A Single-Tube Quantitative Assay for mRNA Levels of Hormonal and Growth Factor Receptors in Breast Cancer Specimens

Ayuko A. Iverson,* Cheryl Gillett,† Paul Cane,‡ Christopher D. Santini,* Thomas M. Vess,* Lauren Kam-Morgan,§ Alice Wang,* Marcia Eisenberg,§ Charles M. Rowland,* Janice J. Hessling,§ Samuel E. Broder,¶ John J. Sninsky,* Andrew Tutt,† Steven Anderson,§ and Sheng-Yung P. Chang*

From Celera,* Alameda, California; the Breast Pathology Laboratory,† Gay’s Hospital, London, United Kingdom; the Histopathology Department,§ St Tom’s Hospital, London, United Kingdom; the Laboratory Corporation of America,¶ Research Triangle Park, North Carolina; Celera,§ Rockville, Maryland; the Breakthrough Breast Cancer Research Unit,¶ Gay’s Hospital, Kings College, London, United Kingdom

Knowledge of estrogen receptor (ER) and progesterone receptor (PR) status has been critical in the evolution of modern targeted therapy of breast cancer and remains essential for making informed therapeutic decisions. Recently, growth factor receptor HER2/neu (ERBB2) status has made it possible to provide another form of targeted therapy linked to the overexpression of this protein. Presently, pathologists determine the receptor status in formalin-fixed, paraffin-embedded sections using subjective, semi-quantitative immunohistochemistry (IHC) assays and quantitative fluorescence in situ hybridization for HER2. We developed a single-tube multiplex TaqMan (mERPR+HER2) assay to quantitate mRNA levels of ER, PR, HER2, and two housekeeping genes for breast cancer formalin-fixed, paraffin-embedded sections. Using data from the discovery sample sets, we evaluated IHC-status-dependent cutoff-point and IHC-status-independent clustering methods for the classification of receptor status and then validated these results with independent sample sets. Compared with IHC-status, the accuracies of the mERPR+HER2 assay with the cutoff-point classification method were 0.98 (95% CI: 0.97–1.00), 0.92 (95% CI: 0.88–0.95), and 0.97 (95% CI: 0.95–0.99) for ER, PR, and HER2, respectively, for the validation sets. Furthermore, the areas under the receiver operating-characteristic curves were 0.997 (95% CI: 0.994–1.000), 0.967 (95% CI: 0.949–0.985), and 0.968 (95% CI: 0.915–1.000) for ER, PR, and HER2, respectively. This multiplex assay provides a sensitive and reliable method to quantitate hormonal and growth factor receptors.

There has been considerable progress in the targeted therapy of breast cancer in the past few decades. Such progress has depended to a substantial degree on determining the levels of the estrogen receptor (ER), encoded by the ESR1 gene, and progesterone receptor (PR), encoded by the PGR gene, in breast cancer cells.1–8 More recently, it has been possible to use overexpression of human epidermal growth receptor type 2 gene (ERBB2 also known as HER2/neu) to guide yet another form of targeted therapy according to the specific features of individual patients with breast cancer.9–11 Indeed, tumor overexpression of HER2 has been used to select women for therapy with trastuzumab, a recombinant monoclonal antibody against this protein. Moreover, high HER2 expression may be associated with high risk of recurrence in women receiving an aromatase inhibitor or tamoxifen as adjuvant therapy.8

It is worth briefly examining tamoxifen and how this agent has shaped modern breast cancer therapy for early and advanced disease. Tamoxifen is one of the oldest and arguably best studied targeted therapies in medical oncology. ER is essentially the only known therapeutic target for tamoxifen. Thus, in the crucial setting of adjuvant therapy for early breast cancer, tamoxifen has been shown to provide clear benefit both in terms of recurrence-free survival and overall survival.2 More recent trials have also shown very important benefits regarding the use of aromatase inhibitors in adjuvant therapy for postmenopausal women with endocrine-responsive early breast cancer, particularly with respect to better disease-free survival compared with tamoxifen (but not as yet for overall survival).4,5,12 Thus, ER and PR

Celera has provided funding to Guy’s Hospital and St. Thomas Tissue and Data bank for acquisition of breast cancer specimens. In addition, LabCorp is offering the ER/PRE Gene Expression Assay.

Accepted for publication November 5, 2008.

Supplemental material for this article can be found on http://jmd.amjpathol.org.

Address reprint requests to Sheng-Yung P. Chang, Celera, 1401 Harbor Bay Parkway, Alameda, CA 94502. E-mail: sheng-yung.chang@celera.com.
status in malignant tissue from women with operable, invasive breast cancer provides critically important classification of outcome and clinical benefit for adjuvant endocrine or chemoendocrine therapies.

The status of ER-regulated PR has been reported to improve outcome prediction over ER status alone. In Caucasians, approximately 60% to 65% of breast cancer cases are ER-positive and PR-positive (ER+/PR+), 15 to 20% are ER+/PR-negative (PR−), 15% to 20% are ER−/PR−, and <5% are ER−/PR+. The response rate to tamoxifen treatment has been reported to be markedly decreased in patients with ER+/PR− breast tumors.

Starting in the mid 1970s, the levels of ER or PR in fresh or frozen breast tissue specimens was determined by quantitative, but technically challenging, radioactive, biochemical ligand-binding assays using dextran-coated charcoal. Almost two decades later, the development of monoclonal antibodies and the semiquantitative immunohistochemical (IHC) method allowed the measurement of ER and PR expression in frozen tissue or formalin-fixed, paraffin-embedded (FFPE) tissue sections. Gradually the IHC method became the method of choice as receptor expression could be related to morphology, and it was demonstrated that the prediction of endocrine therapy response was equivalent to the ligand-binding assay method. The status of hormonal receptors detected by the IHC method can be affected by many factors, such as fixation, antigen retrieval, antibody type, and subjective evaluation criteria. A variety of tools have been developed to improve the performance of IHC testing for ER and PR, including methods for both manual and image-based scoring of staining results. One example is a semiquantitative IHC interpretation system, the Allred score, developed to grade immunostained slides based on the percentage and intensity of positively stained tumor cells. This approach remains subjective, semi-quantitative, and can be labor intensive. Many groups reported poor interlaboratory reproducibility and reliability for testing of hormonal receptors and growth factor receptors, and the number of samples with IHC Allred scores for each sample set are listed in Table 1.

### TaqMan RT-PCR assays to quantify ER, PR, or HER2 mRNA level individually in archived FFPE specimens

All groups have reported good concordance between mRNA level and protein level for ER or PR. Moreover, several groups developed quantitative PCR assays for HER2 DNA amplification and RT-PCR for over-expression of HER2 mRNA in frozen or FFPE breast tumor specimens.

