

Agreement in Breast Cancer Classification between Microarray and Quantitative Reverse Transcription PCR from Fresh-Frozen and Formalin-Fixed, Paraffin-Embedded Tissues

MICHAEL MULLINS,¹ LAURENT PERREARD,² JOHN F. QUACKENBUSH,¹ NICHOLAS GAUTHIER,¹ STEVEN BAYER,¹ MATTHEW ELLIS,³ JOEL PARKER,⁴ CHARLES M. PEROU,⁵ ANIKO SZABO,⁶ and PHILIP S. BERNARD^{1,2*}

Background: Microarray studies have identified different molecular subtypes of breast cancer with prognostic significance. To transition these classifications into the clinical laboratory, we have developed a real-time quantitative reverse transcription (qRT)-PCR assay to diagnose the biological subtypes of breast cancer from fresh-frozen (FF) and formalin-fixed, paraffin-embedded (FFPE) tissues.

Methods: We used microarray data from 124 breast samples as a training set for classifying tumors into 4 previously defined molecular subtypes: Luminal, HER2⁺/ER⁻, Basal-like, and Normal-like. We used the training set data in 2 different centroid-based algorithms to predict sample class on 35 breast tumors (test set) procured as FF and FFPE tissues (70 samples). We classified samples on the basis of large and minimized gene sets. We used the minimized gene set in a real-time qRT-PCR assay to predict sample subtype from the FF and FFPE tissues. We evaluated primer set performance

between procurement methods by use of several measures of agreement.

Results: The centroid-based algorithms were in complete agreement in classification from FFPE tissues by use of qRT-PCR and the minimized “intrinsic” gene set (40 classifiers). There was 94% (33 of 35) concordance between the diagnostic algorithms when comparing subtype classification from FF tissue by use of microarray (large and minimized gene set) and qRT-PCR data. We found that the ratio of the diagonal SD to the dynamic range was the best method for assessing agreement on a gene-by-gene basis.

Conclusions: Centroid-based algorithms are robust classifiers for breast cancer subtype assignment across platforms and procurement conditions.

© 2007 American Association for Clinical Chemistry

Expression-based classifications are important for determining risk of relapse and making treatment decisions in breast cancer (1–4). Classifications are often developed using microarray data and then further validated on the same or different platforms by use of minimized gene sets. For instance, van't Veer et al. (4) and van de Vijver et al. (5) used microarray data in training and test sets to validate a 70-gene signature that predicts relapse in early-stage estrogen receptor (ER)⁷-positive and ER-negative tumors. In addition, Paik et al. (2) developed a 16-gene classifier that predicts relapse in ER-positive

¹ Department of Pathology, University of Utah School of Medicine, Salt Lake City, UT.

² The ARUP Institute for Clinical and Experimental Pathology, Salt Lake City, UT.

³ Siteman Cancer Center, Washington University, St. Louis, MO.

⁴ Constella Group, Durham, NC.

⁵ Departments of Genetics and Pathology and Laboratory Sciences, Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC.

⁶ Department of Oncological Sciences, Huntsman Cancer Institute, Salt Lake City, UT.

* Address correspondence to this author at: University of Utah, Department of Pathology, Salt Lake City, UT 84112. Fax ; e-mail phil.bernard@hci.utah.edu.

Received December 5, 2006; accepted May 3, 2007.

Previously published online at DOI: 10.1373/clinchem.2006.083725

⁷ Nonstandard abbreviations: ER, estrogen receptor; qRT, quantitative reverse transcription; FFPE, formalin-fixed, paraffin-embedded; PAM, prediction analysis of microarray; SSP, single sample predictor; FF, fresh-frozen; Cy, cyanine; *m*, diagonal bias; *d*, diagonal spread; *dsd*, diagonal SD; *r_d*, diagonal correlation; *ccc*, concordance correlation coefficient; *DWD*, distance weighted discrimination; *IHC*, immunohistochemistry.

tumors by use of quantitative reverse transcription (qRT)-PCR on formalin-fixed, paraffin-embedded (FFPE) tissues. Furthermore, Perou et al. (3) and Sorlie et al. (6) showed that hierarchical clustering of microarray data separates breast tumors into different “biological” subtypes (Luminal, HER2⁺/ER⁻, Basal-like, and Normal-like) and that these subtypes are prognostic. The biological classification has been validated on multiple patient cohorts by use of cross-platform microarray analyses and qRT-PCR (7–9).

