PRECLINICAL STUDY

Development of a robust RNA-based classifier to accurately determine ER, PR, and HER2 status in breast cancer clinical samples

Timothy R. Wilson · Yuanyuan Xiao · Jill M. Spoerke · Jane Fridlyand ·
Hartmut Koeppen · Eloisa Fuentes · Ling Y. Huw · Ilma Abbas ·
Arjan Gower · Erica B. Schleifman · Rupal Desai · Ling Fu · Teiko Sumiyoshi ·
Joyce A. O'Shaughnessy · Garret M. Hampton · Mark R. Lackner

Abstract Breast cancers are categorized into three sub-types based on protein expression of estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor-2 (HER2/ERBB2). Patients enroll onto experimental clinical trials based on ER, PR, and HER2 status and, as receptor status is prognostic and defines treatment regimens, central receptor confirmation is critical for interpreting results from these trials. Patients enrolling onto experimental clinical trials in the metastatic setting often have limited available archival tissue that might better be used for comprehensive molecular profiling rather than slide-intensive reconfirmation of receptor status. We developed a Random Forests-based algorithm using a training set of 158 samples with centrally confirmed IHC status, and subsequently validated this algorithm on multiple test sets with known, locally determined IHC status. We observed a strong correlation between target mRNA expression and IHC assays for HER2 and ER, achieving an overall accuracy of 97 and 96 %, respectively. For determining PR status, which had the highest discordance between central and local IHC, incorporation of expression of co-regulated genes in a multivariate approach added predictive value, outperforming the single, target gene approach by a 10 % margin in overall accuracy. Our results suggest that multiplexed qRT-PCR profiling of *ESR1*, *PGR*, and *ERBB2* mRNA, along with several other subtype associated genes, can effectively confirm breast cancer subtype, thereby conserving tumor sections and enabling additional biomarker data to be obtained from patients enrolled onto experimental clinical trials.

Keywords Breast cancer · Receptor concordance · ER · PR · HER2 · Random forest · IHC · Biomarkers · qPCR

Timothy R. Wilson and Yuanyuan Xiao contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s10549-014-3163-8) contains supplementary material, which is available to authorized users.

T. R. Wilson · J. M. Spoerke · L. Y. Huw · I. Abbas · A. Gower · E. B. Schleifman · R. Desai · L. Fu · T. Sumiyoshi · G. M. Hampton · M. R. Lackner (✉)
Department of Oncology Biomarker Development, Genentech Inc., 1 DNA Way, South San Francisco, CA, USA
e-mail: mlackner@gene.com; lackner.mark@gene.com

Y. Xiao · J. Fridlyand
Department of Biostatistics, Genentech Inc., 1 DNA Way, South San Francisco, CA, USA

H. Koeppen · E. Fuentes
Department of Research Pathology, Genentech Inc., 1 DNA Way, South San Francisco, CA, USA

J. A. O'Shaughnessy
US Oncology, Dallas, TX, USA

J. A. O'Shaughnessy
Baylor Sammons Cancer Center, Dallas, TX, USA

J. A. O'Shaughnessy
Texas Oncology, Dallas, TX, USA

Introduction

Breast cancer is a heterogeneous disease that is routinely categorized by immunohistochemical (IHC) staining of estrogen receptor (ER), progesterone receptor (PR), and the human epidermal growth factor receptor-2 (HER2) [1]. Hormone receptor positive cancers stain positive for either ER or PR and are generally treated with anti-hormone

therapies such as tamoxifen and letrozole [2]. HER2 positive breast cancers overexpress the *HER2/neu* oncogene and are treated with HER2-directed therapies, such as trastuzumab, lapatinib, pertuzumab, and trastuzumab emtansine, which have improved outcomes and changed the natural history of this previously poor prognostic patient subgroup [3]. Triple negative breast cancers (TNBC) lack ER, PR, and HER2 staining and currently have no approved targeted therapies. TNBC is generally treated with combinations of cytotoxic agents such as anthracyclines and taxanes [4].

Over a decade ago, Perou and others provided evidence that breast cancer could be divided molecularly into distinct subgroups based on RNA microarray experiments [5–8]. Follow-up studies demonstrated that a 50-gene signature (subsequently designated the PAM50) could recapitulate the initial studies by Perou and colleagues [9]. Using this assay, breast cancer can be stratified into luminal A and luminal B subgroups that mostly comprise hormone receptor positive breast cancers; basal-like subgroup that mostly comprises triple negative breast cancers; HER2-enriched subgroup that mostly comprises HER2+ breast cancers, and a normal-like subgroup that has been proposed to mostly comprise the normal surrounding stroma [10]. PAM50 analysis has been shown to provide independent prognostic information compared to standard IHC classification. However, the PAM50 intrinsic subtypes show imperfect agreement with IHC classification, not surprisingly since they were designed to provide additional orthogonal information [11], suggesting that predictors that can recapitulate IHC status may still have independent clinical utility in the context of currently approved therapies.

Specifically, despite the promise of molecular assays such as PAM50 and other classifiers, patients are still treated and routinely enrolled onto clinical trials of experimental anti-cancer agents based on IHC determination of ER, PR, and HER2. Such testing requires at least three tissue sections, and often more if staining needs to be repeated. Moreover, if HER2 status is deemed equivocal (i.e., 2+), additional testing is required using a fluorescence in situ hybridization (FISH) assay [12]. Moreover, IHC is subject to variation in inter-pathologist interpretation, and as such, concordance between laboratories can vary. In several studies, discordant results were noted between IHC staining between local and central assessment. In a study by Martinez and colleagues, ER showed a concordance rate of 92 %, PR showed a concordance rate of 78 %, and HER2 showed a concordance rate of 83 % [13]. In a similar study, Orlando et al. demonstrated concordance rates of 82 % for ER, 86 % for PR, and 73 % for HER2 [14]. Based on these considerations, a facile method of subtype assignment that uses minimal slides and shows high concordance with central IHC testing would be highly desirable from a clinical biomarker perspective.

Patients enrolled in experimental clinical trials are often late line metastatic patients who may have undergone multiple rounds of diagnostic testing and often have limited tissue remaining. Comprehensive biomarker profiling of study-enrolled patients is thus challenging and could benefit from multiplexed technologies rather than more traditional individual slide-based assays. In this study, we showed that ER, PR, and HER2 status could be accurately confirmed using multiplexed RNA expression profiling, thereby conserving tissue sections and maximizing the value of limited tissue samples from clinical trial patients.

