefore, completely extraneous to—the development and training of the model. The proportions of the clinical stages (I–III) in the total set of 93 subjects were maintained in both the discovery and validation subsets of subjects. Moreover, subjects with all possible combinations of receptor classifications (ER, PR, and HER2) were included in both the discovery and the validation study with approximately equal proportions. Table 1 shows all clinical information regarding the stage and the receptor status of all 93 subjects, including that of the subjects misclassified by our prognostic biomarker model.

Statistical methods

We processed the original raw intensity data (CEL files) using the Expression Console software by Affymetrix and choosing the RMA algorithm (510K FDA approved) with the standard settings.

In order to reduce the dimensionality of the data and zero in on those variables (transcripts) that are most significant in the process of treatment-response in the case of breast cancer patients, we applied our bioinformatic methods that we have developed, presented, and explained in a great detail in our previous studies.[4–8] Briefly, we performed ROC curve analysis in order to assess the discriminating capability of all variables with respect to our two groups, namely, R (responders) and NR (non-responders). In the final round, we selected only those variables with an AUC $\geq$ 0.770. Fourteen variables fulfilled this

Table 1. Clinical information pertaining to the stage and receptor status of all 93 subjects, including that of the subjects misclassified by the F_1 prognostic biomarker model.

Group	No. of subjects	Stage I	Stage II	Stage III	Unknown stage	ER(+)	ER(–)	PR(+)	PR(–)	HER2(+)	HER2(–)
A. Discovery											
R	10	0	2	5	3	2	8	1	9	3	7
NR	40	6	16	14	4	27	13	25	15	4	36
B. Validation											
R	10	0	1	7	2	2	8	1	9	4	6
NR	33	5	13	10	5	20	13	17	16	4	29
C. Misclassified subjects											
Discovery											
R	1	0	0	0	1	1	0	0	1	0	1
NR	4	1	1	2	0	2	2	2	2	0	4
Validation											
R	1	0	0	1	0	0	1	0	1	1	0
NR	2	0	0	0	2	0	2	0	2	0	2

Notes: The information and the results shown in Table 1 provide evidence that the accuracy of the F_1 prognostic biomarker model is independent of the receptor status of the subject. **A.** Clinical information pertaining to the stage and receptor status of the 50 subjects [10 responders (R) and 40 non-responders (NR)] used in the discovery study. **B.** Clinical information pertaining to the stage and receptor status of the 43 unknown subjects [10 responders (R) and 33 non-responders (NR)] used in the validation study. **C.** Clinical information pertaining to the stage and receptor status of all 8 subjects (5 in the discovery study and 3 in the validation study) misclassified by the F_1 prognostic biomarker model.

criterion, and they constituted the final pool of the most significant variables.

Generation of the F_1 Super Variable (*mathematical function*): From the aforementioned 14 most significant variables, 9 became the input variables to the complex mathematical function F_1, also referred to here as super variable. We should point out that one other super variable was generated employing the remaining of the aforementioned 14 most significant variables, but, following final assessment, it proved to be not as robust as the F_1, and it is consequently not presented here. The 9 input variables (transcripts) to the F_1 super variable correspond to 9 different genes. The F_1 super variable, therefore, is a function of the following 9 genes, presented here in alphabetical order:

$$F_1 = f(\text{CCND1, CELSR1, DKFZp566H0824, FAAH, IGKV1-5, LAMA5, OXCT1, RARA, UBE2J1}) \quad (1)$$

The F_1 super variable (complex mathematical function) (Equation 1) constitutes the final prognostic biomarker model of treatment response of breast cancer patients (clinical stages I–III), and its 9 input variables (genes) are listed in Table 2, along with their name, relative differential expression, and other properties.

Computer programs

Computer programs were written using MATLAB R2011b by The MathWorks, Inc., Natick, MA, USA.

Results

Discovery study

As was mentioned earlier, from the total number of 93 subjects [20 responders (R) and 73 non-responders (NR)] used in this study, we randomly selected 50 subjects [10 responders (R) and 40 non-responders (NR)] for the development and training of the prognostic biomarker model (F_1); and we will henceforward refer to those 50 subjects as the 50 original subjects. After the development of the prognostic biomarker model, we assessed its accuracy using the aforementioned 50 original subjects, which were employed for its development. This constitutes an important first step in the assessment of a prognostic test.

The cut-off score of the F_1 prognostic biomarker model was determined by taking into account the results of the following two analyses: (1) calculation of the optimal point on the ROC curve based on the 50 scores of the 50 original subjects used in the discovery study [optimal point is defined as the point with the highest sensitivity and the lowest false positive rate (1-specificity)] and (2) calculation of the 99.99% confidence intervals for the mean F_1 scores of the two groups (R and NR) and their respective standard deviations. Based on that, the cut-off score of the F_1 model was determined to be 4.6683. If a subject has an F_1 score less than 4.6683, then that subject is classified as an R (responder); otherwise ($\geq$4.6683), that subject is classified as an NR (non-responder). As can be seen from Figure 1, the F_1 model correctly identified (9/10) R subjects and (36/40) NR subjects. Assuming that we are interested in identifying the responders (R) to the T/FAC chemotherapy, our target group is the R group and our reference group is the NR group. It follows, then, that for the discovery study, the F_1 model exhibited a sensitivity = 9/10 = 0.900 and a specificity = 36/40 = 0.900. Figure 1 and Table 3 show all pertinent statistical results of the F_1 prognostic biomarker model in connection with the discovery study in great detail.