In the present study, we report the development and the performance of a single-tube, one-step, multiplex TaqMan RT-PCR assay (mERPR+HER2) to detect ER, PR, and HER2 mRNA levels in breast cancer FFPE specimens using two Real-Time PCR instrument platforms. In addition, we also report the concordance between hormonal receptor and growth factor receptor status determined by the IHC assay and the results of the mERPR+HER2 assay classified with two different methods using two independent sample sets.

### Materials and Methods

#### Study Subjects

Three sets of FFPE breast tumor sections were used to develop the RT-PCR assay for ER, PR, and HER2. Two contemporary sets (sample set 1 and sample set 2) were provided by Laboratory Corporation of America (LabCorp), and a third set of archived FFPE breast tumor samples (sample set 3) was provided by Guy’s and St. Thomas’ Tissue and Data bank (London, United Kingdom). The cohort of 291 subjects was diagnosed between 1975 and 2001 with tumor size <3 cm, lymph node negative and ER+ primary breast tumors, and the use of this cohort was approved by Guy’s Research Ethics Committee (04/Q0704/137). The use of these three sample sets for the development of classification methods of hormonal and growth factor receptors, and the number of samples with IHC Allred scores for each sample set are listed in Table 1.

### IHC Assays

#### Hormonal Receptors

For the IHC assay performed at LabCorp, the FFPE tissue specimens were mounted on SuperFrost Plus slides (Fisher Scientific, Hampton, NH) and dried for 30 minutes in a 60°C slide drier. An H&E-stained section was prepared for each specimen and evaluated for the presence of tumor cells. The FFPE slides were processed on

<table>
<thead>
<tr>
<th>Table 1. Description of Sample Sets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Set</td>
</tr>
<tr>
<td>Set 1*</td>
</tr>
<tr>
<td>Set 2</td>
</tr>
<tr>
<td>Set 3</td>
</tr>
</tbody>
</table>

*HER2 IHC status was not available. †ER and PR IHC Allred scores were available for 42 of 55 samples. ‡ER, PR, and HER2 IHC Allred scores were available for 291, 279, and 272 of 291 samples, respectively. A total of 400, 388, and 327 samples with ER, PR, and HER2 IHC status, respectively, were used for data analyses.
staining in <10% of the tumor cells). For sample set 3, HER2 IHC was scored according to the new ASCO-CAP guidelines.33

RNA Extraction from FFPE Sections

All FFPE section slides used for this study were 4- or 10-μm thick with ~60% to 80% breast tumor cells. The FFPE section slides were deparaffinized by soaking them in xylene for 10 minutes with occasional agitation and repeated with fresh xylene. The slides were then washed consecutively with 100% ethanol, 90% ethanol, and 70% ethanol with 2 minutes for each wash. The slides were then air dried at room temperature for 5 minutes. Fifteen microliters of Proteinase K digestion solution [2 mg/ml Proteinase K (Ambion, Austin, TX), 0.1 mol/L NaCl, 10 mmol/L Tris pH 8.0, 1 mmol/L EDTA, and 0.5% SDS], was applied to the dried tissue on the slide. The tissue was then scraped with a sterile surgical blade and transferred into a 1.5-ml tube containing 185 μl Proteinase K digestion solution, and incubated overnight at 55°C for 18 to 24 hours. After incubation, the samples were spun at 14,000 rpm for 5 minutes, and the supernatant was transferred to a new tube. A mixture of 600 μl of 100% ethanol and 400 μl of extraction buffer (5 mol/L guanidium thiocyanate, 31.25 mmol/L Na Citrate, pH 7.0, 0.625% Sarcosyl, and 0.125 mol/L β-mercaptoethanol) was added to the supernatant of each sample, loaded into Zymo-Spin II Columns (Zymo Research, Orange, CA), spun at 12,000 rpm for one minute, and repeated until the entire sample had been spun through the column. The column was washed once with 200 μl of wash buffer (80% ethanol in 10 mmol/L Tris-HCl and 0.1 mmol/L EDTA, pH 8.0), followed by 13.5 Kunitz units DNase (QIAGEN, Valencia, CA) treatment at room temperature for 30 minutes. The columns were washed with 200 μl wash buffer twice and then dried by centrifugation for 2 minutes at 12,000 rpm. The total RNA was then eluted twice with 50 μl of Tris-EDTA buffer that had been heated to 65°C.

The amount of PCR-amplifiable RNA was quantitated by one-step RT-PCR using primers for the housekeeping (HSK) gene, NUP214, and compared with a serially diluted control, Universal Human Reference RNA (Stratagene, La Jolla, CA). The recovery of amplifiable RNA depends on the age of the FFPE specimen and RNA extraction methods. The recovery of amplifiable RNA from one 4-μm breast cancer FFPE section ranges from 0.5 ng to 25 ng.

A New Approach for Determining Normalization Factor

The top two most stable HSK genes, PPIG and NUP214, were previously determined by the profiling of 138 breast cancer FFPE samples,53 and they were used to validate the novel approach of determining the normalization factor for RNA amount in each RT-PCR reaction. Fifty-eight human total RNA samples (see supplemental Table S4 at http://jmd.amjpathol.org) from various tissue types were used to demonstrate the feasibility of using two TaqMan
Table 2. Genes and Information of RT-PCR Primers and TaqMan Probes in the mERPR+HER2 Assay for the 7500 System

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Gene symbol</th>
<th>Forward primer sequence</th>
<th>Reverse primer sequence</th>
<th>Reporter</th>
<th>Probe sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>2099</td>
<td>ESR1*</td>
<td>5'CTCTACAGAAGAGGATTTG-3'</td>
<td>5'-GTCCTTTTCTTCCTCAGAGCTT-3'</td>
<td>FAM</td>
<td>5'-TTCCTTATCACTTAATCTA-3'</td>
</tr>
<tr>
<td>5241</td>
<td>PGR†</td>
<td>5'-TCGACTTTCCTGCTCAGAAGAT-3'</td>
<td>5'-CCCAGATGAAGACACGCTCA-3'</td>
<td>VIC</td>
<td>5'-TGAGCTCTGGAATTTGCTAT-3'</td>
</tr>
<tr>
<td>2064</td>
<td>ERBB2‡</td>
<td>5'-CAGTCGACTGTTCACATTCA-3'</td>
<td>5'-GGAAGGCTGACATCCCGACAT-3'</td>
<td>VIC</td>
<td>5'-TCAGGAAATTCGGCGCTC-3'</td>
</tr>
<tr>
<td>8021</td>
<td>NUP214§</td>
<td>5'-CACTTCTTCTTATGACCCACGT-3'</td>
<td>5'-CAAAGGAAATTCGGCGCCTT-3'</td>
<td>VIC</td>
<td>5'-ATGTTCAAGTTTTGGCTT-3'</td>
</tr>
<tr>
<td>9360</td>
<td>PPIG</td>
<td>5'-GACCGAGAAGAAGGATA-3'</td>
<td>5'-AGGATCGTGGCTTCTTTTAT-3'</td>
<td>VIC</td>
<td>5'-ATGTTCAAGTTTTGGCTT-3'</td>
</tr>
</tbody>
</table>