Although there are few genes in common between those used to determine the biological subtypes and those used in other classifications for breast cancer prognosis, the different tests identify similar properties that predict tumor behavior (1). The classification for biological subtypes is based on hierarchical clustering, a major difference between it and other classifications for breast cancer. The unsupervised nature of hierarchical clustering is effective for discovery (10), but it is not suitable for predicting a new sample’s class since dendrogram associations can change when new data are introduced. However, it is possible to classify samples within the framework of hierarchical clustering by centroid-based methods (7, 11–13). For instance, Tibshirani et al. (13) showed that the nearest shrunken centroid method, used in prediction analysis of microarray (PAM), can classify samples as accurately as statistical approaches such as artificial neural networks. In addition, Hu et al. (7) used another simple centroid method called single sample predictor (SSP) to classify subtypes of breast cancer.

We have shown that a minimized intrinsic gene set can be used in a qRT-PCR assay to recapitulate the microarray classification of breast cancer subtypes (8). In this study, we refined our minimal gene set by using data from Hu et al. and compared 2 centroid-based methods for our breast cancer classification across platforms (microarray and qRT-PCR) and procurement methods [fresh-frozen (FF) and FFPE]. In addition, we performed a gene-by-gene analysis of the PCR data to compare agreement between FF and FFPE tissues. Our methods have general application to developing other multigene qRT-PCR tests for cancer diagnostics.

Materials and Methods

TISSUE PROCUREMENT AND PROCESSING

All tissues and data used in this study were collected and handled in compliance with federal and institutional guidelines. Breast samples received in pathology were flash-frozen in liquid nitrogen and stored at -80°C . We procured samples at the University of North Carolina at Chapel Hill, Thomas Jefferson University, University of Chicago, and University of Utah. The 159 breast samples included a 124-sample microarray training set and a 35-sample test set profiled by microarray and real-time qRT-PCR (FF and FFPE). Total RNA from FF samples was isolated using the RNeasy Midi Kit (Qiagen) and treated on-column with DNase I to eliminate contaminating

DNA. The RNA was stored at -80°C until used for cDNA synthesis.

We compared each FF sample in the test set to the clinical FFPE tissue block. We used a hematoxylin and eosin-stained slide to confirm the presence of $>50\%$ tumor and prepared 20- μm cuts with a microtome. Tissue blocks were 1 to 5 years old (i.e., early-age FFPE). The FFPE cut was deparaffinized in Hemo-De (Scientific Safety Solvents) and washed with 100% ethanol. We isolated total RNA by use of the High Pure RNA Paraffin Kit (Roche Molecular Biochemicals). We followed manufacturer’s instructions for RNA extraction except that the reagents were increased 2-fold for the 1st proteinase K digestion. Samples were treated with TURBO DNA-free (Ambion, #1906) and stored at -80°C until cDNA synthesis.

FIRST-STRAND cDNA SYNTHESIS

We performed cDNA synthesis for each sample in 40- μL total volume reactions containing 600 ng total RNA. Total RNA was first mixed with 2 μL gene-specific mixture containing 55 primers (each antisense primer at 1 $\mu\text{mol/L}$) and 2 μL of 10 mmol/L dNTP Mix (10 mmol/L each dATP, dGTP, dCTP, dTTP at pH 7). Reagents were heated at 65°C for 5 min in a PTC-100 Thermal Cycler (MJ Research) and briefly centrifuged. We added the following reagents to each tube: 8 μL of 5 \times First-Strand Buffer [250 mmol/L Tris-HCl (pH 8.3 at room temperature), 375 mmol/L KCl, 15 mmol/L MgCl₂], 2 μL of 0.1 mol/L dithiothreitol, 2 μL RNase Out (Invitrogen), and 2 μL SuperScript III polymerase (200 units/ μL). The reaction was thoroughly mixed by pipetting and incubated at 55°C for 45 min followed by 15 min at 70°C for enzyme inactivation. After cDNA synthesis, samples were purified with the QIAquick PCR Purification Kit (Qiagen). We adjusted the samples to a final concentration of 1.25 mg/L cDNA with Tris-EDTA (10 mmol/L Tris-HCl, pH 8.0, 0.1 mmol/L EDTA).