Validation study

As was mentioned earlier, from the total number of 93 subjects [20 responders (R) and 73 non-responders (NR)] used in this study, we had randomly segregated 43 subjects [10 responders (R) and 33 non-responders (NR)] for the sole and express purpose of testing our prognostic biomarker model. Those 43 unknown subjects were completely extraneous to the model, that is to say they were new and different from the original 50 subjects used for the development of the model, and they had never before been encountered by it. This, validation by unknown and different subjects, constitutes the most important test in the assessment of a prognostic test.

As can be seen from Figure 2 and Table 4, our prognostic biomarker model (F_1) correctly identified (9/10) R subjects and (31/33) NR subjects from the total of 43 unknown subjects used in the validation study. More specifically, 9/10 R subjects had F_1 scores that were less than the 4.6683 cut-off value, and 31/33 NR subjects had F_1 scores that were $\geq$4.6683. Therefore, in connection

with the validation study, the sensitivity of the F_1 prognostic model was (9/10) = 0.900, and the specificity was (31/33) = 0.939.

Table 4, in addition to other pertinent statistical results of our prognostic biomarker model, shows the observed mean F_1 scores of the two groups (R and NR) of the 43 unknown subjects in the validation study. As can be seen, both of those group mean scores, as observed in the validation study with the 43 unknown subjects, fall within the 99.99% confidence intervals of the respective group mean scores as predicted in the discovery study (Table 3).

Overall prognostic biomarker model performance

If we combined the discovery study results with those of the validation study, then the overall performance of our F_1 prognostic biomarker model would be as follows. Overall sensitivity = 0.900 (18/20 R subjects) and overall specificity = 0.918 (67/73 NR subjects). Figure 3 and Table 5 depict those overall results, along with additional pertinent statistical results of the F_1 prognostic biomarker model.

Significant genes

In connection with the aforementioned 9 significant genes that constitute the input variables to the F_1 function (Equation 1), we conducted an Ingenuity Pathway Analysis (IPA) search. We sought to ascertain information about those 9 genes pertaining to their known interactions with other genes; their known interactions with drugs, chemicals, and/or hormones; and their known associations with various types of cancer as derived from the findings of scientific, peer-reviewed studies. The IPA search results are listed in Table 2, along with the direction of the statistically significant differential expression (over-expression or under-expression) of those 9 genes in the NR group (non-responders) relative to that of the R group (responders).

The CCND1 (also cyclin D1) gene encodes a protein that belongs to the cyclin family, the members of which are regulators of CDK kinases. Overexpression of the CCND1 gene, which alters cell cycle progression, has been observed in a variety of tumors and may contribute to tumorigenesis. Moreover, the CCND1 gene has been observed to interact with the BRCA1 and BRCA2 genes, known to be familial breast and ovarian cancer susceptibility genes.[9,10] Over-expression of CCND1 has been shown to play a crucial role in the development and progression of several types of cancers, such as breast, esophageal, bladder, and lung cancer.[11–15] Furthermore, and more importantly, over-expression of CCND1 has been linked to the development of resistance to endocrine drugs in breast cancer cells.[11,16,17] Over-expression of CCDN1 has also been shown to contribute to the progression of breast tumor cells to invasive carcinomas.[18] Those findings are in agreement with our results: we found that the CCND1 gene was significantly over-expressed in the NR group (non-responders) relative to the R group (responders) (Table 2).

The LAMA5 (laminin, alpha 5) gene encodes a protein that belongs to the alpha subfamily of laminin proteins, which constitute a major component of basement membranes, and which affect tissue development in many organs. Over-expression of the LAMA5 gene has been observed in various types of cancer, such as glioma, melanoma, hepatocellular carcinoma, lung adenocarcinoma, breast cancer, ovarian cancer, etc., especially in connection with tumor cell migration and invasiveness.[19–24] In addition to oncogenesis and metastatic colonization, over-expression of laminin has been linked to cytotoxic drug resistance in the case of lung cancer cell lines.[25] Furthermore, in the case of breast cancer cells, it has been shown that over-expression of laminin inhibits estrogen action and leads to resistance to hormonal drugs without the loss of hormone receptors on the part of the breast cancer cells.[23] The above findings are in accord with our results: we found that the LAMA5 gene was significantly over-expressed in the NR group (non-responders) relative to the R group (responders) (Table 2).

The gene FAAH (fatty acid amide hydrolase) encodes a protein that is responsible for the hydrolysis of a number of primary and secondary fatty acid amides. In connection with cancer, it has been observed that over-expression of the FAAH gene resulted in cell invasion and cell migration in prostate carcinoma cells.[26] Moreover, tumor over-expression of FAAH has been associated with prostate cancer severity and outcome,[27] and it has been shown that anti-proliferative effects could be observed in prostate cancer cell lines by inhibiting the FAAH enzyme.[28] In connection with breast cancer, and more specifically regarding

Table 2. The 9 genes (constituent variables) of the F_1 prognostic biomarker model, ranked according to their ROC AUC value.