*ESR1 and PGR have at least four alternative splice variants. NM_0001215, AF258449, AF258450, and AF258451 are the accession numbers of the variants for ESR1. NM_009296, AB056833, AB058444, and AB058445, are the accession numbers of the variants for PGR. NM_004448 and NM_01055862 are the annotated accession numbers of the variants for ERBB2. NM_005085 and NM_004792 are the accession numbers of the variants for NUP214 and PPIG, respectively. For each of these genes, RT-PCR primers were designed to amplify a region shared by all listed splice variants. The amplicon sizes are 104-bp, 80-bp, 95-bp, 123-bp, and 61-bp, for ESR1, PGR, ERBB2, NUP214, and PPIG, respectively. †TRE and PHO labeled probes were provided by Applied Biosystems. §All five TaqMan probes have minor-groove-binding and non-fluorescent quencher at 3' termini.

probes labeled with identical fluorescent reporter dye (see supplemental Table S6 at http://jmd.amipathol.org) to determine the normalization factor of total RNA input amount in each sample. The concentration of each RNA sample was determined using RiboGreen quantitation assay (Invitrogen, Carlsbad, CA), and 20 ng of total RNA was used for each reaction. The expression levels of two HSK genes, NUP214 or PPIG, were quantitated in independent simplex reactions using either NUP214 probe or PPIG probe labeled with the same fluorescent reporter dye using the 7900 Real-Time PCR System (7900 system; Applied Biosystems, Foster City, CA). The average of NUP214 and PPIG expression levels was then compared with the composite NUP214 and PPIG expression level quantitated using both NUP214 and PPIG TaqMan probes in a single reaction.

Single-Tube, One-Step, Multiplex TaqMan Assays

mERPR+HER2 RT-PCR Assay on the 7500 System

Table 2 lists gene IDs, gene symbols, the oligonucleotide sequences, fluorescent reporters, and quenchers of all TaqMan probes for the 7500 Real-Time PCR System (7900 system; Applied Biosystems, Foster City, CA).

Quantitative detection of mRNA levels of ESR1, PGR, ERBB2 (HER2), and the two HSK genes, NUP214 and PPIG, and the oligonucleotide sequences and fluorescent reporters of all TaqMan probes for the 7500 Real-Time PCR System (7500 system; Applied Biosystems, Foster City, CA).

Quantitative detection of mRNA levels of ESR1, PGR, ERBB2 (HER2), and two HSK genes in a single tube was accomplished through one-step multiplex TaqMan RT-PCR with a 384-well plate using the 7900 system. Each 15-μL reaction contained 50 mmol/L of Tricine, 115 mmol/L KOAc (pH 8.0), 4.5 mmol/L Mn(OAc)₂, 9.6% glycerol, 400 μmol/L dATP, 400 μmol/L dGTP, 400 μmol/L dCTP, 800 μmol/L dUTP, 1% dimethyl sulfoxide, 0.3 μmol/L 6-ROX (Invitrogen, Carlsbad, CA) in 5% Tween-20, 0.12 μmol/L enhancer (Abbott), 0.08 unit/μL uracil N-glycosylase, 0.4 unit/μL Z50 DNA polymerase (Abbott), 500 nmol/L of each primer, 200 nmol/L TET-labeled (or NED-labeled) TaqMan probe for each HSK gene, 250 nmol/L FAM-labeled TaqMan probe for ER, 250 nmol/L VIC-labeled TaqMan probe for PR, and approximately 0.5 to 1 ng of amplifiable RNA extracted from FFPE specimen. The thermocycling parameters for the 7900 system are as follows: 50°C for 2 minutes; 95°C for 1 minute; 95°C for 15 seconds and 58°C for 35 seconds for 42 cycles for the 7500 system. In addition to each RNA sample from the FFPE specimen, 25 ng of the Universal Human Reference RNA was included as the control in each amplification plate, and all samples were run in duplicate reactions.

mERPR RT-PCR Assay on the 7900 System

A single-tube multiplex TaqMan assay for ER, PR, and two HSKs (mERPR assay) was developed for the 7900 system. The mERPR+HER2 assay for the 7900 system was not developed due to the unavailability of a compatible fluorescent dye for HER2 for the optical system on the 7900 system. The supplemental Table S5 at http://jmd.amipathol.org lists the oligonucleotide sequences, orientations, fluorescent reporters, and quenchers of all TaqMan probes for the 7900 system.

Quantitative detection of mRNA levels of ER, PR, and two HSK genes in a single tube was also accomplished through one-step multiplex TaqMan RT-PCR with a 384-well plate using the 7900 system. Each 15-μL reaction contained 50 mmol/L of Tricine, 115 mmol/L KOAc (pH 8.0), 4.5 mmol/L Mn(OAc)₂, 9.6% glycerol, 400 μmol/L dATP, 400 μmol/L dGTP, 400 μmol/L dCTP, 800 μmol/L dUTP, 1% dimethyl sulfoxide, 0.3 μmol/L 6-ROX (Invitrogen, Carlsbad, CA) in 5% Tween-20, 0.12 μmol/L enhancer (Abbott), 0.08 unit/μL uracil N-glycosylase, 0.4 unit/μL Z50 DNA polymerase (Abbott), 500 nmol/L of each primer, 200 nmol/L TET-labeled (or NED-labeled) TaqMan probe for each HSK gene, 250 nmol/L FAM-labeled TaqMan probe for ER, 250 nmol/L VIC-labeled TaqMan probe for PR, and approximately 0.5 to 1 ng of amplifiable RNA extracted from FFPE specimen. The thermocycling parameters for the 7900 system are as follows: 50°C for 2 minutes; 95°C for 1 minute; 60°C for 30 minutes; 95°C for 15 seconds and 58°C for 35 seconds for 42 cycles for the 7500 system. In addition to each RNA sample from FFPE specimens, 25 ng of the Universal Human Reference RNA (Stratagene) was included as the control in each amplification plate. All samples on the plate were run in duplicate. The results from the 7900 system are described in Supplemental Tables S6 to S10 and Figures S1A, S1B, S2A, and S2B at http://jmd.amipathol.org.
FFPE Section-to-Section Reproducibility

To determine FFPE section-to-section reproducibility, five sequential sections from each of 10 breast cancer tumor FFPE samples (BioChain Institute, Hayward, CA) were obtained. Before RNA was isolated, the slide was checked to ensure that all sections from each sample were identical in size and shape. Total RNA was extracted from these 50 sections and the recovery was determined using NanoDrop (Thermo Scientific, Wilmington, DE). The amplifiable RNA was determined by a TaqMan RT-PCR assay for the HSK gene, NUP214. ER, PR, and HER2 mRNA levels in each section were determined using the mERPR+HER2 assay.