PRIMER DESIGN AND OPTIMIZATION

We designed primers with Roche LightCycler Probe Design Software 2.0. We obtained reference gene sequences through National Center for Biotechnology Information LocusLink and found optimal primer sites with the aid of Evidence Viewer (<http://www.ncbi.nlm.nih.gov>). We selected primer sets to avoid known insertions/deletions and mismatches while including all isoforms possible. Amplicons were limited to 60 to 100 bp in length because of the degraded condition of the FFPE mRNA. When possible, RNA-specific amplicons were localized between exons spanning large introns (>1 kb). Finally, we used National Center for Biotechnology Information BLAST to verify gene target specificity of each primer set. Primer sequences are presented in Table 1 in the Data Supplement that accompanies the online version of this article at <http://www.clinchem.org/content/vol53/issue7>. Primers were synthesized by Operon, resuspended in Tris-

EDTA to a final concentration of 60 $\mu\text{mol/L}$, and stored at -80°C . We assessed each new FFPE primer set for performance through qRT-PCR runs with 3 serial 10-fold dilutions of reference cDNA in duplicate and 2 no-template control reactions. Primers were verified for use when they fulfilled the following criteria: (a) target crossing point <30 in 10 ng reference cDNA; (b) PCR efficiency >1.75 ; (c) no primer-dimers in presence of template as determined through postamplification melting curve analysis; and (d) no primer-dimers in negative template control before cycle 40.

REAL-TIME qRT-PCR

We carried out PCR amplification on the Roche LightCycler 2.0. Each reaction contained 2 μL cDNA (2.5 ng) and 18 μL PCR master mix with the following final concentrations of reagents: 1 unit Platinum Taq, 50 mmol/L Tris-HCl (pH 9.1), 1.6 mmol/L $(\text{NH}_4)_2\text{SO}_4$, 400 mg/L BSA, 4 mmol/L MgCl_2 , 0.2 mmol/L dATP, 0.2 mmol/L dCTP, 0.2 mmol/L dGTP, 0.6 mmol/L dUTP, 1:40 000 dilution of SYBR Green I dye (Molecular Probes), and 0.4 $\mu\text{mol/L}$ of both forward and reverse primers for the selected target. The PCR was done with an initial denaturation step at 94°C for 90 s and then 50 cycles of denaturation (94°C , 3 s), annealing (58°C , 6 s), and extension (72°C , 6 s). Fluorescence acquisition (530 nm) was taken once each cycle at the end of the extension phase. After PCR, we initiated a postamplification melting curve program by heating to 94°C for 15 s, cooling to 58°C for 15 s, and slowly increasing the temperature (0.1 $^\circ\text{C/s}$) to 95°C while continuously measuring fluorescence.

Each PCR run contained a no-template control, a calibrator reference in triplicate, and each sample in duplicate. The calibrator reference sample comprised 3 breast cancer cell lines (MCF7, SKBR3, and ME16C2) and Stratagene Universal Human Reference RNA (Stratagene) represented in equal parts. The target crossing point, defined as the cycle at which the fluorescence of a sample rises above the background, was automatically calculated for each reaction by Roche LightCycler Software 4.0. For relative quantification, we imported an external efficiency curve ($\text{Eff} = 1.89$) and set the calibrator at 10 ng for each gene. To correct for differences in sample quality and cDNA input, we adjusted copy numbers to the arithmetic mean of 5 housekeeper genes [*ACTB*⁸ (β -actin), *PSMC4* (proteasome 26S subunit, ATPase, 4), *PUM1* (pumilio homolog 1, *Drosophila*), *MRPL19* (mitochondrial ribosomal protein L19), and *SF3A1* (splicing factor 3a, subunit 1, 120 kDa)]. Values from replicate samples were aver-

aged, and data were \log_2 transformed. Raw copy numbers (i.e., not housekeeper-adjusted) for all genes analyzed are provided in Table 2 in the online Data Supplement.