Rank	ROC AUC	Affymetrix transcript no.	Gene symbol	Gene name	Gene function/ process	Gene signif. diff. expr. non-responders (NR)	Known interactions	Known drugs/ chemicals/ hormones
1	0.80342	208712_at	CCND1	Cyclin D1	G1/S transition of mitotic cell cycle, positive regulation of cyclin-dependent protein kinase activity, response to estrogen stimulus—(Mutations, amplification and overexpression of this gene, which alters cell cycle progression, are observed frequently in a variety of tumors and may contribute to tumorigenesis. [provided by RefSeq, Jul 2008])	↑	ESR1, BRCA1, CDK4, CDK6, TP53, RB1, MCM10	Beta-estradiol, tretinoin, fulvestrant, sirolimus, troglitazone
2	0.80342	210150_s_at	LAMA5	Laminin, alpha 5	Angiogenesis, cell proliferation, cytoskeleton organization, cell migration	↑	MYOC, ACTB, MAPK1, MAPK8, TNF, CD44	Heparin, L-glutamic acid
3	0.80068	204231_s_at	FAAH	Fatty acid amide hydrolase	Fatty acid catabolic process	↑	DINP, PXN	Progesterone, beta-estradiol, Fsh, bucladesine
4	0.79658	216300_x_at	RARA	Retinoic acid receptor, alpha	Cell proliferation, transcription from RNA polymerase II promoter, estrogen receptor signaling pathway, response to estradiol stimulus	↑	ESR1, CDK7, SRC, SLC10A1, HOXB1, HOXA1, CD38, MED1	Tretinoin, beta-estradiol, retinoid, arsenic trioxide
5	0.79452	41660_at	CELSR1	Cadherin, EGF LAG seven-pass G-type receptor 1	G-protein coupled receptor protein signaling pathway, signal transduction	↑	ESR1, PSAP, EGFR	Interferon alpha

6	0.78425	214768_x_at	IGKV1-5	Immunoglobulin kappa variable 1–5	Immune response, innate immune response, complement activation pathway	↓	AGTR1, AGT, MDH1, IGHA1	
7	0.78151	207470_at	DKFZp566H0824	Hypothetical LOC54744		↑		
8	0.77877	217825_s_at	UBE2J1	Ubiquitin-conjugating enzyme E2, J1, U	Post-translational protein modification, ATP binding, adaptive immunity signaling	↓	BRCA1, NFKBIA, BIRC2, TRAF7, TRAF6	Dexamethasone
9	0.77329	202780_at	OXCT1	3-oxoacid CoA transferase 1	CoA-transferase activity, adipose tissue development, cellular lipid metabolic process	↓	HRAS, CNTF, CFTR, PRKAG3, PPARGC1A	Beta-estradiol, bromobenzene, lipopolysaccharide

Notes: The following properties of those 9 genes are shown here: Affymetrix transcript number; gene symbol; gene name; gene function and/or process; significant differential expression [over-expression (↑) or under-expression (↓)] as observed in the NR group (Non-Responders) relative to the R group (Responders); known gene interactions; and known drug/chemical/hormone interactions.

treatment response, it has been observed that FAAH was significantly over-expressed in the subjects that failed to respond to the T/FAC treatment.[29] Those findings are in agreement with those of our study: we found that the FAAH gene was significantly over-expressed in the NR group (non-responders) relative to the R group (responders) (Table 2).

We should point out here that in our previous study on treatment response in ovarian cancer,[5] one of the three gene networks we discovered to play a significant role in the response to the taxol/platinum chemotherapy and survival also pertained to lipid breakdown metabolism (LYPLA2 and OSBPL8), and that strengthens our hypothesis in that study, namely, that aggressive tumor cells effect extensive remodeling of lipid metabolism, presumably for energy purposes.

The RARA gene encodes a protein (retinoic acid receptor alpha) that regulates transcription. In the case of acute promyelocytic leukemia, over-expression of RARA has been shown to induce cell proliferation via direct up-regulation of c-MYC in mice.[30] Over-expression of RARA has also been observed in human ovarian tumor cells.[31] In connection with human breast cancer cells, it has been widely observed that the expression of ER receptor α and that of RARA are coordinated; more specifically, over-expression of the former induces over-expression of the latter in ER-positive breast cancer cells.[32,33] More interestingly, however, regarding our findings, it has also been observed that the crucial biological effects exerted by RARA on human breast cancer cells are mediated regardless of the ER status of those cells.[34,35] Those findings are in agreement with those of our study: we found that the RARA gene was significantly over-expressed in the NR group (non-responders) relative to the R group (responders) (Table 2).

The CELSR1 (cadherin, EGF LAG seven-pass G-type receptor 1) gene encodes a protein that is a member of the flamingo subfamily, which is part of the cadherin superfamily. The flamingo cadherins are located at the plasma membrane and are thought to be receptors involved in contact-mediated cell communication. In squamous cell carcinoma cells, it has been shown that over-expressed G protein-coupled receptor proteins, via communication with EGFR (epidermal growth factor receptor) signaling systems, induce cell