Data Analysis

The ER, PR, and HER2 mRNA expression levels in each FFPE clinical sample were calculated using the \( \Delta \Delta C_T \) method.\(^{54}\) First, the average \( C_T \) (cycle threshold) of duplicate reactions of each gene of interest was calculated for each sample and the control sample, Universal Human Reference RNA. Then the ER, PR, and HER2 mRNA expression levels were normalized with the HSK gene expression level for each FFPE and the control sample. Finally, the HSK-normalized ER, PR, and HER2 expression levels in each FFPE sample were further compared with the HSK-normalized ER, PR, and HER2 expression levels in the control sample, respectively. Therefore, the relative expression level of each gene of interest in each FFPE sample, is presented as \( \Delta \Delta C_T = (\Delta C_T_{\text{sample}} (C_T \text{ of gene of interest} - C_T \text{ of HSK genes}) - \Delta C_T_{\text{control}} (C_T \text{ of gene of interest} - C_T \text{ of HSK genes})) \). A minus one factor is included to graphically illustrate higher expression above lower expression. When \( C_T \) value was not reported, then a \( C_T \) of 42 was used for the calculation of \( \Delta \Delta C_T \).

Statistical Analysis

For ER and PR classification, the results of the mERPR+HER2 assay from sample set 1 and combined sample sets 2 and 3 were used as the discovery and validation sets, respectively. For HER2 classification, the results of the mERPR+HER2 assay from sample sets 2 and 3 were used as the discovery and validation sets, respectively.

Area under the receiver operating characteristic curve (AUC) measures the ability of the assay to discriminate between positive and negative status of ER, PR, or normal- and over-expression status of HER2 across the entire range of \( \Delta \Delta C_T \) values. AUC was computed based on the receiver operating characteristic function available from the Mayo Clinic (http://mayoresearch.mayo.edu/mayo/research/biostat/splustestsfunctions.cfm) and confidence intervals (CI) for the AUC were calculated using the variance estimate described by Delong et al.\(^{55}\)

We used two different methods to classify the status of ER, PR, and HER2. An IHC-status-dependent \( \Delta \Delta C_T \) cutoff-point method was used to determine the hormonal and growth factor receptor status. Using IHC status as the gold standard, an Allred score \( \geq 3 \) defines positive hormonal status (ER+ or PR+).\(^{23}\) and an intensity score of HER2 3+ defines HER2 overexpression.\(^{33}\) The \( \Delta \Delta C_T \) cutoff point for classification of each marker was empirically selected based on the diagnostic metrics of sensitivity, specificity, positive predictive value, negative predictive value, and accuracy from the comparisons with IHC status using various \( \Delta \Delta C_T \) cutoff points. A \( \Delta \Delta C_T \) cutoff point for classification of each marker was selected using the data from their respective discovery sets. The selected \( \Delta \Delta C_T \) cutoff points were then applied to classify ER, PR and HER2 status of samples in their respective validation sets.

An IHC-status-independent classification method was established by developing Gaussian mixture models as implemented in MCLUST software for the R programming language (http://www.R-project.org)\(^{56}\) based solely on ER \( \Delta \Delta C_T \), PR \( \Delta \Delta C_T \), and HER2 \( \Delta \Delta C_T \) measurements of subjects in their respective discovery sets.\(^{57-61}\) The Bayesian Information Criterion was used to determine the best fitting model. For ER and HER2 measures, the best model was a mixture of two Gaussian distributions with equal variance. For PR, since the best model by Bayesian Information Criterion was a single Gaussian distribution, which would not be helpful for classification purposes, a mixture model of two Gaussian distributions with equal variance was used. The mixture models estimated from the discovery data were then used to classify an independent set of validation subjects to the cluster for which they had the highest probability of membership based on their \( \Delta \Delta C_T \) measurements.

The diagnostic metrics of sensitivity, specificity, positive predictive value, negative predictive value, and accuracy were calculated for both discovery and validation sets. The agreement coefficient, Cohen’s kappa,\(^{62}\) was used to evaluate the agreement between the IHC status and the status determined using the results from the mERPR+HER2 assay for the \( \Delta \Delta C_T \) cutoff-point and clustering methods. In addition, the square of Pearson’s correlation coefficient was used to assess the degree of correlation between two instrument platforms.

Results

A New Approach for Determination of Normalization Factor

To obtain more accurate normalization of RNA input amount and to accommodate three genes of interest, ESR1, PGR, and ERBB2, in a multiplex TaqMan assay with four different fluorescent reporters, we designed a novel approach of determining the expression levels of two HSK genes using two TaqMan probes labeled with the same fluorescent reporter.

Two HSK genes, NUP214 and PPIG, expressed at relatively constant levels in breast tumor FFPE specimens were selected to validate our approach. mRNA levels of NUP214 and PPIG were averaged from independent reactions with NUP214 or PPIG probes, and compared with the NUP214 and PPIG composite mRNA level in a single co-amplification reaction. The comparison of 58 total RNA samples from various tissues using two amplification for-
mats is shown in Figure 1. The two different formats of determining HSK gene expression levels correlated well with a correlation coefficient, $r^2$, of 0.9742 ($P < 0.0001$).

**FFPE Section-to-Section Reproducibility**

Total RNA and amplifiable RNA from each of five sequential sections of 10 breast cancer tumor FFPE samples were determined by absorbance at 260 nm and the TaqMan RT-PCR assay for the housekeeping gene NUP214. The recoveries of amplifiable RNA from each of 5 sections of 10 FFPE samples are shown in Figure 2. The average percentage of the amplifiable RNA relative to the total RNA measured by absorbance at 260 nmol/L of each sample is shown above each bar. The average amplifiable RNA from 10 FFPE samples varied from 70 ng (S4) to 1300 ng (S1). Figure 3 illustrates the distributions of ER, PR, and HER2 $\Delta C_T$ values from 5 sections of 10 FFPE samples, respectively. The relatively larger variations of the PR $\Delta C_T$ values in samples S2, S4, and S8 were due to later $C_T$ resulting from lower PR expression levels. There was no correlation between the variation of amplifiable RNA recovery and ER, PR, or HER2 $\Delta C_T$ values.

**Classification of Hormonal Receptor Status**

Three breast cancer tumor FFPE sample sets with available ER and PR IHC Allred total scores listed in Table 1 were used to determine the classifications of ER and PR status. Sample set 1 with 67 samples, and combined sample sets 2 and 3 with 333 samples were used as the discovery and validation sets, respectively. Both ER mRNA and PR mRNA were detected in all clinical specimens using the mERPR+HER2 assay.