Results

Comparison of local and central IHC testing for ER, PR, and HER2 status

As ER, PR, and HER2 status is typically determined by different pathologists in various institutions (local testing) and is subject to inter-observer variability, we re-analyzed the expression of ER, PR, and HER2 using a single pathologist (central testing) from 158 patients enrolled in the United States Oncology (USO) phase III study 01062 (USO 01062), which tested the addition of capecitabine to standard adjuvant therapy in high-risk breast cancer [15]. Receptor positivity, here and throughout the manuscript, is defined as an IHC3+ score for HER2 and an Allred cutoff of 3 and above for ER and PR. In general, a strong concordance was observed between local and central testing for ER, PR, and HER2 status (Table 1: 92, 84 and 94 % accuracy, respectively), which is similar to previously published studies [13, 14]. However, when breaking down concordances into positive predictive values (PPV) and negative predictive values (NPV), discrepancies were high for certain subgroups of patients. Specifically, a low PPV of local testing was observed in the HER2+ subgroup of patients. Only 16 of 24 samples positive by local testing were confirmed by central testing (PPV = 67 %), although 132 of 134 samples negative by local testing were negative by central testing (NPV = 99 %). Similarly for PR testing, a low NPV was observed for local testing of PR negative patients (NPV = 68 %), with only 50 of the 73 negative samples by local testing confirmed by central testing.

Generation of a training set to develop the predictive algorithm for ER, PR, and HER2 status

Receptor status is critical for enrollment and stratification of breast cancer patients onto clinical trials, therefore we sought to determine whether a molecular classifier could accurately predict molecular subtype with the benefit of obtaining additional valuable biomarker data that ER, PR,

Table 1 HER2, ER, and PR status by local and central IHC for 158 USO 01062 study Samples

HER2	Central IHC			ER	Central IHC			PR	Central IHC		
	−	+			−	+			−	+	
Local IHC				Local IHC				Local IHC			
−	132	2	NPV 99 %	−	50	8	NPV 86 %	−	50	23	NPV 68 %
+	8	16	PPV 67 %	+	4	96	PPV 96 %	+	3	82	PPV 96 %
	Spec 94 %	Sens 88 %	Acc 94 %		Spec 93 %	Sens 92 %	Acc 92 %		Spec 94 %	Sens 78 %	Acc 84 %

Spec specificity, *Sens* sensitivity, *PPV* positive predictive value, *NPV* negative predictive value, *Acc* accuracy

and HER2 IHC assays would not capture. Using the 158 IHC centrally confirmed patient samples, we extracted RNA and profiled the samples using a previously reported 96-gene BioMark[TM] Fluidigm microfluidics quantitative RT-PCR-based platform that was developed to capture expression of breast cancer genes involved in proliferation and various aspects of breast cancer signaling [16].

Several studies have shown that mRNA expression of the three receptors correlates with IHC assays for the corresponding protein expression [11, 17]. Similarly, we observed a strong correlation between the centrally confirmed IHC status of ER, PR, and HER2 and their corresponding target gene expression, *ESR1*, *PGR*, and *ERBB2* (Fig. 1 and Supplemental Fig. 1, N = 158). Pearson correlation coefficients of IHC and target gene expression were 0.91, 0.80, and 0.60 for ER, PR, and HER2, respectively. In a receiver operating characteristic (ROC) analysis, the area under the curve (AUC) for predicting HER2 IHC3+ samples using *ERBB2* expression was 0.998 (Fig. 1a). To confirm that *ERBB2* qRT-PCR data could accurately determine HER2 status, we developed a DNA-based HER2 copy number assay that showed a high correlation with *ERBB2* gene expression that more accurately captured the HER2 amplified cases based on central confirmation (Supplemental Fig. 2). Notably, three of the cases that were HER2 IHC positive by local but not central testing clustered with the IHC negatives when assessed using this methodology.

For the determination of ER and PR status, we utilized the Allred scoring criteria for positivity [18, 19] and applied ROC analysis to examine the positive predictability defined by different Allred score cutoffs using RNA expression level. For ER and PR status, an AUC of 0.98 and 0.90, respectively, was observed at the FDA-recommended Allred cutoff of 3 and above (Fig. 1b, c).

Performance of target gene prediction

Encouraged by the strong RNA IHC correlations observed for the three target genes (Fig. 1) and to formally address the question whether mRNA target expression can be used to determine central IHC status reliably in a statistical framework, we applied a two-component Gaussian mixture model approach to the empirical distributions of *ERBB2*, *ESR1*, and *PGR* expression, motivated by the bimodal marginal distribution of these genes (Fig. 2). A cutoff point between high and low expression was selected at the value where the posterior probabilities for the two components were equal. For the determination of HER2, ER, and PR status, this target gene prediction (TGP) approach achieved an overall accuracy of 97, 96, and 81 %, respectively (Table 2). Although PR status determination remained challenging by TGP, there was a pronounced improvement in determining HER2 and ER status by TGP compared to local IHC. The PPV of HER2, which was 67 % for local testing was increased to 100 % for TGP. This high concordance between TGP and central IHC status, especially in the HER2 and ER groups, suggests that the PCR-based assay is a reliable surrogate for central IHC to confirm receptor status.

Performance of multivariate gene prediction

To investigate whether incorporating information from other genes in addition to target genes in the BioMark[TM] panel will further improve the prediction performance, especially in predicting the more challenging PR status, we applied and contrasted several multivariate classifiers. These include Random Forests (RF) [20], prediction analysis of microarrays (PAM) [21], and a Random Forests and K-Nearest Neighbors combination approach (RF-KNN), which utilizes RF for variable selection and KNN for formalizing prediction (see "Materials and methods" section and Table 3). Both RF-based approaches performed better than PAM, and achieved an overall cross-validated accuracy of 98, 95, and 91 % for HER2, ER, and PR status prediction. Variable importance measures (VIM) produced by RF were elicited to quantify the relative importance of genes in contributing to prediction accuracy (Fig. 3). For HER2 status prediction, the genes that predicted this group were *ERBB2* and *GRB7*, which is often co-amplified with *ERBB2* [22], with a sensitivity and specificity of 83 and 100 %, respectively (Table 3). For ER status prediction, the gene with the highest VIM and was most often selected by RF-KNN was *ESR1*,