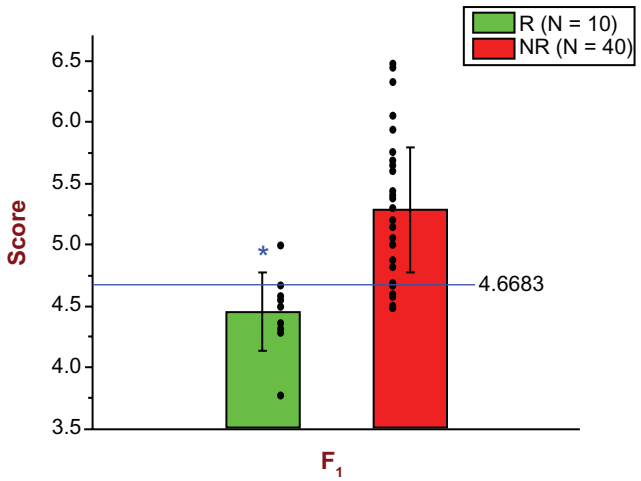

Figure 1. Scatter plot and bar graph of all 50 original subjects [10 responders (R) and 40 non-responders (NR)] used in the discovery study in connection with the F_1 prognostic biomarker model.
Notes: As can be seen, 9/10 R subjects [Responders (green color)] had F_1 scores lower than the determined cut-off score of 4.6683 and were therefore identified correctly by the F_1 prognostic biomarker model [sensitivity = 9/10 = 0.900]. Regarding the NR group [Non-Responders (red color)], 36/40 subjects had F_1 scores greater than the determined cut-off score of 4.6683 and were therefore identified correctly by the F_1 prognostic biomarker model [specificity = 36/40 = 0.900]. For the discovery study, the mean F_1 score of the 10 R subjects (Responders) was 4.4528 (top of the green bar) and their standard deviation (whiskers above or below the top of the green bar) was 0.3156. The mean F_1 score of the 40 NR subjects (Non-Responders) was 5.2833 (top of the red bar) and their standard deviation (whiskers above or below the top of the red bar) was 0.5074. The significance level was set at $\alpha = 0.001$ (two-tailed), and the probability of significance for the F_1 was $P = 1.06 \times 10^{-5}$ (independent *t*-Test with T-value = 4.9195). The F_1 is parametrically distributed with respect to both groups.

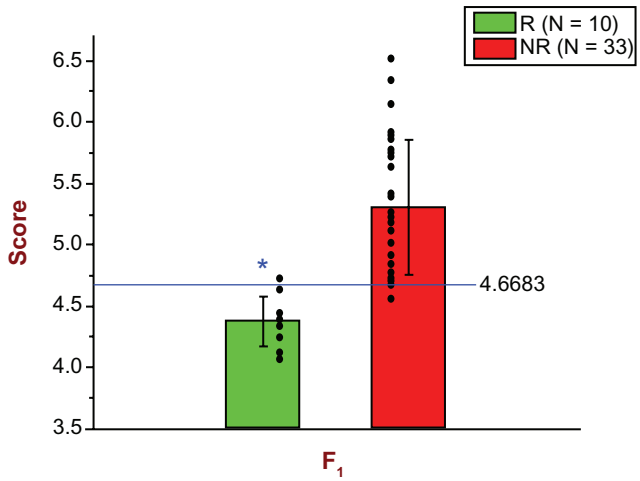

Figure 2. Scatter plot and bar graph of all 43 unknown (new and different) subjects [10 responders (R) and 33 non-responders (NR)] used in the validation study in connection with the F_1 prognostic biomarker model.
Notes: As can be seen, 9/10 R subjects [Responders (green color)] had F_1 scores lower than the determined cut-off score of 4.6683 and were therefore identified correctly by the F_1 prognostic biomarker model [sensitivity = 9/10 = 0.900]. Regarding the NR group [Non-Responders (red color)], 31/33 subjects had F_1 scores greater than the determined cut-off score of 4.6683 and were therefore identified correctly by the F_1 prognostic biomarker model [specificity = 31/33 = 0.939]. For the validation study, the mean F_1 score of the 10 R subjects (Responders) was 4.3766 (top of the green bar) and their standard deviation (whiskers above or below the top of the green bar) was 0.2041. The mean F_1 score of the 33 NR subjects (Non-Responders) was 5.3028 (top of the red bar) and their standard deviation (whiskers above or below the top of the red bar) was 0.5476. The significance level was set at $\alpha = 0.001$ (two-tailed), and the probability of significance for the F_1 was $P = 5.83 \times 10^{-6}$ (independent *t*-Test with T-value = 5.2029). The F_1 is parametrically distributed with respect to both groups.

proliferation and migration.[36] In the case of breast cancer cells, it has been observed that CELSR1 interacts with estrogen receptor (ER).[37] Our findings agree with those observations: the CELSR1 gene was significantly over-expressed in the NR group (non-responders) relative to the R group (responders) (Table 2).

The IGKV1-5 (immunoglobulin kappa variable 1–5) gene encodes a protein whose molecular function is antigen binding, and which is involved in compliment activation, innate immune response, and in regulation of immune response, in general. Although little is known about the exact function of IGKV1-5, it has been shown that it is expressed in leukocytes in human peripheral blood,[38] and that various types of cancer cells effect significant reduction of the expression of immune-response related genes, such as those involved in antigen presentation pathway,[39] genes in the B-cell receptor complex,[40] genes in the human leukocyte antigen (HLA) class,[41] etc. More specifically, in connection with breast cancer, it has been observed

Table 3. Statistical results of the F_1 prognostic biomarker model in the discovery study {identification of the 50 original subjects [10 responders (R) and 40 non-responders (NR)]}.

Prognostic test	ROC AUC	95% CI of ROC AUC	T-value	*P* (2-tailed) $\alpha = 0.001$	R group [99.99% CI of mean] (SD)	NR group [99.99% CI of mean] (SD)
Discovery study						
F_1	0.9425	[0.8159, 0.9829]	4.9195	1.06×10^{-5}	[4.1611, 4.7981] (0.3156)	[5.0103, 5.5300] (0.5074)

Notes: The ROC AUC value, the 95% confidence interval of the ROC AUC Value, the T value and probability of significance (*P*) of the independent *t*-Test, the 99.99% confidence interval for the mean score of the R group (Responders) and that of the NR group (Non-Responders), along with their respective standard deviations, of the F_1 prognostic biomarker model in the discovery study are shown.

Table 4. Statistical results of the F_1 prognostic biomarker model in the validation study {identification of the 43 unknown subjects [10 responders (R) and 33 non-responders (NR)], which were new and different from the 50 original subjects}.