MICROARRAY

We analyzed all samples by use of DNA microarray (Agilent Human A1, Agilent Human A2, and Agilent custom oligonucleotide microarrays). We labeled and hybridized RNA for microarray analysis with the Agilent low RNA input linear amplification reagent set (<http://www.chem.agilent.com/Scripts/PDS.asp?IPage=10003>) as described in Hu et al. (14). Only RNA from FF tissue was used for microarray experiments. Each sample was assayed vs a common reference that was Stratagene's Human Universal Reference total RNA enriched with equal amounts of RNA from the MCF7 and ME16C cell lines. Microarray hybridizations were carried out on Agilent Human oligonucleotide microarrays by using 2 μg cyanine 3 (Cy3)-labeled reference sample and 2 μg Cy5-labeled experimental sample.

We scanned all microarrays by use of an Axon Scanner 4000B (Axon Instruments). We analyzed the image files with GenePix Pro 4.1 (Axon Instruments) and uploaded them into the UNC Microarray Database at the University of North Carolina at Chapel Hill (<https://genome.unc.edu/>), where a Lowess normalization procedure was performed to adjust the Cy3 and Cy5 channels (15). Microarray data for this study have been submitted to Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/>) under accession no. GSE6130.

CLINICAL IMMUNOHISTOCHEMISTRY AND PCR

At the time of diagnosis, samples were scored for protein expression of ER, progesterone receptor, and HER2/*neu* by use of standard operating procedures established at each institution. Nuclei staining $>10\%$ positive were considered positive for ER and progesterone receptor. Staining and scoring criteria for HER2 were according to HercepTestTM (Dako). For quantitative PCR to determine DNA copy number of the *ERBB2* (*v-erb-b2* erythroblastic leukemia viral oncogen homolog 2) gene, we used a clinical assay from ARUP Laboratories (cat. no. 00049390).

SELECTING GENES FOR REAL-TIME qRT-PCR

The real-time qRT-PCR assay consisted of 5 housekeeper genes (16), 5 proliferation genes for risk stratification of the Luminal (ER-positive) tumors (8), and 40 intrinsic genes important for distinguishing biological subtypes of breast cancer (7). We statistically selected the minimal 40 intrinsic classifiers from a larger 1393 intrinsic gene set previously reported in Hu et al. (7) by use of minimization methods described by Dudoit and Fridlyand (17). Briefly, we used a semisupervised classification method in which samples are hierarchically clustered and assigned subtypes on the basis of the sample-associated dendrogram (7, 11–13). We designated samples Luminal, HER2⁺/ER⁻, Basal-like, or Normal-like. We identified the

⁸ Human genes: *ACTB*, β -actin; *PSMC4*, proteasome 26S subunit, ATPase, 4; *PUM1*, pumilio homolog 1, *Drosophila*; *MRPL19*, mitochondrial ribosomal protein L19; *SF3A1*, splicing factor 3a, subunit 1, 120 kDa; *ERBB2*, *v-erb-b2* erythroblastic leukemia viral oncogen homolog 2; *ESR1*, estrogen receptor 1; *IGBP1*, immunoglobulin binding protein 1.

best class distinguishers according to the ratio of between-group to within-group sums of squares. We performed a 10-fold cross-validation by using a nearest centroid classifier and testing overlapping gene sets of varying sizes. We selected the smallest gene set that provided the highest class prediction accuracy compared with the classifications made by the complete microarray-based intrinsic gene set.

ASSESSING qRT-PCR AGREEMENT BETWEEN FF AND FFPE TISSUES

We analyzed 35 matched FF and FFPE samples (70 samples total) by qRT-PCR using the same primer sets. Agreement in the quantitative data was determined using diagonal bias (m), diagonal spread (d), diagonal SD (d_{sd}), diagonal correlation (r_d), and concordance correlation coefficient (ccc).

In diagonal bias, a best-fitting line parallel to the diagonal (slope = 1) is made from a plot of the qRT-PCR data (FF vs FFPE). Numerically, if (x_i, y_i) , $i = 1, \dots, n$ denote the measurement pairs, then the best-fitting line parallel to the diagonal is given by the following expression:

$$y = x + \bar{y} - \bar{x}$$

where \bar{x} and \bar{y} denote the sample means of the x and y measurements, respectively.