Fig. 1 Target gene expression of **a** ERBB2 **b** ESR1 **c** PGR by central IHC status. *Left panel*: *boxplot* of target gene expression by central HER2 status. IHC positive group is colored in *gray*. *Right panel*: ROC analysis for predicting IHC positivity defined by different cutoffs using target gene expression. Figure legend indicates cutoffs and AUCs with 95 % confidence intervals in *parentheses*
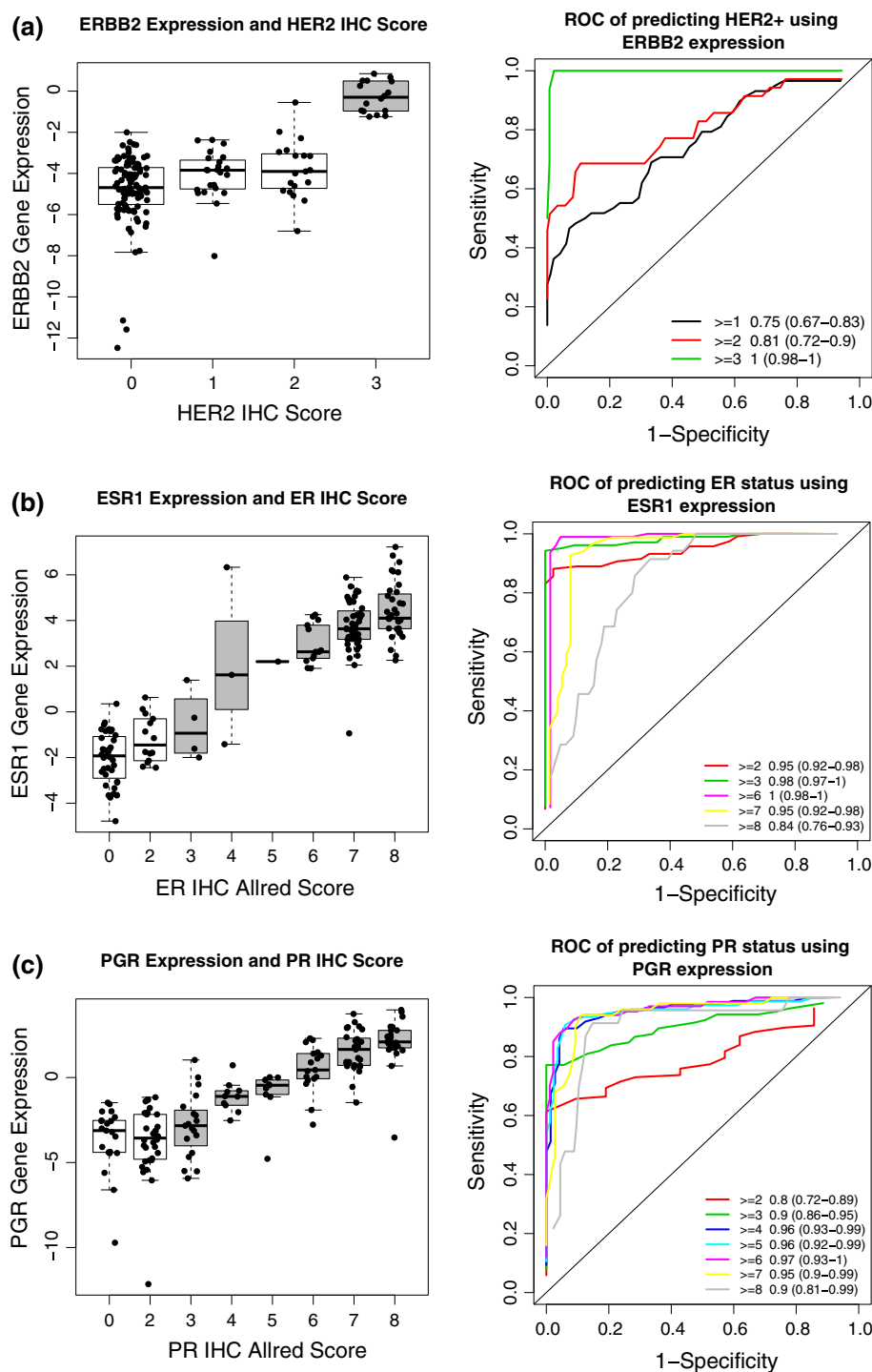
followed by known *ER*-regulated genes, such as *GATA3*, *PGR*, *SCUBE2*, and *FOXA1* [7, 23–25], with a sensitivity and specificity of 93 and 98 %, respectively. For PR status prediction, the most predictive gene was *ESR1*, followed by *PGR* and several other *ER*-regulated genes, with a sensitivity and specificity of 90 and 92 %, respectively. Notably, using a cutoff of Allred score of 4 and above for PR positivity showed a stronger correlation with the expression level of

PGR (Fig. 1c and Supplemental Fig. 3; AUC = 0.96, 95 % CI 0.93–0.99), compared to the FDA-recommended cutoff of 3 and above (AUC = 0.9, 95 % CI 0.86–0.95). This superior AUC for Allred score of 4 and above is supported by a significant increase in *PGR* expression between Allred scores 3 and 4 (*t*-test $P = 0.003$), and conversely, a lack of change in PGR expression between Allred scores 2 and 3 (*t*-test $P = 0.12$).
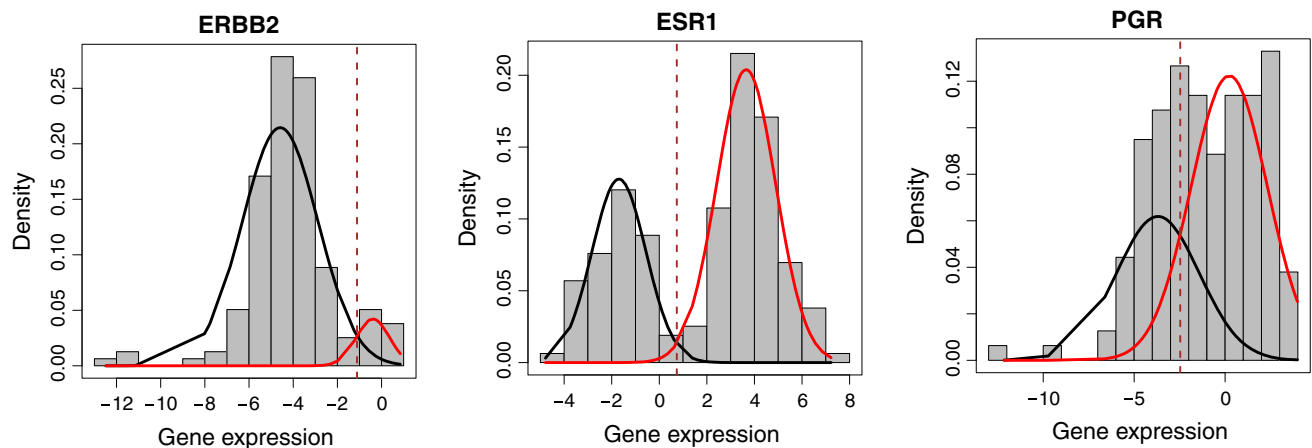
Fig. 2 Target gene prediction. A bimodal, 2-component Gaussian mixture distribution fit was superimposed to the actual data summarized in the histogram. The two mixture distributions are depicted in *red* and *black lines*. The *dotted line* indicates the cutoff between the positive and negative groups