Prognostic test	ROC AUC	Sensitivity	Specificity	T-value	*P* (2-tailed) $\alpha = 0.001$	R group Mean ± SD	NR group Mean ± SD
Validation study							
F_1	0.9788	0.9000	0.9394	5.2029	5.83×10^{-6}	4.3766 ± 0.2041	5.3028 ± 0.5476

Notes: The ROC AUC value, the sensitivity, the specificity, the T value and probability of significance (*P*) of the independent *t*-Test, and the mean score of the R group (Responders) and that of the NR group (Non-Responders), along with their respective standard deviations, of the F_1 prognostic biomarker model in the validation study are shown. As can be seen, both of those group mean scores, as observed in the validation study with the 43 unknown subjects, fall within the 99.99% confidence intervals of the respective group mean scores as predicted in the discovery study (Table 3).

that significant down-regulation of immune-response related genes was significantly associated with tumor progression, nodal involvement, lymphatic invasion, and risk of breast cancer reccurence.[41] Those findings are in accord with our results: we found that the IGKV1-5 gene was significantly under-expressed in the NR group (non-responders) relative to the R group (responders) (Table 2).

The Affymetrix HG-U133 A probe set 207470_at corresponds to DKFZp566H0824 (hypothetical

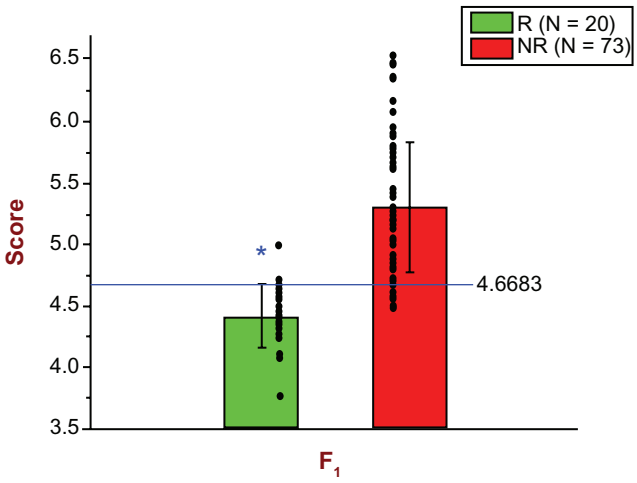

Figure 3. Scatter plot and bar graph of all 93 subjects [20 responders (R) and 73 non-responders (NR)] used in the entire study (discovery and validation) in connection with the F_1 prognostic biomarker model.
Notes: As can be seen, 18/20 R subjects [Responders (green color)] had F_1 scores lower than the determined cut-off score of 4.6683 and were therefore identified correctly by the F_1 prognostic biomarker model [sensitivity = 18/20 = 0.900]. Regarding the NR group [Non-Responders (red color)], 67/73 subjects had F_1 scores greater than the determined cut-off score of 4.6683 and were therefore identified correctly by the F_1 prognostic biomarker model [specificity = 67/73 = 0.918]. The mean F_1 score of the 20 R subjects (Responders) was 4.4148 (top of the green bar) and their standard deviation (whiskers above or below the top of the green bar) was 0.2617. The mean F_1 score of the 73 NR subjects (Non-Responders) was 5.2921 (top of the red bar) and their standard deviation (whiskers above or below the top of the red bar) was 0.5223. The significance level was set at $\alpha = 0.001$ (two-tailed), and the probability of significance for the F_1 was $P = 1.36 \times 10^{-10}$ (independent *t*-Test with T-value = 7.2454). The F_1 is parametrically distributed with respect to both groups.

LOC54744). According to our results, this unknown gene was significantly over-expressed in the NR group (non-responders) relative to the R group (responders) (Table 2).

The UBE2J1 (ubiquitin-conjugating enzyme E2, J1, U) gene encodes a protein that is a member of the E2 ubiquitin-conjugating enzyme family. The modification of proteins with ubiquitin is an important cellular mechanism that targets abnormal or short-lived proteins for degradation. It has been shown that BRCA1, via its binding to UBE2J1, as well as to other members of the E2 family, directs the synthesis of specific polyubiquitin chain linkages.[42] Given that BRCA1 functions as tumor suppressor and plays a role in DNA damage repair,[43] it follows that an abnormal down-regulation of BRCA1 would most likely entail a down-regulation of UBE2J1. That would be consistent with our findings: the UBE2J1 gene was significantly under-expressed in the NR group (non-responders) relative to the R group (responders) (Table 2).

The OXCT1 (3-oxoacid CoA transferase 1) gene encodes a protein that is a mitochondrial matrix enzyme and plays a central role in ketone metabolism. Among other biological processes, OXCT1 is involved in adipose tissue development and cellular lipid metabolism. It has been observed that HRAS, a well-known oncogene involved in many different types of cancer, suppresses the expression of OXCT1.[44] In connection with breast cancer, it has been shown that 69% of breast cancer tumors exhibit an over-expression of HRAS, which is associated positively with disease progression and lymph node involvement and negatively with response to treatment.[45] It has also been shown that over-expression of HRAS in breast cancer tumors can be constitutively mediated via deregulation of HER2, ER, EGFR, and other receptors.[46–48] That, therefore, over-expression of HRAS in aggressive breast tumor cells leads to sup-

Table 5. Overall statistical results of the F_1 prognostic biomarker model with respect to both the discovery and the validation studies combined {identification of 93 subjects [20 responders (R) and 73 non-responders (NR)]}.

Prognostic test	ROC AUC	Sensitivity	Specificity	T-value	*P* (2-tailed) $\alpha = 0.001$	R group Mean ± SD	NR group Mean ± SD
Overall results (discovery and validation studies)							
F_1	0.9616	0.9000	0.9178	7.2454	1.36×10^{-10}	4.4148 ± 0.2617	5.2921 ± 0.5223

Notes: The overall ROC AUC value, the overall sensitivity, the overall specificity, the T value and probability of significance (*P*) of the independent *t*-Test, and the mean score of the R group (Responders) and that of the NR group (Non-Responders), along with their respective standard deviations, of the F1 prognostic biomarker model with respect to both the discovery and the validation studies combined are shown.

pression of the expression of OXCT1 accords with our finding: the OXCT1 gene was significantly underexpressed in the NR group (non-responders) relative to the R group (responders) (Table 2).