Then we calculate diagonal bias as:

$$m = \frac{\bar{y} - \bar{x}}{\sqrt{2}}$$

The d_{sd} was calculated as follows:

$$S_d = \sqrt{\frac{\sum_{i=1}^n d_i^2}{n-1}}$$

where d_i is the distance to the best fit line calculated as follows:

$$d_i = \frac{|(y_i - x_i) - (\bar{y} - \bar{x})|}{\sqrt{2}}$$

Let d represent the mean deviation from the best fit line calculated as:

$$d = \frac{\sum_{i=1}^n d_i}{n}$$

Diagonal correlation was used to determine the spread of points around the diagonal line:

$$r_d = \frac{2Cov(X,Y)}{Var(X) + Var(Y)}$$

This method does not provide information about the extent of deviation but allows measurements with different units to be compared. Furthermore, if we let ρ denote

the correlation coefficient and σ_X and σ_Y the respective SDs, then:

$$r_d = \rho \frac{2}{\frac{\sigma_X}{\sigma_Y} + \frac{\sigma_Y}{\sigma_X}}$$

That is, the diagonal correlation penalizes the correlation coefficient if there is a scale shift σ_X [neq] σ_Y . We measured the combined effect of the bias and scale shift by use of the ccc proposed by Lin (18):

$$ccc = \frac{2Cov(X,Y)}{Var(X) + Var(Y) + (\bar{Y} - \bar{X})^2}$$

ASSESSING AGREEMENT BETWEEN MICROARRAY AND qRT-PCR FOR CLASSIFICATION

A breast cancer subtype predictor was developed in PAM (<http://www-stat.stanford.edu/~tibs/PAM/>) and SSP (<https://genome.unc.edu/cgi-bin/sai/ssp/ssp.pl>) (13, 19). PAM and SSP are both nearest centroid classifiers that use prototype samples in the training set to develop centroids. Test samples are then assigned the class of the nearest centroid as measured by Euclidean distance. The major difference between the methods lies in how the centroids are constructed. SSP uses a simple unstandardized centroid created from a subset of genes identified during cross-validation, whereas PAM creates standardized and shrunken or “denoised” centroids. The amount of shrinkage is determined in cross-validation. We used a training set with prototype samples for Luminal (64 samples), HER2⁺/ER⁻ (23 samples), Basal-like (28 samples), and Normal-like (9 samples) subtypes. We classified an independent test set (35 matched FF and FFPE samples) by use of the large (1393 genes) and minimized (40 genes) versions of the microarray intrinsic gene set (see *Selecting Genes for Real-Time qRT-PCR*).

The qRT-PCR data from the test set were merged with the microarray data of the training set before classification by use of distance weighted discrimination (DWD), a method that adjusts for systematic biases between different platforms (20). The gold standard for classification of the training and test samples was based on FF tissue RNA and the classifications obtained when performing hierarchical clustering analysis using the 1393 gene intrinsic gene set from microarray data.

Results

ASSESSMENT OF qRT-PCR PRIMER SET PERFORMANCE BY COMPARING AGREEMENT BETWEEN FF AND FFPE TISSUES

We evaluated the dataset of 35 matched FF and FFPE tissues (70 samples) for 50 genes with the same PCR conditions. Agreement between FF and FFPE tissues was assessed for diagonal bias (m), diagonal correlation (r_d), diagonal SD (d_{sd}), and ccc . Fig. 1 shows an agreement plot for the relative quantification of the ER gene [*ESR1*