Table 2 HER2, ER, and PR status by central IHC and TGP for the training set

HER2	Central IHC			ER	Central IHC			PR	Central IHC		
	−	+			−	+			−	+	
ERBB2				ESR1				PGR			
−	139	4	NPV 97 %	−	53	5	NPV 90 %	−	39	16	NPV 71 %
+	0	14	PPV 100 %	+	1	99	PPV 99 %	+	14	89	PPV 86 %
	Spec 100 %	Sens 77 %	Acc 97 %		Spec 98 %	Sens 95 %	Acc 96 %		Spec 74 %	Sens 84 %	Acc 81 %

Spec specificity, *Sens* sensitivity, *PPV* positive predictive value, *NPV* negative predictive value, *Acc* accuracy

As the target genes were often the most predictive genes for the determination of ER, PR, or HER2 status, we contrasted the performance of Random Forests-based prediction (RFP; Table 3) with the target gene prediction approach (TGP; Table 2). Both approaches showed similar overall accuracy in predicting central IHC score for HER2 and ER. In the case of PR prediction, RFP outperformed the TGP approach based on *PGR* expression by a 10 % margin in overall accuracy (91 vs. 81 %, respectively). Taken together, these results suggest that inclusion of additional target or co-regulated *PGR* genes in a multivariate approach has a greater predictive power in determining PR status than TGP.

Application of the ER, PR, and HER2 predictive algorithm to test sets

To determine how the predictive algorithm performs in different cohorts of patients, we applied RFP to three separate breast cancer sample sets. First, we applied RFP to an additional 630 patients enrolled onto the USO 01062 study where local ER, PR, and HER2 status was available.

The overall concordance between RFP and local IHC testing for determining HER2, ER, and PR status was 92, 92, and 82 %, respectively (Table 4 and Supplemental Table 1). Figure 4 shows the distribution of target genes *ERBB2, ESR1, and PGR* categorized by the predicted and the local IHC subtypes for ER, PR, and HER2. Similar to the training set, lower concordance was observed between local HER IHC and RFP within the subset of HER2+ patients determined by local IHC (Table 4). Among the 83 HER2+ patients determined by local IHC, only 41 were predicted to be positive by RFP, with the rest exhibiting very low target gene *ERBB2* expression levels (Fig. 4). To examine whether this discrepancy has clinical implication, we correlated local IHC and RFP with disease-free survival in the USO 01062 study (Fig. 5). We defined HER2+ patients as those being HER2 positive regardless of ER or PR status, HR+ patients as those being HER2 negative and either ER or PR positive, and TNBC patients as those being HER2, ER, and PR negative. Significant survival disadvantages was observed for HER2+ patients compared to the HR+ or TNBC patients by RFP subtyping (log-rank $P = 0.02$ vs. log-rank $P = 0.40$), but not by local IHC,

Table 3 Performance of multivariate prediction methods for the training set

	Accuracy	Specificity	Sensitivity	Genes picked (number of times)
HER2				
RF	0.98	1(140/140)	0.83(15/18)	
RF-KNN	0.99	1(140/140)	0.94(17/18)	ERBB2(10) GRB7(1)
PAM	0.94	1(140/140)	0.5(9/18)	
ER				
RF	0.95	0.98(53/54)	0.93(97/104)	
RF-KNN	0.95	0.96(52/54)	0.94(98/104)	ESR1(10) GATA3(8) TFF1(4) FOXA1(3) SCUBE2(3) PGR(2) LYN(1) VAV3(1)
PAM	0.95	0.94(51/54)	0.95(99/104)	
PR				
RF	0.91	0.92(49/53)	0.90(95/105)	
RF-KNN	0.91	0.91(48/53)	0.91(96/105)	ESR1(10) GATA3(7) PGR(7) FOXA1(6) SCUBE2(6) TFF1(6) IGF1R(3) BCL2(2) BUB1(2) XBP1(2) CTSL2(1) ERBB3(1) IRS1(1)
PAM	0.91	0.91(48/53)	0.90(95/105)	
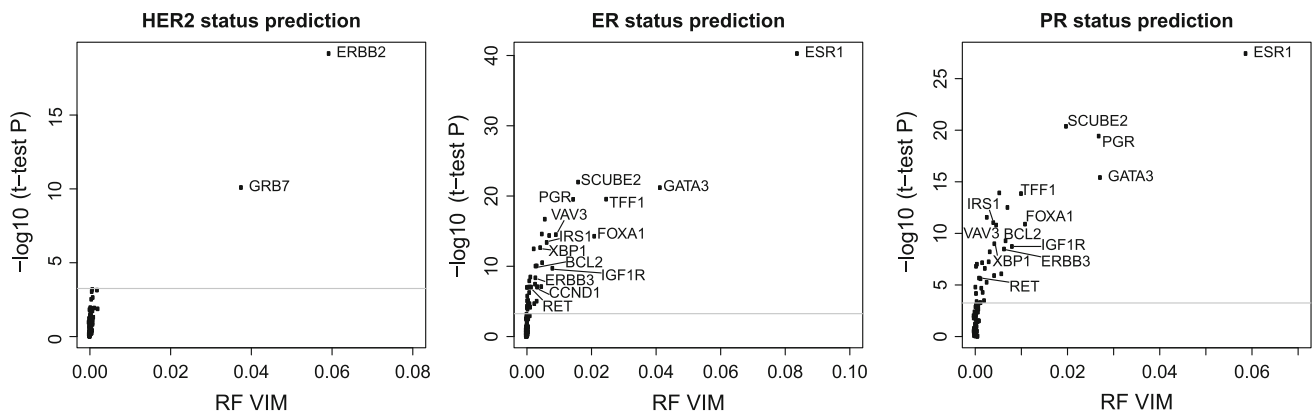
Fig. 3 Multivariate variable importance measures (VIM) by RF for HER2, ER, and PR prediction. *Y* axes are –log10 based *P* values of the two group *t*-test between central IHC positive and negative groups, and (Bonferroni) adjusted *P* value 0.05 is marked with *gray lines*. Genes with two sample *t*-test adjusted *P* values ≤0.05 and fold change ≥2 were marked with *gene symbols*

suggesting that some of the 83 HER2+ patients by local IHC were likely misclassified. Second, when the algorithm was applied to a commercially procured sample set of 136 HR+ breast cancers with locally determined ER status [16], 129 were determined to be HR+ with an overall accuracy of 95 % (Supplemental Fig. 4 and Supplemental Table 2). Among the seven samples that were predicted as ER and PR negatives, two were a primary and metastatic pair, exhibiting high *ERBB2* expression. Finally, in a commercially procured sample set of 37 locally assessed triple negative breast cancers, 34 were determined to be triple negative with an overall accuracy of 92 % (Supplemental Table 2). These results suggest that multiplexed RNA assays can be leveraged to validate local IHC status, obviating the need to centrally confirm IHC receptor status for patients enrolled onto studies.