---

**Estrogen Receptor**

The ER $\Delta C_T$ values of 67 RNA samples of the discovery set using the mERPR+HER2 assay were calculated, and the distribution of ER $\Delta C_T$ values in the discovery set was bimodal, as reported previously. The AUC of ER $\Delta C_T$ values from the discovery set was 0.989 (95% CI: 0.972–1.000). The performance measurements of sensitivity, specificity, positive predictive value, negative predictive value, and accuracy for the ER classification based on the IHC ER status were compared using various $\Delta C_T$ cutoff points (Figure 4A). A $\Delta C_T$ cutoff point of 1.5 with 94% accuracy was empirically selected to divide 67 ER $\Delta C_T$-values into two groups in Figure 4B. The distribution of 67 IHC ER Allred scores and the classifications of ER status by both the IHC-status-dependent $\Delta C_T$ cutoff-point and the IHC-status-independent clustering methods are listed in Table 3. Two Allred TS0 samples and two Allred TS3 samples were classified as ER+ and ER−, respectively, by the $\Delta C_T$ cutoff-point method. All Allred TS0 samples were classified as ER− correctly, and two Allred TS3 samples were classified as ER− by the clustering method. When compared with IHC ER status, the kappa coefficient of the clustering method, 0.924 (95% CI: 0.821–1.000) was higher than that of the $\Delta C_T$ cutoff-point method, 0.842 (95% CI: 0.693–0.992) (Table 4).

Both the $\Delta C_T$ cutoff point of 1.5 and the model parameters for the clustering method derived from the discovery set were applied to classify the ER status of samples in the validation set. The validation set consisted of two independent subsets, sample set 2 and sample set 3 listed in Table 1. Forty-two samples with ER IHC Allred scores in sample set 2 and 291 samples with ER IHC Allred scores in sample set 3 were used to validate ER classification. The 291 archived specimens in sample set 3 were originally identified as ER+ between 1975 and 2001. The ER and PR status was re-evaluated in these specimens with contemporary IHC.
assays, and 8 of 291 samples (3%) were re-classified as IHC ER−. The AUC of ER \( \Delta \Delta C_T \) values from the validation set was 0.997 (95% CI: 0.994–1.000). The distribution of IHC Allred total scores of the entire 333 samples and the classifications of ER status by both the IHC-status-dependent \( \Delta \Delta C_T \) cutoff-point and the IHC-status-independent clustering methods of the validation set are listed in Table 3. One Allred TS0 sample and four Allred TS3 samples were classified as ER− and ER+ respectively, by the \( \Delta \Delta C_T \) cutoff-point method. All IHC ER− samples were correctly classified as ER− by the clustering method. However, an additional six Allred TS4 to TS6, and one Allred TS8 samples were classified as ER− by the clustering method. When compared with IHC ER status, the kappa coefficient of the clustering method was 0.759 (95% CI: 0.623–0.895), lower than 0.870 (95% CI: 0.758–0.982) of the \( \Delta \Delta C_T \) cutoff-point method (Table 4).

Figure 3. Reproducibility of the status of hormonal and growth factor receptors in sequential FFPE sections. A, B, and C: The \( \Delta \Delta C_T \) distributions of ER, PR, and HER2, respectively, of 5 sequential sections from 10 FFPE breast tumor specimens.

Figure 4. ER classification. A: Performance measurements of classifications based on ER IHC status using various \( \Delta \Delta C_T \) cutoff points of the discovery sample set. B: Classification of the discovery sample set using the \( \Delta \Delta C_T \) cutoff point of 1.5. C: Distribution of ER \( \Delta \Delta C_T \) values with IHC status of the discovery and validation sample sets. PPV: positive predictive value; NPV: negative predictive value.
The performance measurements of the PR classification of 67 \( \Delta \Delta C_T \) values based on the IHC PR status were compared using various \( \Delta \Delta C_T \) cutoff points (Figure 5A). The AUC of PR \( \Delta \Delta C_T \) values from the discovery set was 0.987 (95% CI: 0.969–1.000). A \( \Delta \Delta C_T \) cutoff point of 0.5 with 94% accuracy was empirically selected to divide 67 PR \( \Delta \Delta C_T \) values into two groups in Figure 5B. The distribution of 67 IHC PR Allred total scores and the classifications of PR status by both the IHC-status-dependent \( \Delta \Delta C_T \) cutoff-point and the IHC-status-independent clustering methods are listed in Table 5. Twelve (11 Allred TS0 and one Allred TS2) were classified as PR+ by both the \( \Delta \Delta C_T \) cutoff-point method and the clustering method, respectively. Fourteen Allred TS3 and TS4 samples were classified as PR− by the \( \Delta \Delta C_T \) cutoff-point method, and additional six samples (four Allred TS3, one Allred TS5 and one Allred TS6) were also classified as PR− by the clustering method. When compared with IHC PR status, the kappa coefficients of the \( \Delta \Delta C_T \) cutoff-point and clustering methods were similar but lower than those of the discovery set, 0.664 (95% CI: 0.544–0.784) and 0.669 (95% CI: 0.556–0.782), respectively (Table 6).

### Classification of ER Status of the Discovery and Validation Sets Using the \( \Delta \Delta C_T \) Cutoff-Point and Clustering Methods

<table>
<thead>
<tr>
<th>Allred TS*</th>
<th>( \Delta \Delta C_T ) Cutoff-Point</th>
<th>Clustering</th>
<th>Allred TS*</th>
<th>( \Delta \Delta C_T ) Cutoff-Point</th>
<th>Clustering</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ER+</td>
<td>ER−</td>
<td>ER+</td>
<td>ER−</td>
<td>IHC (% of total)</td>
</tr>
<tr>
<td>0</td>
<td>17</td>
<td>2</td>
<td>15</td>
<td>0</td>
<td>17</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ER− ¹</td>
<td>17 (25%)</td>
<td>2</td>
<td>15</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>12</td>
<td>12</td>
<td>0</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>31</td>
<td>31</td>
<td>0</td>
<td>31</td>
<td>0</td>
</tr>
<tr>
<td>ER+ ²</td>
<td>50 (75%)</td>
<td>48</td>
<td>2</td>
<td>48</td>
<td>2</td>
</tr>
</tbody>
</table>

*Allred total score. †Total number of specimens with Allred TS0 and TS2 in each set. ‡Total number of specimens with Allred TS3 to TS8 in each set.