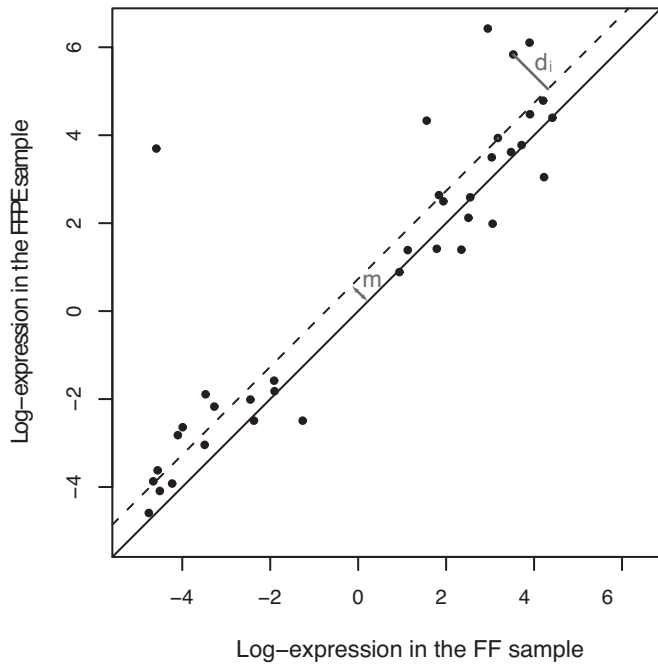


Fig. 1. Agreement plot between FF and FFPE for the ER gene. We analyzed gene expression in 35 breast tumors procured as FF and FFPE tissues, using the same conditions on the matched samples for reverse transcription and PCR. A best-fit line (dashed) is compared with the ideal line (solid), and the distance between them is the diagonal bias (m). The distance of each point to the best-fit line is represented as d_i .

(estrogen receptor 1)] between FF and FFPE tissues. This is a typical plot that was used to assess each classifier gene. In the case of *ESR1*, there is a particularly large dynamic range, and tumors are clearly divided into 2 populations. This separation highly associates with immunohistochemistry (IHC) status for ER, even without normalization (see Fig. 1 in the online Data Supplement). We have previously shown that *ESR1* alone measured from FF tissue has very high sensitivity and specificity by use of ER status by IHC as the gold standard (8).

For each gene, the agreement between FF and FFPE was analyzed using the raw data, housekeeper-normalized data, and DWD-adjusted normalized data. Scatter

plots are provided in Fig. 2 in the online Data Supplement, and values are presented in Table 3 in the online Data Supplement. The line graphs in Fig. 2 show the effects at each step of data processing. The raw (prenormalized) data show a negative bias for all genes, likely due to lower RNA quality in the FFPE tissue (Fig. 2A). Much of the bias was corrected by normalization to the housekeeper genes and DWD adjustment. As expected, DWD had a significant effect on bias (m) but did not affect other measurements of agreement (Fig. 2, B and C).

Genes with the highest diagonal correlation between FF and FFPE usually had the largest dynamic range in expression (e.g., *ESR1*, *TFF3*, *COX6C*, and *FBP1*). Housekeeper genes and other genes with low variability in expression [*IGBP1* (immunoglobulin binding protein 1)] had the lowest diagonal correlation since they form more of a cloud than a line around the diagonal. The housekeeper genes all had high agreement in terms of having low variability in expression across samples in the FF and FFPE tissues.

The ccc considers both bias and scale shift when determining agreement. The median ccc between FF and FFPE for the raw data of the 45 genes (housekeepers excluded) was 0.28. Normalization to housekeepers raised the ccc median to 0.48, and adjusting with DWD brought the median to 0.61. A comparison of the ccc value to the ratio of the dsd over the dynamic range identified many of the same primer sets as good (or poor) performers from the FFPE-derived samples.

BREAST CANCER SUBTYPE CLASSIFICATION OF TEST SET BY USE OF PAM AND SSP

Hierarchical clustering of the 124-sample training set by use of the minimized intrinsic gene set identified from Hu et al. (7) shows 4 distinct classes representing Luminal, HER2⁺/ER⁻, Basal-like, and Normal-like (see Fig. 5 in the online Data Supplement). We developed centroid classifiers from the microarray expression data by use of PAM and SSP (7, 13, 19). We made class predictions on the test set by use of microarray (large and minimized “intrinsic” sets) and qRT-PCR data (see Table 4 in the online Data