Discussion

In our current study, we demonstrated that multiplexed PCR-based methods can accurately predict ER, PR, and HER2 status in breast cancer patients when coupled with a RF-based approach that takes into consideration multiple genes that are associated with the unique biology of breast cancer. Most notably, for predicting PR status, incorporation of PR target or co-regulated genes such as *GATA3*, *ESR1*, and *FOXA1* adds predictive benefit when compared to *PGR* alone. In the USO 01062 training set that incorporated central IHC determination, the accuracy of assigning breast cancer subtypes was much superior compared to local testing. In our training set of 24 HER2+ samples locally assessed from the USO 01062 trial, 16 were positive by central staining. This large discordance,

Table 4 HER2, ER and PR status by local IHC and RFP for test set 1, an additional set of USO 01062 study samples

HER2	Local IHC			ER	Local IHC			PR	Local IHC		
	−	+			−	+			−	+	
RFP				RFP				RFP			
−	538	42	NPV 93 %	−	249	20	NPV 93 %	−	246	21	NPV 92 %
+	7	41	PPV 85 %	+	28	337	PPV 92 %	+	94	273	PPV 74 %
	Spec 99 %	Sens 49 %	Acc 92 %		Spec 90 %	Sens 94 %	Acc 92 %		Spec 72 %	Sens 93 %	Acc 82 %

Spec specificity, *Sens* sensitivity, *PPV* positive predictive value, *NPV* negative predictive value, *Acc* accuracy
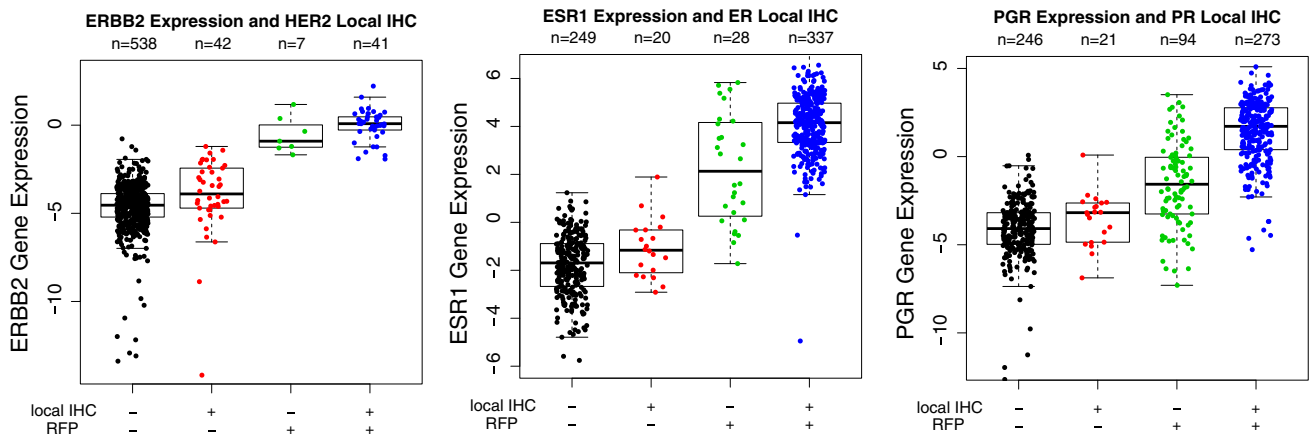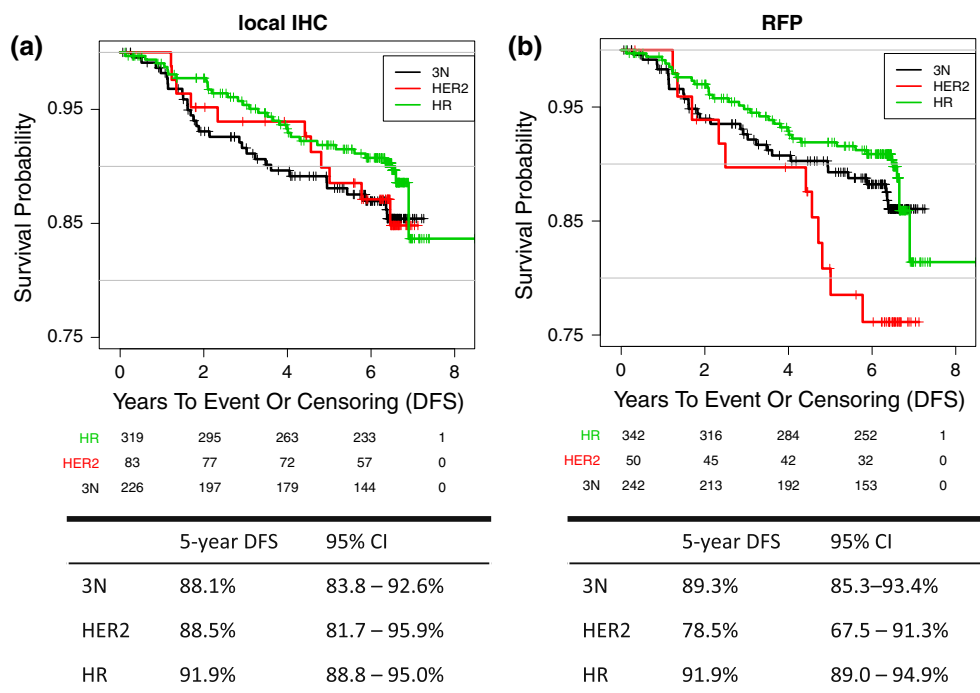

Fig. 4 Boxplots of *ERBB2*, *ESR1*, and *PGR* mRNA by local IHC and RFP results for test set 1, an additional set of USO 01062 study samples with local IHC status. *Black points* are local IHC and RFP negatives, *red points* are local IHC positives and RFP negatives, *green points* represent local IHC negatives and RFP positives, and *blue points* represent local IHC and RFP positives

Fig. 5 Kaplan–Meier curves showing disease-free survival for disease subtypes by local IHC (**a**) and RFP (**b**)

Local IHC (a):

HR	319	295	263	233	1
HER2	83	77	72	57	0
3N	226	197	179	144	0

	5-year DFS	95% CI
3N	88.1%	83.8 – 92.6%
HER2	88.5%	81.7 – 95.9%
HR	91.9%	88.8 – 95.0%

RFP (b):

HR	342	316	284	252	1
HER2	50	45	42	32	0
3N	242	213	192	153	0

	5-year DFS	95% CI
3N	89.3%	85.3–93.4%
HER2	78.5%	67.5 – 91.3%
HR	91.9%	89.0 – 94.9%

even utilizing similar methodologies, may perhaps explain the lack of concordance between HER2 IHC local determination and our HER2 RFP algorithm in the analysis of the larger test set. Similarly, a high degree of discordance between HER2 status between local and central/reference sites has been shown in independent studies, ranging from 13 to 26.6 % [13, 14]. Of note, we found significant survival differences in the HER2 positive group between those patients classified by IHC or RFP, an observation that was not observed in the HR+ or TNBC groups. This poorer disease-free survival observed with the HER2 RFP is more consistent with the literature suggesting a negative prognostic impact of HER2 status prior to the approval of trastuzumab-containing regimens, with only approximately 30 % of the HER2 positive patients receiving trastuzumab-based therapies following completion of the experimental adjuvant therapy [6]. A similar gene expression analysis was carried out by Prat et al. in HER2 IHC positive breast cancer patients treated with a trastuzumab-containing regimen, in which the HER2-enriched group, as defined by PAM50, derived a significant survival benefit compared to the group defined as non-HER2-enriched [26].