Discussion

To the best of our knowledge, we are not aware of the existence of any prognostic tests that can predict which breast cancer patients will respond to chemotherapy with both sensitivity and specificity $> 80\%$, as mandated by the latest FDA requirements. Having employed 50 subjects [10 responders (R) and 40 non-responders (NR)], we were able to develop a prognostic test that—based on global gene expression analysis of tumor tissue collected during biopsy and prior to the commencement of chemotherapy—can identify with a high accuracy those patients with breast cancer (clinical stages I–III) who will respond to the T/FAC chemotherapy and will experience pathological complete response (Responders), as well as those breast cancer patients (clinical stages I–III) who will not do so (Non-Responders). Following validation with 43 unknown (new and different) subjects [10 responders (R) and 33 non-responders (NR)], our prognostic test (F_1) exhibited an overall sensitivity = 0.900 (18/20 R subjects) and overall specificity = 0.918 (67/73 NR subjects). Given the relatively small number of responders in this study (20 R vs. 73 NR), it stands to reason that the robust performance of our prognostic test should be further tested using a wider pool of subjects not only in terms of a larger number of responders but also in terms of demographics, family history, and syndromic associations.

Furthermore, we are equally unaware of the existence of any prognostic tests that can predict which breast cancer patients will respond to chemotherapy with both sensitivity and specificity $\geq 90\%$ and, at the same time, regardless of the status of the three hormone receptors (ER, PR, and HER2); and that constitutes the highest contribution of our study. As can be seen from the information and results shown in Table 1, our prognostic test with both sensitivity and specificity $\geq 90\%$ is applicable, and can be administered, to all breast cancer patients independently of the status of the hormone receptors ER, PR, and HER2, as well as of the ethnicity and age of the patients. In contrast, other breast cancer prognostic tests currently in the market not only have limited accuracy (sensitivity and specificity $< 80\%$) but also limited applicability: they can be administered only to specific combinations of the aforementioned three hormone receptors, that is to say, they can be administered to a small subset of the population of the breast cancer patients. Conversely, that also means that a large fraction of the women with breast cancer cannot avail themselves of those prognostic tests, and that, therefore, they cannot be enabled to make accurate decisions about treatment and management of their disease.

Once again, provided there is further and more extensive validation, the clinical significance of our prognostic test in the field of breast cancer can be summarized in the following. (1) Our prognostic test could be applied to all breast cancer patients in spite of receptor status, age, or ethnicity. (2) Physicians will have the ability to identify with a high degree of accuracy both the responders and the non-responders to current chemotherapy at the outset (at the time of the biopsy and prior to the commencement of chemotherapy). (3) Alternative therapies may be provided to those patients identified as non-responders to chemotherapy at the beginning, saving, thus, valuable and critical time, and increasing the probability of a favorable outcome. (4) In connection with providing the non-responders to the T/FAC

chemotherapy with effective drugs, our prognostic test and the findings of our study pertaining to the aforementioned nine important genes can assist pharmaceutical companies to test and develop new analogs of chemotherapeutic agents or new cocktails of small molecules that can modulate most, if not all, of those nine genes.

In the end, all of the above may significantly extend the survival of all breast cancer patients, regardless of whether they are deemed responders or non-responders to the current chemotherapy, and regardless of their status of the three hormone receptors, their age, or their ethnicity.

Author Contributions

JBN conceived and developed the mathematical theory of super variables and generated the F_1 super variable in this study. JBN co-conceived, designed, performed the analysis, and executed this project; and he wrote and co-edited the manuscript. PAB and WCL co-conceived this project, participated in the discussions, and co-edited the manuscript. PAB provided the necessary support and resources for this project.

Competing Interests

JBN, WCL and PAB hold stock and have received funding from AIS for travel, accommodation or meetings unrelated to activities disclosed here and their employer. The authors are named inventors in a patent application related to this study.

Disclosures and Ethics

As a requirement of publication author(s) have provided to the publisher signed confirmation of compliance with legal and ethical obligations including but not limited to the following: authorship and contributorship, conflicts of interest, privacy and confidentiality and (where applicable) protection of human and animal research subjects. The authors have read and confirmed their agreement with the ICMJE authorship and conflict of interest criteria. The authors have also confirmed that this article is unique and not under consideration or published in any other publication, and that they have permission from rights holders to reproduce any copyrighted material. Any disclosures are made in this section. The external blind peer reviewers report no conflicts of interest.