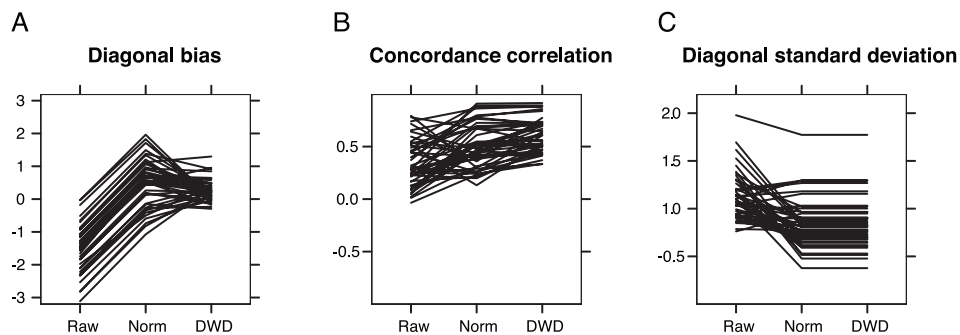


Fig. 2. Line graphs showing the effects of data processing across different methods of assessing agreement between FF and FFPE tissues. The raw data (Raw), housekeeper-normalized data (Norm), and DWD-adjusted normalized data are shown for diagonal bias (A), concordance correlation (B), and diagonal SD (C). The raw (prenormalized) data show a negative bias for all genes, likely because of lower RNA quality in the FFPE tissue. Much of the bias is corrected by normalization to the housekeeper genes and DWD adjustment. Although DWD had a significant effect on bias, it did not affect the other measurements of agreement.

Supplement). Each individual microarray (large and minimized) and PCR dataset was DWD merged with the training set before subtype class prediction.

Agreement in classification between large and minimized microarray gene sets. Of 35 samples, 33 (94%) were classified the same between PAM and SSP when using the large intrinsic microarray dataset for classification. In both discrepant cases, IHC data agreed with the PAM classification. There was the same agreement (94%) when performing the analysis with the minimized version of the microarray data. Interestingly, there was 1 sample that was called HER2⁺/ER⁻ by both PAM and SSP when using the large microarray dataset but called Basal-like by both methods when using the minimized microarray dataset. Additional analysis of this sample by quantitative PCR showed no DNA amplification of HER2/ERBB2 amplicon.

Agreement in classification between FF and FFPE. By qRT-PCR, there was 97% (34 of 35) concordance between FF and FFPE using PAM and 91% (32 of 35) concordance using SSP. There was 94% (33 of 35) concordance between the diagnostic algorithms from FF tissue and complete agreement in classification from FFPE tissue. Because the FFPE samples were obtained from the clinical block, it is likely that there was a higher tumor percentage in those samples than in the matched FF sample, which could affect the agreement. Indeed, 2 of the 3 discrepancies in classification made by SSP were when the FF tissue sample was called Normal-like (microarray and PCR) and the FFPE sample was called Luminal (PCR). These samples were ER positive by IHC and likely Luminal. The only discrepancy in PAM was in a sample classified as Normal-like from FF tissue and Luminal from FFPE.

Overall concordance across methods. Overall, PAM diagnosed 33 of 35 samples (94%) the same across microarray and qRT-PCR, whereas SSP diagnosed 30 of 35 samples (86%) the same across platforms and procurement methods. Discrepancies were of several types, including Luminal tumors classified as Normal-like, HER2⁺/ER⁻ tumors classified as Luminal, and Basal-like tumors classified as HER2⁺/ER⁻.

Discussion

Translating large-scale microarray studies into clinical tests requires several critical steps, including gene set minimization, cross-platform validation, and development of robust classification algorithms.

Several centroid-based algorithms have been developed for predicting sample subtypes from microarray data (13, 17, 19, 21). Programs that are simple and intuitive in design, such as linear discriminant analysis, are preferred owing to their transparency (19). PAM adds a feature selection to linear discriminant analysis in which *t*-statistics are computed for each gene to determine its

contribution to the assigned subtypes (13). The *t*-statistics for each gene are then ranked, and the gene set can be minimized by selecting the top genes that provide a minimal false discovery value. The main difference between SSP and PAM is that PAM shrinks the centroid toward the overall mean for classification. Here we directly compared PAM with SSP by use of the large microarray dataset applied in Hu et al. (7) and also a minimized version. On this dataset, PAM performed slightly better than SSP for classification across gene sets and conditions, although both methods performed well.

Determining agreement between methods is a complex issue. Cronin et al. (22) used Pearson correlation to show that the genes with the highest correlation in microarray maintained their association with qRT-PCR. They used short amplicons and control housekeeper genes in the qRT-PCR assay to correct biases between FF and FFPE tissues. Although correlation provides information about the linearity and slope (positive or negative correlation) of the data, it does not indicate the amount of bias, scale shift, or data spread. These additional measurements are helpful in determining whether the discrepancies in the data can be compensated for experimentally (e.g., housekeeper genes) or by use of software algorithms.