### Classification of Overexpression of Growth Factor Receptor HER2

The HER2 \( \Delta \Delta C_T \) values of 55 samples of the HER2 discovery set (sample set 2 in Table 1) using the mERPR+HER2 assay were determined. The AUC of HER2 \( \Delta \Delta C_T \) values from the discovery set was 0.968 (95% CI: 0.924–1.000). The HER2 \( \Delta \Delta C_T \) values were compared with HER2 IHC scores with HER2 IHC 3+(HER2-over) defined as samples expressing above the normal level of HER2 (HER2-norm). The performance measurements of HER2 classification based on the HER2 IHC status were compared using various HER2 \( \Delta \Delta C_T \) values from the validation set was 0.967 (95% CI: 0.949–0.985). The distribution of IHC Allred total scores and the classifications of PR status of 321 validation samples by both the \( \Delta \Delta C_T \) cutoff-point and the IHC-status-independent clustering methods are listed in Table 5. Twelve (11 Allred TS0 and one Allred TS2) and eight samples (seven Allred TS0 and one Allred TS2) were classified as PR+ by the \( \Delta \Delta C_T \) cutoff-point method and the clustering method, respectively. Fourteen Allred TS3 and TS4 samples were classified as PR− by the \( \Delta \Delta C_T \) cutoff-point method, and additional six samples (four Allred TS3, one Allred TS5 and one Allred TS6) were classified as PR− by the clustering method. When compared with IHC PR status, the kappa coefficients of the \( \Delta \Delta C_T \) cutoff-point and clustering methods were similar but lower than those of the discovery set, 0.664 (95% CI: 0.544–0.784) and 0.669 (95% CI: 0.556–0.782), respectively (Table 6).

### Summary of the Performance of ER Classification

<table>
<thead>
<tr>
<th>Discovery (n = 67)</th>
<th>Validation (n = 333)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Delta \Delta C_T ) Cutoff-Point</td>
<td>Clustering</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.96 (0.86–1.00)</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.88 (0.64–0.99)</td>
</tr>
<tr>
<td>PPV</td>
<td>0.96 (0.86–1.00)</td>
</tr>
<tr>
<td>NPV</td>
<td>0.88 (0.64–0.99)</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.94 (0.85–0.98)</td>
</tr>
<tr>
<td>Kappa</td>
<td>0.842 (0.693–0.992)</td>
</tr>
</tbody>
</table>
cutoff points (Figure 6A). A \( \Delta \Delta C_T \) cutoff point of 3.5 with 91% accuracy was empirically selected to divide 55 HER2 \( \Delta \Delta C_T \) values into two groups in Figure 6B. The distribution of HER2 IHC scores and the classification of HER2 status by both \( \Delta \Delta C_T \) cutoff-point and clustering methods of the discovery set are listed in Table 7. Using a \( \Delta \Delta C_T \) cutoff point of 3.5 for the classification of HER2 expression status, one HER2 IHC 2 sample were classified as HER2-over, and four samples with HER2 IHC 3 were classified as HER2-norm. Using the clustering method, all 38 samples with HER2 IHC 0 to 2 were classified correctly. Nine of 17 samples with HER2 IHC 3 were classified as HER2-norm. When compared with IHC HER2 expression status, the kappa coefficients of the \( \Delta \Delta C_T \) cutoff-point and clustering methods for classification of HER2 expression status of the discovery set were 0.776 (95% CI: 0.592–0.961) and 0.551 (95% CI: 0.312–0.791), respectively (Table 8).

Both the \( \Delta \Delta C_T \) cutoff point of 3.5 and the model parameters for the clustering method derived from the discovery set were applied to classify HER2 expression status of 272 samples in the validation set. The AUC of HER2 \( \Delta \Delta C_T \) values from the validation set was 0.968 (95% CI: 0.915–1.000). The distribution of 272 HER2 IHC scores and the classification of HER2 status by both \( \Delta \Delta C_T \) cutoff-point and clustering methods of the validation set are listed in Table 7. Using the \( \Delta \Delta C_T \) cutoff point of 3.5, four samples (two HER2 IHC 0 and

Table 5. Classification of PR Status of the Discovery and Validation Sets Using the \( \Delta \Delta C_T \) Cutoff-Point and Clustering Methods

<table>
<thead>
<tr>
<th>Allred TS*</th>
<th>IHC (% of total)</th>
<th>( \Delta \Delta C_T ) cutoff-point</th>
<th>Clustering</th>
<th>IHC (% of total)</th>
<th>( \Delta \Delta C_T ) cutoff-point</th>
<th>Clustering</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PR+</td>
<td>PR−</td>
<td>PR+</td>
<td>PR−</td>
<td>PR+</td>
<td>PR−</td>
</tr>
<tr>
<td>0</td>
<td>20</td>
<td>1</td>
<td>1</td>
<td>19</td>
<td>1</td>
<td>19</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>PR−</td>
<td>0 (30%)</td>
<td>1</td>
<td>1</td>
<td>19</td>
<td>1</td>
<td>19</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>11</td>
<td>9</td>
<td>2</td>
<td>8</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>11</td>
<td>11</td>
<td>10</td>
<td>1</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>0</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>0</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>PR−</td>
<td>47 (70%)</td>
<td>44</td>
<td>41</td>
<td>6</td>
<td>277 (86%)</td>
<td>263</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>14</td>
</tr>
</tbody>
</table>

*Allred total score. †Total number of specimens with Allred TS0 and TS2 in each set. ‡Total number of specimens with Allred TS3 to TS8 in each set.
two HER2 1+) were classified as HER2-over, and three HER2 IHC 3+ samples was classified as HER2-norm. Using the clustering method, all 255 HER2-norm samples were classified correctly, but 12 of 17 HER2 IHC 3+ samples were classified as HER2-norm. When compared with IHC HER2 expression status, the kappa coefficients of the $\Delta\Delta C_T$ cutoff-point and clustering methods for classification of HER2 overexpression of the validation set were 0.786 (95% CI: 0.633–0.940) and 0.439 (95% CI: 0.182–0.696), respectively (Table 8).

Diagnostic Metrics of mERPR+HER2 Assay

The performance measurements of the mERPR+HER2 assay, sensitivity, specificity, positive predictive value, negative predictive value, accuracy, and kappa coefficient, for ER, PR, and HER2 overexpression with the discovery and validation sets are listed in Tables 4, 6, and 8, respectively.

All $\Delta\Delta C_T$ values from the discovery and validation sets were sorted, and then plotted using $\Delta\Delta C_T$ of 1.5, 0.5, and 3.5 as the cutoff points for ER, PR, and HER2, respectively, and compared with IHC ER, PR, and HER2 status in Figures 4C, 5C, and 6C, respectively.