Molecular subtyping using PAM50 is not an effective surrogate for IHC status since it identifies distinct subtypes (luminal A, luminal B, basal-like, and HER2-enriched) with different prognostic implications [9]. In a similar study assessing *ESR1*, *PGR*, and *ERBB2* gene expression using quantitative RT-PCR, Du, and colleagues determined that single gene expression had a concordance rate of 87 % within the luminal subtype, 75 % within the HER2+ subtype, and 48 % within the triple negative subtype [17]. Of the 52 % discordant cases in the triple negative subtype, 12 out of the 17 (71 %) cases were determined to be luminal by RNA expression. The authors demonstrate the prognostic differences between the three IHC subtypes, with the luminal subgroup being the more favorable group followed by the triple negative and then HER2+ group. Interestingly, by RT-PCR subtyping, the HER2+ and triple negative subgroups had reversed prognostic trend, although, this may have been driven by the small HER2+ patient population in this sample set, comprising of only 9 cases. In a second study, Haibe-Kains et al., suggested that the simplest three gene classification model, SCMGENE, which comprised *ESR1*, *PGR*, and *AURKA* was largely concordant with other more complex gene expression models in terms of subtype prediction [27]. In addition, the authors show that their SCMGENE model demonstrated similar prognostic implications when compared to more established models such as PAM50, MAMMAPRINT, and ONCO-TYPE gene expression signatures, and may be adequate for clinical management of patients. Comparing RT-PCR

methods to IHC, Bastien et al. demonstrated an AUC, sensitivity, and specificity for ER to be 0.90, 0.96, and 0.74, respectively; PR to be 0.90, 0.84, and 0.85, respectively; and HER2 to be 0.95, 0.94, and 0.85, respectively [11]. When the authors compared the prognostic differences between the different methodologies, the ER+/*ESR1*+ and ER−/*ESR1*+ subgroups had similar clinical outcomes, and performed better than patients who were ER−/*ESR1*− or ER+/*ESR1*−, suggesting that RT-PCR assignment of *ESR1* status was more prognostic and accurate than IHC for ER [11]. This observation may be driven by the ability of RT-PCR technologies to accurately detect low percentage ER+ cells within the tumor population. As such, tumors that have ER IHC expression 1–9 % have a worse prognosis compared to tumors that have an ER IHC expression of 10 % or greater [28], which may, in part, explain the results of Bastien and colleagues. Interestingly, we observed notable differences in the RFP for PR predicted genes based on the Allred cutoff. Using an Allred cutoff of 3, *ESR1* was the gene that best predicted PR IHC status. However, with the cutoff extended to 4, *PGR* became the most predictive gene. These results suggest that extending the Allred score to 4 for a positive PR result may more accurately predict PR status, an observation that warrants further clinical investigation.

In conclusion, central confirmation of breast cancer receptor status is critical for the interpretation of clinical data from experimental trials. Our study suggests that local and central testing show substantial overall concordance, but that a PCR-based classifier offers a reasonable strategy to identify the set of samples that may have been erroneously classified and subject them to further testing. Such an approach has added valued in that it conserves 3–5 slides that would be used for repeat IHC testing, and also yields expression data on up to 90 breast cancer-related genes. As tissue can be limited in patients who have gone through multiple lines of therapy in the metastatic setting, multiplexed assays provide an attractive method of obtaining valuable biomarker data for association with clinical outcomes. We proposed two RF-based approaches, RF and RF-KNN. While both approaches performed similarly, the former classifier requires the usage of the same 96 genes for test samples, and the latter, through implementing a preliminary feature selection step, relies on a smaller set of genes for prediction, providing flexibility in panel development. In the current report, we demonstrate the utility of PCR-based multiplexed assays to accurately confirm ER, PR, and HER2 status and obviate the need for central IHC confirmation of subtype, while obtaining additional biomarker data on a diverse set of breast cancer-related genes.

Materials and methods

Tissue collections

Formalin-fixed paraffin-embedded (FFPE) tumor samples were obtained from ($n = 158 + 630$) breast cancer patients as part of a completed phase III study ("A Randomized, Open-Label, Multicenter, Phase Ill Trial Comparing Regimens of Adriamycin plus Cytoxan Followed by Either Taxotere or Taxotere plus Xeloda as Adjuvant Therapy for Female Patients with High-Risk Breast Cancer") (http://clinicaltrials.gov/show/NCT00089479). Tissue samples were collected and analyzed following approval by the US Oncology, Inc. Institutional Review Board and appropriate confirmation of written informed consent. ER, PR, and HER2 status was determined by local testing.

Breast cancer tumor blocks were procured for 173 breast cancer tumors. Tissue samples were obtained from Cureline, Inc (South San Francisco, CA) following approval of the Ethics Committee of Saint Petersburg City Clinical Oncology Hospital and appropriate confirmation of written informed consent. Tissue samples were also obtained from The MT Group (Van Nuys, CA) following IRB approval (http://www.sterlingirb.com). The IRB waived the need for written informed consent per FDA guidelines, as this was a retrospective study with anonymized patient data. ER, PR, and HER2 status was determined by local testing.

Gene expression

Hematoxylin–eosin sections were prepared for all samples and were reviewed by a pathologist to confirm diagnosis and assess tumor content. RNA extraction and gene expression analysis were performed as previously described [16]. Briefly, FFPE sections were macrodissected to enrich for neoplastic tissue followed by RNA extraction using the High Pure FFPE RNA Micro Kit (Roche Applied Sciences, Indianapolis, IN). RNA was then subjected to a one-step cDNA synthesis/preamplification reaction using the Invitrogen Platinum Taq/Reverse Transcriptase enzyme mix and pooled TaqMan® Gene Expression Assays (Life Technologies, Carlsbad, CA). Quantitative PCR (qPCR) was then conducted on Fluidigm 96.96 Dynamic Arrays using the BioMark™ HD system (Fluidigm Corporation, South San Francisco, CA). Cycle threshold (Ct) values were normalized and converted to relative expression values (negative delta Ct) by subtracting the median gene expression estimated using all 96 genes on the array.