References

1. American Cancer Society. Cancer Facts and Figures 2010. Atlanta: American Cancer Society; 2010.
2. Jemal A, Siegel R, Xu J, Ward E. Cancer statistics, 2010. *CA Cancer J Clin*. 2010;60:277–300.
3. Tabchy A, Valero V, Vidaurre T, Lluch A, Gomez H, Martin M, et al. Evaluation of a 30-gene paclitaxel, fluorouracil, doxorubicin, and cyclophosphamide chemotherapy response predictor in a multicenter randomized trial in breast cancer. *Clin Cancer Res*. 2010;16:5351–61.
4. Nikas JB, Low WC. Linear discriminant functions in connection with the micro-RNA diagnosis of colon cancer. *Cancer Informatics*. 2012;11:1–14.
5. Nikas JB, Boylan KLM, Skubitz APN, Low WC. Mathematical prognostic biomarker models for treatment response and survival in epithelial ovarian cancer. *Cancer Informatics*. 2011;10:233–47.
6. Nikas JB, Low WC. Application of clustering analyses to the diagnosis of Huntington disease in mice and other diseases with well-defined group boundaries. *Computer Methods and Programs in Biomedicine*. 2011;104(3):e133–47.
7. Nikas JB, Low WC. ROC-supervised principal component analysis in connection with the diagnosis of diseases. *American Journal of Translational Research*. 2011;3(2):180–96.
8. Nikas JB, Keene CD, Low WC. Comparison of analytical mathematical approaches for identifying key nuclear magnetic resonance spectroscopy biomarkers in the diagnosis and assessment of clinical change of diseases. *Journal of Comparative Neurology*. 2010;518:4091–112.
9. Wang H, Shao N, Ding QM, Cui J, Reddy ESP, Rao VN. BRCA1 proteins are transported to the nucleus in the absence of serum and splice variants BRCA1a, BRCA1b are tyrosine phosphoproteins that associate with E2F, cyclins and cyclin dependent kinases. *Oncogene*. 1997;15:143–57.
10. Plevova P, Cerna D, Balcar A, Foretova L, Zapletalova J, Silhanova E, et al. CCND1 and ZNF217 gene amplification is equally frequent in BRCA1 and BRCA2 associated and non-BRCA breast cancer. *Neoplasma*. 2010;57(4):325–32.
11. Alao JP. The regulation of cyclin D1 degradation: roles in cancer development and the potential for therapeutic invention. *Molecular Cancer*. 2007;6:24.
12. Gillett C, Smith P, Gregory W, Richards M, Millis R, Peters G, et al. Cyclin D1 and prognosis in human breast cancer. *Int J Cancer*. 1996;69(2):92–9.
13. Knudsen KE, Diehl JA, Haiman CA, Knudsen ES. Cyclin D1: polymorphism, aberrant splicing and cancer risk. *Oncogene*. 2006;25(11):1620–8.
14. Motokura T, Arnold A. Cyclin D and oncogenesis. *Curr Opin Genet Dev*. 1993;3(1):5–10.
15. Weinstat-Saslow D, Merino MJ, Manrow RE, Lawrence JA, Bluth RF, Wittenbel KD, et al. Overexpression of cyclin D mRNA distinguishes invasive and in situ breast carcinomas from non-malignant lesions. *Nat Med*. 1995;1(12):1257–60.
16. Hui R, Finney GL, Carroll JS, Lee CS, Musgrove EA, Sutherland RL. Constitutive overexpression of cyclin D1 but not cyclin E confers acute resistance to antiestrogens in T-47D breast cancer cells. *Cancer Res*. 2002;62(23):6916–23.
17. Kenny FS, Hui R, Musgrove EA, Gee JM, Blamey RW, Nicholson RI, et al. Overexpression of cyclin D1 messenger RNA predicts for poor prognosis in estrogen receptor-positive breast cancer. *Clin Cancer Res*. 1999;5(8):2069–76.
18. Yang C, Trent S, Ionescu-Tiba V, Lan L, Shioda T, et al. Identification of cyclin D1- and estrogen-regulated genes contributing to breast carcinogenesis and progression. *Cancer Res*. 2006;66(24):11649–58.
19. Kawatakia T, Yamanec T, Naganumab H, Rousselled P, Andurena I, Tryggvason K, et al. Laminin isoforms and their integrin receptors in glioma cell migration and invasiveness: Evidence for a role of α5-laminin(s) and α3β1 integrin. *Exp Cell Res*. 2007;313:3819–31.
20. Oikawa Y, Hansson J, Sasaki T, Rousselle P, Domogatskaya A, Rodin S, et al. Melanoma cells produce multiple laminin isoforms and strongly migrate on α5 laminin(s) via several integrin receptors. *Exp Cell Res*. 2011;317(8):1119–33.
21. Kikkawa Y, Sudo R, Kon J, Mizuguchi T, Nomizu M, Hirata K, et al. Laminin alpha 5 mediates ectopic adhesion of hepatocellular carcinoma through integrins and/or Lutheran/basal cell adhesion molecule. *Exp Cell Res*. 2008;314(14):2579–90.