Discussion

Several factors, such as the size of the tissue biopsy, the type of the fixative, the age of the paraffin block, or the degree of chemical modification, affect the recovery of amplifiable RNA from FFPE sections. Insufficient amounts of recovered RNA from FFPE sections for profiling of a panel of genes for breast cancer studies have been reported. Therefore, we set out to develop a multiplex TaqMan assay to quantitate mRNA levels of $ESR1$, $PGR$, $ERBB2$ (HER2), and two HSK genes in a single tube for breast cancer diagnosis. We evaluated the performance of the mERPR+HER2 assay with three sets of

<table>
<thead>
<tr>
<th></th>
<th>Discovery (n = 67)</th>
<th>Validation (n = 321)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\Delta\Delta C_T$ cutoff-point</td>
<td>Clustering</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.94 (0.82–0.99)</td>
<td>0.87 (0.74–0.95)</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.95 (0.75–1.00)</td>
<td>0.95 (0.75–1.00)</td>
</tr>
<tr>
<td>PPV</td>
<td>0.98 (0.88–1.00)</td>
<td>0.98 (0.87–1.00)</td>
</tr>
<tr>
<td>NPV</td>
<td>0.86 (0.65–0.97)</td>
<td>0.76 (0.55–0.91)</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.94 (0.85–0.98)</td>
<td>0.90 (0.80–0.96)</td>
</tr>
<tr>
<td>Kappa</td>
<td>0.861 (0.730–0.993)</td>
<td>0.767 (0.607–0.928)</td>
</tr>
</tbody>
</table>

Table 6. Summary of the Performance of PR Classification
breast cancer number specimens with HER2 IHC scores 0, 1+, and 2+. †The number of specimens with HER2 IHC score 3+.

A recent review discussed the current issues in ER and HER2 testing for breast cancer by IHC.31 In addition, the intratumoral heterogeneity and the various degree of HER2 DNA amplification in breast tumor specimens using IHC and fluorescence in situ hybridization assays have been reported.65,66 Therefore, it is important to demonstrate the reproducibility of sequential FFPE sections using the mERPR+HER2 assay. Our results in Figure 3 demonstrated good reproducibility for samples with ER+, PR+, or HER2-over status, and better than that of the group of ER−, PR−, or HER2-norm, respectively, because of the later Ct values resulting from the relatively low abundance of mRNA levels.

The lack of intermediate Allred scores in the ER discovery sample set (only two Allred TS3 and no Allred TS2 or TS4 samples) rendered the ΔΔCt cutoff-point selection more challenging; therefore the more conservative lower ΔΔCt cutoff point of 1.5 was selected. Approximately two thirds of breast cancer has ER+ status, however sample set 3 of the validation sample set in our study was mostly ER+ (97%). Consequently, the percentage of samples with HER2 overexpression (HER2 IHC 3+) in this set was also lower than the generally observed 25% to 30% with HER2 overexpression.17 The kappa coefficients of ER classification using the ΔΔCt cutoff-point method for the discovery and validation sets were similar, 0.842 and 0.870, respectively (Table 4). In contrast, the kappa coefficient of ER classification using the clustering method dropped from 0.924 to 0.759 for the validation set (Table 4). The discordant results between the IHC ER assay and the mERPR+HER2 assay were 9 (2%) and 13 (3%) of a total of 400 samples using the ΔΔCt cutoff-point method (Figure 4C) and the clustering method, respectively.

The ER mRNA expression in breast tumor specimens is bimodal as represented by the sigmoidal transition between RT-PCR–and RT-PCR+ groups (see Figure 4C). Both IHC ER–/PCR ER+ and IHC ER+/PCR ER− groups were identified by IHC methods with different antibodies used by the two clinical sites. Therefore, it is likely that the performance of the different antibodies was similar even though the SP1 clone used by Guy’s Hospital has been indicated to have higher affinity and a more robust performance.31,67 The group of IHC ER− but PCR ER+ subjects, currently not being identified, may merit consideration of endocrine therapy, but clinical data do not exist on the benefit of endocrine therapy in this situation.

The total number of IHC ER− samples, 36 of 400 (9%) in the discovery and validation sets was relatively small, therefore the cutoff point of 1.5 requires additional validation with a larger number of samples even though the kappa coefficients of the agreement of ER status between the IHC assay and the mERPR+HER2 assay with the ΔΔCt cutoff-point method was “almost perfect”68 based on the interpretation of Cohen’s kappa for both discovery and validation sets. The agreement of ER status between the IHC assay and the mERPR+HER2 assay with the ΔΔCt cutoff-point method was slightly higher than those reported by Cronin et al63 (kappa = 0.825; n = 62) and Ma et al64 (kappa = 0.83; n = 852). Subsequently, two additional groups reported the agreement of ER status between the IHC assay and the ER TaqMan assay in the Oncotype DX as kappa = 0.81 (n = 149)63 and kappa = 1.0 (n = 80).64

In contrast to ER mRNA expression, PR mRNA expression is more continuous as represented by the gradual

| Table 7. Classification of HER2 Overexpression of the Discovery and Validation Sets Using the ΔΔCt Cutoff-Point and Clustering Methods |
|----------------|----------------|----------------|----------------|----------------|----------------|
| HER2 IHC score | ΔΔCt cutoff-point | Clustering | HER2 IHC score | ΔΔCt cutoff-point | Clustering |
| IHC (% of total) | HER2-over | HER2-norm | HER2-over | HER2-norm | IHC (% of total) | HER2-over | HER2-norm |
| 0 | 10 | 0 | 0 | 0 | 0 | 0 | 200 |
| 1+ | 20 | 0 | 0 | 0 | 0 | 0 | 53 |
| 2+ | 8 | 1 | 7 | 0 | 0 | 0 | 2 |
| HER2-norm* | 38 (69%) | 38 (69%) | 38 (69%) | 38 (69%) | 255 (94%) | 255 (94%) |
| 3+ | 17 | 13 | 4 | 8 | 9 | 17 |
| HER2-over* | 17 (31%) | 17 (31%) | 17 (31%) | 17 (31%) | 51 |

*Total number of specimens with HER2 IHC scores 0, 1+, and 2+. †The number of specimens with HER2 IHC score 3+.

| Table 8. Summary of the Performance of HER2 Classification |
|----------------|----------------|----------------|----------------|----------------|
|                | ΔΔCt cutoff-point | Clustering | ΔΔCt cutoff-point | Clustering |
| Sensitivity    | 0.76 (0.50–0.93) | 0.53 (0.28–0.77) | 0.53 (0.28–0.77) | 0.71 (0.44–0.90) |
| Specificity    | 0.97 (0.86–1.00) | 1.00 (0.91–1.00) | 1.00 (0.91–1.00) | 1.00 (0.99–1.00) |
| PPV            | 0.93 (0.66–1.00) | 1.00 (0.63–1.00) | 1.00 (0.63–1.00) | 1.00 (0.48–1.00) |
| NPV            | 0.90 (0.77–0.97) | 0.81 (0.67–0.91) | 0.99 (0.97–1.00) | 0.96 (0.92–0.98) |
| Accuracy       | 0.91 (0.80–0.97) | 0.84 (0.71–0.92) | 0.97 (0.95–0.99) | 0.96 (0.92–0.98) |
| Kappa          | 0.776 (0.592–0.961) | 0.551 (0.312–0.791) | 0.786 (0.633–0.940) | 0.439 (0.182–0.696) |
increase of $\Delta \Delta C_T$ values from the RT-PCR – group to the RT-PCR+ group in Figure 5C. The kappa coefficients of PR status between the IHC assay and the mERPR+HER2 assay dropped from the discovery to validation set using both $\Delta \Delta C_T$ cutoff-point and the clustering methods (Table 6). When compared with ER discordant results, the percentage of samples with discordant results between PR IHC assay and the mERPR+HER2 assay were larger, 30 (8%) and 25 (6%) of a total of 388 samples using $\Delta \Delta C_T$ cutoff-point method (Figure 5C) and the clustering method, respectively, which is likely due to the more continuous values for expression of PR. Therefore, we will continue to evaluate the performance of both classification methods for PR status. The agreement of PR status between the IHC assay and the mERPR+HER2 assay with the $\Delta \Delta C_T$ cutoff-point method for the validation set, was similar to those reported by Cronin et al.63 (kappa = 0.674; n = 62) and Ma et al.64 (kappa = 0.70; n = 852). However, subsequently two groups reported lower agreement for PR status, kappa of 0.48 (n = 149)65 and kappa of 0.57 (n = 80), using the PR TaqMan assay in the Oncotype DX.66