Immunohistochemical staining

A total of 158 samples from the phase III trial (24 HER2+ , 43 ER−/PR−/HER2−, and 91 ER and/or PR+/HER2− based on local testing) were chosen at random and centrally confirmed for ER, PR, and HER2. Antibodies for ER (SP1), PR (1E2), and HER2 (4B5) were obtained from Ventana Medical Systems, Inc., (Tucson, AZ). IHC was performed using the Discovery XT (ER, PR) or Bench-Mark XT (HER2) systems (Ventana). All samples were scored by a single pathologist (E.F.). Standard guidelines for HER2 protein overexpression assessment were used; samples were scored as 0, 1, 2, or 3, where a score of 0 and 1 are considered negative for HER2 protein overexpression, 2 is weakly positive and equivocal, and 3 is strongly positive. For ER and PR assessment, the Allred score was calculated by adding the proportion score (PS) with the intensity score (IS). The PS scores are as follows: $0 =$ no staining, $1 \geq 0–1$ %, $2 \geq 1–10$ %, $3 \geq 10–33$ %, $4 \geq 33–67$ %, $5 \geq 67–100$ % cells stained. IS scores are as follows: $0 =$ no intensity, $1 =$ weak, $2 =$ intermediate, and $3 =$ strong. Interpretation is positive or negative based on total score (PS + IS) where total score of 0, 1 or 2 is negative and a score of 3 or greater is positive.

Target gene prediction algorithm

A two-component Gaussian mixture model was applied to the empirical *ERBB2, ESR1*, and *PGR* expression data, and model fitting made recourse to the R library *mclust* [29]. Posterior conditional probabilities of the component membership for each sample were computed using the fitted parameters, and a threshold of ≥ 0.5 was used to classify a sample as having positive expression of the receptor.

Multivariate prediction algorithm

We applied and contrasted the following multivariate classification algorithms to predict HER2, ER, and PR IHC positivity separately. Both PAM and RF are often used for high-dimensional data settings where the number of variables exceeds the number of observations.

PAM

The nearest shrunken centroid method (also called PAM [21]) was applied to normalized gene expression of 158 samples from the USO 01062 trial with central IHC status. Prediction performance was valuated using 10-fold cross validation. The application of the PAM algorithm utilized the R library *pamr*.

Random Forests

The RF classification was performed using the *random-Forest* library in R. Forests were created with 10,000 trees and default settings. Advantages of the RF method include its ability to handle a large number of variables, provide variable importance measures (VIM), and produce accurate and unbiased estimates of prediction performance without the utilization of a separate test set. The VIM from RF we used is the mean decrease in accuracy generated by this predictor whenever it was selected for splitting, compared to when the values for the predictor was permuted.

Random Forests + KNN

The hybrid approach combining RF and KNN (K-nearest neighbors, $K = 3$) utilizes RF for variable selection and KNN for formalizing prediction based on RF-selected variables. This procedure was carried out using 10-fold cross validation. Briefly, (i) the whole dataset was randomly subdivided into 10 subsets, 9 of which were used to construct an RF predictor. (ii) Variables were ordered by VIM and the N variables that were at least 1/10 of the highest VIM were selected. (iii) To select the smallest possible set of genes for minimal misclassification error rate, out-of-bag error rates using the N RF classifiers sequentially fitted using the top M ($M = 1, …,N$) variables were obtained and contrasted. The set of variables (assuming size is P ($1 \leq P \leq N$)) variables that had lowest out-of-bag error rate were selected. (iv) A classifier was then built using KNN ($K = 3$) and the P variables, and its unbiased prediction error was obtained by predicting the left-out subset from the whole dataset. (v) This process was repeated for each left-out fold. The frequency of genes picked within each fold was recorded.

ERBB2 copy number estimation by real-time quantitative PCR

FFPE tumor DNA was prepared by QIAamp DNA FFPE Tissue Kit from macrodissected tumor sections as previous described [16]. Genomic FFPE DNA (200 ng) was then subjected to 17 cycles of preamplification using pooled gene specific primers at 50 nM each and TaqMan® Pre-amplification Master Mix (Life Technologies) according to the manufacture protocol. The preamplified samples were diluted 5 fold and qPCR was performed using Fluidigm 96.96 Dynamic Arrays on the BioMark™ system according to the manufacture instruction. In brief, sample mix contains DNA, TaqMan® Gene Expression Master Mix (Life Technologies), DNA binding sample loading reagent (Fluidigm) and EvaGreen dye (Biotium, Hayward, CA). Assay mix contains gene specific primer pairs and sample loading reagent (Fluidigm). The Ct determination and melt curve analyses were carried out by Fluidigm Gene Analysis Software. Relative gene copy numbers of ERBB2 were calculated by the Delta Delta Ct method as described previously [30]. Reference genes used for Delta Ct calculation were RPPH1, GPG15 and ZNF80 and median of Delta Ct from each gene was used as a calibrator for Delta Delta Ct calculation. Primers used were: ERBB2_1F 5′-GCAGTTACCAGTGCCAATATCC-3′ and ERBB2_1R 5′-ATCAAAGCTCTCCGGCAGAA-3′; ERBB2-_2F 5′-CTGGTCACCTACAACACAGACA-3′ and ERBB2_2R 5′-AGCTGGCGCCGAATGTATA-3′; RPPH1_1F 5′-GCCAGCGAAGTGAGTTCAA-3′ and RPPH1_1R 5′-GCGGAGGAGAGTAGTCTGAA-3′; RPPH1_2F 5′-GCCAGC GAAGTG AGTTCAA-3′ and RPPH1_2R 5′-GCGGAGGAGAGTAGTCTGAA-3′; GPR15_F 5′-CCCTTTGTTGACAT TGTG ACCTG-3′ and GPR15_R 5′-TGGTAATGGGCACACAGCTTCCTT-3′; ZNF80_F 5′-CAGCTCATCCTCACTT GGCATTGA-3′ and ZNF80_R 5′-GGCCTTCCCACATATCTCATAGAGT-3′.