22. Moriya Y, Niki T, Yamada T, Matsuno Y, Kondo H, Hirohashi S. Increased expression of laminin-5 and its prognostic significance in lung adenocarcinomas of small size. An immunohistochemical analysis of 102 cases. *Cancer*. 2001;91(6):1129–41.
23. Woodward TL, Lu H, Haslam SZ. Laminin inhibits estrogen action in human breast cancer cells. *Endocrinology*. 2000;141(8):2814–21.
24. Capo-chichi CD, Roland IH, Vanderveer L, Bao R, Yamagata T, Hirai H, et al. Anomalous expression of epithelial differentiation-determining GATA factors in ovarian tumorigenesis. *Cancer Res*. 2003;63(16):4967–77.
25. Fridman R, Giaccone G, Kanemoto T, Martin GR, Gazdar AF, Mulshine JL. Reconstituted basement membrane (matrigel) and laminin can enhance the tumorigenicity and the drug resistance of small cell lung cancer cell lines. *Proc Natl Acad Sci U S A*. 1990;87(17):6698–702.
26. Endsley MP, Thill R, Choudhry I, Williams CL, Kajdacsy-Balla A, Campbell WB, et al. Expression and function of fatty acid amide hydrolase in prostate cancer. *Int J Cancer*. 2008;123(6):1318–26.
27. Thors L, Bergh A, Persson E, Hammarsten P, Stattin P, Egevad L, et al. Fatty acid amide hydrolase in prostate cancer: association with disease severity and outcome, CB1 receptor expression and regulation by IL-4. *PLoS One*. 2010;5(8):e12275.
28. Brown I, Cascio MG, Wahle KW, Smoum R, Mechoulam R, Ross RA, et al. Cannabinoid receptor-dependent and -independent anti-proliferative effects of omega-3 ethanolamides in androgen receptor-positive and -negative prostate cancer cell lines. *Carcinogenesis*. 2010;31(9):1584–91.
29. Ayers M, Symmans WF, Stec J, Damokosh AI, Clark E, Hess K, et al. Gene expression profiles predict complete pathologic response to neoadjuvant paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide chemotherapy in breast cancer. *J Clin Oncol*. 2004;22:2284–93.
30. Rice KL, Hormaeche I, Doulatov S, Flatow JM, Grimwade D, Mills KI, et al. Comprehensive genomic screens identify a role for PLZF-RARalpha as a positive regulator of cell proliferation via direct regulation of c-MYC. *Blood*. 2009;114(27):5499–511.
31. Katsetos CD, Stadnicka I, Boyd JC, Ehya H, Zheng S, Soprano CM, et al. Cellular distribution of retinoic acid receptor-alpha protein in serous adenocarcinomas of ovarian, tubal, and peritoneal origin: comparison with estrogen receptor status. *Am J Pathol*. 1998;153(2):469–80.
32. Lu M, Mira-y-Lopez R, Nakajo S, Nakaya K, Jing Y. Expression of estrogen receptor alpha, retinoic acid receptor alpha and cellular retinoic acid binding protein II genes is coordinately regulated in human breast cancer cells. *Oncogene*. 2005;24(27):4362–9.
33. Ross-Innes CS, Stark R, Holmes KA, Schmidt D, Spyrou C, Russell R, et al. Cooperative interaction between retinoic acid receptor-alpha and estrogen receptor in breast cancer. *Genes Dev*. 2010;24(2):171–82.
34. Fitzgerald P, Teng M, Chandraratna RA, Heyman RA, Allegretto EA. Retinoic acid receptor alpha expression correlates with retinoid-induced growth inhibition of human breast cancer cells regardless of estrogen receptor status. *Cancer Res*. Jul 1, 1997;57(13):2642–50.
35. Schneider SM, Offterdinger M, Huber H, Grunt TW. Activation of retinoic acid receptor alpha is sufficient for full induction of retinoid responses in SK-BR-3 and T47D human breast cancer cells. *Cancer Res*. 2000;60(19):5479–87.
36. Gschwind A, Hart S, Fischer OM, Ullrich A. TACE cleavage of proamphiregulin regulates GPCR-induced proliferation and motility of cancer cells. *EMBO J*. 2003;22(10):2411–21.
37. Carroll JS, Liu XS, Brodsky AS, Li W, Meyer CA, Szary AJ, et al. Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell*. Jul 15, 2005; 122(1):33–43.
38. Palmer C, Diehn M, Alizadeh AA, Brown PO. Cell-type specific gene expression profiles of leukocytes in human peripheral blood. *BMC Genomics*. 2006;7:115.
39. Yoshihara K, Tsunoda T, Shigemizu D, Fujiwara H, Hatae M, Fujiwara H, et al. High-risk ovarian cancer based on 126-gene expression signature is uniquely characterized by down-regulation of antigen presentation pathway. *Clin Cancer Res*. 2012. doi:10.1158/1078-0432.CCR-11-2725.
40. Huang X, Takata K, Sato Y, Tanaka T, Ichimura K, Tamura M, et al. Downregulation of the B-cell receptor signaling component CD79b in plasma cell myeloma: a possible post transcriptional regulation. *Pathol Int*. 2011;61(3):122–9.
41. Kaneko K, Ishigami S, Kijima Y, Funasako Y, Hirata M, Okumura H, et al. Clinical implication of HLA class I expression in breast cancer. *BMC Cancer*. 2011;11:454.
42. Christensen DE, Brzovic PS, Klevit RE. E2-BRCA1 RING interactions dictate synthesis of mono- or specific polyubiquitin chain linkages. *Nat Struct Mol Biol*. 2007;14(10):941–8.
43. Wang C, Fan S, Li Z, Fu M, Rao M, Ma Y, et al. Cyclin D1 antagonizes BRCA1 repression of estrogen receptor alpha activity. *Cancer Res*. 2005;65(15):6557–67.
44. Yoon JW, Kita Y, Frank DJ, Majewski RR, Konicek BA, Nobrega MA, et al. Gene expression profiling leads to identification of GLI1-binding elements in target genes and a role for multiple downstream pathways in GLI1-induced cell transformation. *J Biol Chem*. 2002;277(7):5548–55.
45. Clair T, Miller WR, Cho-Chung YS. Prognostic significance of the expression of a ras protein with a molecular weight of 21,000 by human breast cancer. *Cancer Res*. 1987;47(20):5290–3.
46. Zeng X, Shaikh FY, Harrison MK, Adon AM, Trimboli AJ, Carroll KA, et al. The Ras oncogene signals centrosome amplification in mammary epithelial cells through cyclin D1/Cdk4 and Nek2. *Oncogene*. 2010;29(36):5103–12.
47. Janes PW, Daly RJ, deFazio A, Sutherland RL. Activation of the Ras signalling pathway in human breast cancer cells overexpressing erbB-2. *Oncogene*. 1994;9(12):3601–8.
48. Hoadley KA, Weigman VJ, Fan C, Sawyer LR, He X, Troester MA, et al. EGFR associated expression profiles vary with breast tumor subtype. *BMC Genomics*. 2007;8:258.