The regulation of PR expression is complex and not fully understood at present. Two estrogen-regulated promoters and several alternative splice variants have been reported.16,69,70 The differences in the agreement between various IHC and RT-PCR assays might be the result of the PR antibodies interrogating different epitopes for the IHC assays and/or different mRNA variants for RT-PCR assays. In addition, the RT-PCR assay doesn’t discern PR-A from PR-B isoforms, which are translated from different translation initiation sites. Several studies reported interesting findings of cross talk between the estrogen and growth factor HER2 signaling pathways, which resulted in decreased PR levels, and the tumors eventually becoming tamoxifen resistant.71–74

The performances of ER and PR classifications using IHC-status-dependent $\Delta \Delta C_T$ cutoff-point and IHC-status-independent clustering methods were similar (Tables 4 and 6). However, it is interesting that the performance of classification of HER2 overexpression between the IHC-status-dependent $\Delta \Delta C_T$ cutoff-point and IHC-status-independent clustering methods were quite different (Table 8). Using the clustering method, 9 of 17 (53%) and 12 of 17 (70%) samples with HER2 IHC 3+ samples were classified as HER2-norm for the discovery and validation sets, respectively. Based on the clustering results, a HER2 $\Delta \Delta C_T$ cutoff point of 5.0 instead of 3.5 could have been selected to classify HER2 status. However, the sensitivity of the HER2 classification of the discovery set would have decreased to 0.47 when compared with the “gold standard” IHC assay. Using the HER2 IHC assay, several groups have reported false-positive HER2 overexpression results, especially in the group with HER IHC 2+.65,66,75 The agreement, kappa, of HER2 overexpression status between the IHC assay and the mERPR+HER2 assay with the $\Delta \Delta C_T$ cutoff-point method for both discovery and validation sets (Table 8) were higher than kappa of 0.60 with the HER2 TaqMan assay in the Oncotype DX.65 The samples with HER2, IHC 2+ and 3+ status in our study were under represented in our HER2 validation set due to the high percentage of ER+

samples in the sample set. Therefore, we will continue to evaluate both $\Delta \Delta C_T$ cutoff-point and clustering methods for classification of HER2 overexpression, and compare the results with that of IHC and/or fluorescence in situ hybridization. Although presently the ASCO 2007 Guideline Update does not recommend testing HER2 level to determine sensitivity to endocrine therapy,32 reporting quantitative ER, PR, and HER2 expression levels might provide a better understanding of the cause of the resistance to tamoxifen and non-responsiveness to trastuzumab treatments.

In conclusion, we developed a sensitive single-tube, one-step multiplex TaqMan assay to quantitate ER, PR, and HER2 expression levels with results shown to be consistent across multiple adjacent sections from the same tumor. The classification of ER, PR, and HER2-overexpression status was evaluated with two methods and compared with IHC results. Based on the interpretation of kappa coefficients, the agreement was “almost perfect” for ER, and the agreement was “substantial” for both PR and HER2.68 This 2-hour RT-PCR assay to determine the ER, PR, and HER2 status can be easily adopted in a clinical laboratory for molecular testing of the most important predictive and prognostic markers for breast cancer.

Acknowledgments

We thank Scott Benson and Dean Tsou at Applied Biosystems for providing TaqMan probes labeled with TRE and PHO and new passive reference dye, NPR. We thank Chris Sigua, Henry Shu, Wontae Kim, and Cindy Christopherson at Celera, Eddie Kallam at LabCorp, and Robert Springall at Guy’s and St. Thomas’ Breast Tissue and Data Bank for their technical supports to this study. We thank Tom White and Tom Lenk for providing comments on the manuscript.

References


6. Viale G, Regan MM, Maier C, Mastropasqua MG, Dell’Orto P,


38. Tong D, Schneeberger C, Czerwenka K, Schmutzler RK, Speiser P, Kucera E, Concini N, Kubista E, Leodolter S, Zeillinger R: Messenger RNA determination of estrogen receptor, progesterone receptor, pS2,
and plasminogen activator, inhibitor-1 by competitive reverse transcrip-
39. Iwao K, Niyoshi Y, Egawa C, Ikeda N, Tsukamoto F, Noguchi S: Quan-
titative analysis of estrogen receptor-alpha and -beta messen-
50. Kostopoulou E, Vageli D, Kaisaridou D, Nakou M, Neticsa M, Vladsic N, Daponte A, Koukoulis G: Comparative evaluation of non-informa-
tive HER-2 immunoreactions (2+) in breast carcinomas with FISH. CISH and QRT-PCR. Breast 2007, 16:615–624
67. Cheang MC, Treaba DO, Speers CH, Olofsson VA, Bajdik CD, Chia SK, Goldenstein LC, Gelmon KA, Huntsman D, Gilks CB, Nielsen TO, Gown AM: Immunohistochemical detection using the new rabbit monoclo-
al antibody SP1 of estrogen receptor in breast cancer is superior to mouse monoclonal antibody TDS in predicting survival. J Clin Oncol 2006, 24:5637–5644
69. Kastner P, Kurtz A, Turcotte B, Stropp U, Tora L, Gronemeier H, Chambon P: Two distinct estrogen-regulated promoters generate transcripts encoding the two functionally different human progester-
one receptor forms A and B. EMBO J 1993, 9:1603–1614
73. Kim HJ, Cui X, Hilsenbeck SG, Lee AV: Progesterone receptor loss correlates with human epidermal growth factor receptor 2 overex-
74. Ciocca DR, Gago FE, Fanelli MA, Calderwood SK: Co-expression of steroid receptors (estrogen receptor alpha and/or progesterone re-
75. Tubbs RR, Pettay JJ, Hoche PC, Stoler MH, Jenkins RB, Grogan TM: Discrepancies in clinical laboratory testing of eligibility for trastu-