References

1. Hammond ME, Hayes DF, Dowsett M, Allred DC, Hagerty KL, Badve S, Fitzgibbons PL, Francis G, Goldstein NS, Hayes M et al (2010) American Society of Clinical Oncology/College Of American Pathologists guideline recommendations for immuno-histochemical testing of estrogen and progesterone receptors in breast cancer. J Clin Oncol 28(16):2784–2795
2. Puhalla S, Bhattacharya S, Davidson NE (2012) Hormonal therapy in breast cancer: a model disease for the personalization of cancer care. Mol Oncol 6(2):222–236
3. Figueroa-Magalhaes MC, Jelovac D, Connolly RM, Wolff AC (2014) Treatment of HER2-positive breast cancer. Breast 23(2):128–136
4. Anders CK, Zagar TM, Carey LA (2013) The management of early-stage and metastatic triple-negative breast cancer: a review. Hematol Oncol Clin North Am 27(4):737–749
5. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA et al (2000) Molecular portraits of human breast tumours. Nature 406(6797):747–752

6. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS et al (2001) Gene expression patterns of breast carcinomas distinguish tumor sub-classes with clinical implications. Proc Natl Acad Sci USA 98(19):10869–10874

7. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT et al (2002) Gene expression profiling predicts clinical outcome of breast cancer. Nature 415(6871):530–536

8. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ et al (2002) A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 347(25):1999–2009

9. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, Davies S, Fauron C, He X, Hu Z et al (2009) Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol 27(8):1160–1167

10. Elloumi F, Hu Z, Li Y, Parker JS, Gulley ML, Amos KD, Troester MA (2011) Systematic bias in genomic classification due to contaminating non-neoplastic tissue in breast tumor samples. BMC Med Genomics 4:54

11. Bastien RR, Rodriguez-Lescure A, Ebbert MT, Prat A, Munarriz B, Rowe L, Miller P, Ruiz-Borrego M, Anderson D, Lyons B et al (2012) PAM50 breast cancer subtyping by RT-qPCR and concordance with standard clinical molecular markers. BMC Med Genomics 5:44

12. Carlson RW, Moench SJ, Hammond ME, Perez EA, Burstein HJ, Allred DC, Vogel CL, Goldstein LJ, Somlo G, Gradishar WJ et al (2006) HER2 testing in breast cancer: NCCN task force report and recommendations. J Natl Compr Canc Netw 4(Suppl 3):S1–S22;quiz S23–24

13. de Duenas EM, Hernandez AL, Zotano AG, Carrion RM, Lopez-Muniz JI, Novoa SA, Rodriguez AL, Fidalgo JA, Lozano JF, Gasion OB et al (2014) Prospective evaluation of the conversion rate in the receptor status between primary breast cancer and metastasis: results from the GEICAM 2009-03 ConvertHER study. Breast Cancer Res Treat 143(3):507–515

14. Orlando L, Viale G, Schiavone P, Fedele P, Nacci A, Rizzo P, Calvani N, Sponziello F, Chetri C, Mazzoni E et al. (2011) Discordance in pathology report after central pathology review in early breast cancer and its impact on treatment choice. J Clinical Oncol, 29 (suppl; abstr 585)

15. O'Shaughnessy J, Paul D, Stoke C, Pippen JJ, Blum JL, Krekow L, Holmes FA, Vukelja S, Lindquist D, Sedlacek S et al. (2010) First efficacy results of a randomized, open-label, phase III study of adjuvant doxorubicin plus cyclophosphamide, followed by docetaxel with or without capecitabine, in high-risk early breast cancer. Cancer Res, 2010;70(Suppl 24):Abstract nr S4-2

16. Schleifman EB, Desai R, Spoerke JM, Xiao Y, Wong C, Abbas I, O'Brien C, Patel R, Sumiyoshi T, Fu L et al (2014) Targeted biomarker profiling of matched primary and metastatic estrogen receptor positive breast cancers. PLoS ONE 9(2):e88401

17. Du X, Li XQ, Li L, Xu YY, Feng YM (2013) The detection of ESR1/PGR/ERBB2 mRNA levels by RT-QPCR: a better approach for subtyping breast cancer and predicting prognosis. Breast Cancer Res Treat 138(1):59–67

18. Mohsin SK, Weiss H, Havighurst T, Clark GM, Berardo M, le Roanh D, To TV, Qian Z, Love RR, Allred DC (2004) Progesterone receptor by immunohistochemistry and clinical outcome in breast cancer: a validation study. Mod Pathol 17(12):1545–1554

19. Harvey JM, Clark GM, Osborne CK, Allred DC (1999) Estrogen receptor status by immunohistochemistry is superior to the ligand-binding assay for predicting response to adjuvant endocrine therapy in breast cancer. J Clin Oncol 17(5):1474–1481

20. Breiman L (2001) Random Forests. Mach Learn 45(1):5–32

21. Tibshirani R, Hastie T, Narasimhan B, Chu G (2002) Diagnosis of multiple cancer types by shrunken centroids of gene expression. Proc Natl Acad Sci USA 99(10):6567–6572

22. Kauraniemi P, Kuukasjarvi T, Sauter G, Kallioniemi A (2003) Amplification of a 280-kilobase core region at the ERBB2 locus leads to activation of two hypothetical proteins in breast cancer. Am J Pathol 163(5):1979–1984

23. Bertucci F, Nasser V, Granjeaud S, Eisinger F, Adelaide J, Tagett R, Loriod B, Giaconia A, Benziane A, Devilard E et al (2002) Gene expression profiles of poor-prognosis primary breast cancer correlate with survival. Hum Mol Genet 11(8):863–872

24. Tozlu S, Girault I, Vacher S, Vendrell J, Andrieu C, Spyratos F, Cohen P, Lidereau R, Bieche I (2006) Identification of novel genes that co-cluster with estrogen receptor alpha in breast tumor biopsy specimens, using a large-scale real-time reverse transcription-PCR approach. Endocr Relat Cancer 13(4):1109–1120

25. Gruvberger S, Ringner M, Chen Y, Panavally S, Saal LH, Borg A, Ferno M, Peterson C, Meltzer PS (2001) Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns. Cancer Res 61(16):5979–5984

26. Prat A, Bianchini G, Thomas M, Belousov A, Cheang MC, Koehler A, Gomez P, Semiglazov V, Eiermann W, Tjulandin S et al (2014) Research-based PAM50 subtype predictor identifies higher responses and improved survival outcomes in HER2-positive breast cancer in the NOAH study. Clin Cancer Res 20(2):511–521

27. Haibe-Kains B, Desmedt C, Loi S, Culhane AC, Bontempi G, Quackenbush J, Sotiriou C (2012) A three-gene model to robustly identify breast cancer molecular subtypes. J Natl Cancer Inst 104(4):311–325

28. Iwamoto T, Booser D, Valero V, Murray JL, Koenig K, Esteva FJ, Ueno NT, Zhang J, Shi W, Qi Y et al (2012) Estrogen receptor (ER) mRNA and ER-related gene expression in breast cancers that are 1% to 10% ER-positive by immunohistochemistry. J Clin Oncol 30(7):729–734

29. Fraley C, Raftery AE (2002) Model-based clustering, discriminant analysis, and density estimation. J Am Stat Assoc 97(458):611–631

30. Spoerke JM, O'Brien C, Huw L, Koeppen H, Fridlyand J, Brachmann RK, Haverty PM, Pandita A, Mohan S, Sampath D et al (2012) Phosphoinositide 3-kinase (PI3 K) pathway alterations are associated with histologic subtypes and are predictive of sensitivity to PI3 K inhibitors in lung cancer preclinical models. Clin Cancer Res 18(24):6771–6783