We found that the most useful analyses for assessing PCR primer set performance across FF and FFPE tissues were the ccc, the diagonal SD, and the dynamic range. Genes with a large dynamic range often had high correlation and were good classifiers across conditions, even with relatively large diagonal SDs. Although genes with a small dynamic range can be good classifiers, the measurement may not be as reproducible if there is a large amount of variation. Thus, we found that the best assessment of a classifier was using a ratio of the diagonal SD to the dynamic range.

Translating an assay from microarray to qRT-PCR provides a 2nd level of gene validation and allows the test to be used on archived FFPE tissue blocks from clinical trials or on samples submitted for routine diagnostics (2, 22). This study demonstrates that a qRT-PCR assay for the biological subtypes of breast cancer can be used with a centroid-based classifier to predict tumor type from FFPE tissues. The assay has application in the clinical laboratory for prognosis in breast cancer.

Grant/funding support: This work was supported by National Cancer Institute Grants R33-CA97769-01 (to P.S.B.) and P50-CA58223-09A1 (to C.M.P.).

Financial disclosures: None declared.

Acknowledgements: We appreciate the help of the core facilities for tissue procurement at the participating institutions. We thank Carlynn Willmore-Payne and Joseph A. Holden for their technical expertise.

References

1. Fan C, Oh DS, Wessels L, Weigelt B, Nuyten DS, Nobel AB, et al. Concordance among gene-expression-based predictors for breast cancer. *N Engl J Med* 2006;355:560–9.
2. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 2004;351:2817–26.
3. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. *Nature* 2000;406:747–52.
4. van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002;415:530–6.
5. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 2002;347:1999–2009.
6. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 2001;98:10869–74.
7. Hu Z, Fan C, Oh DS, Marron JS, He X, Qaqish BF, et al. The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics* 2006;7:96.
8. Perreard L, Fan C, Quackenbush JF, Mullins M, Gauthier NP, Nelson E, et al. Classification and risk stratification of invasive breast carcinomas using a real-time quantitative RT-PCR assay. *Breast Cancer Res* 2006;8:R23.
9. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* 2003;100:8418–23.
10. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* 1998;95:14863–8.
11. Bair E, Tibshirani R. Semi-supervised methods to predict patient survival from gene expression data. *PLoS Biol* 2004;2:E108.
12. Bullinger L, Dohner K, Bair E, Frohling S, Schlenk RF, Tibshirani R, et al. Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. *N Engl J Med* 2004;350:1605–16.
13. Tibshirani R, Hastie T, Narasimhan B, Chu G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc Natl Acad Sci U S A* 2002;99:6567–72.
14. Hu Z, Troester M, Perou CM. High reproducibility using sodium hydroxide-stripped long oligonucleotide DNA microarrays. *Biotechniques* 2005;38:121–4.
15. Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, et al. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res* 2002;30:e15.
16. Szabo A, Perou CM, Karaca M, Perreard L, Quackenbush JF, Bernard PS. Statistical modeling for selecting housekeeper genes. *Genome Biol* 2004;5:R59.
17. Dudoit S, Fridlyand J. A prediction-based resampling method for estimating the number of clusters in a dataset. *Genome Biol* 2002;3:RESEARCH0036.
18. Lin LI. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 1989;45:255–68.
19. Dudoit S, Fridlyand J, Speed TP. Comparison of discrimination methods for the classification of tumors using gene expression data. *J Am Stat Assoc* 2002;97:77–87.
20. Benito M, Parker J, Du Q, Wu J, Xiang D, Perou CM, et al. Adjustment of systematic microarray data biases. *Bioinformatics* 2004;20:105–14.
21. Dabney AR. Classification of microarrays to nearest centroids. *Bioinformatics* 2005;21:4148–54.
22. Cronin M, Pho M, Dutta D, Stephans JC, Shak S, Kiefer MC, et al. Measurement of gene expression in archival paraffin-embedded tissues: development and performance of a 92-gene reverse transcriptase-polymerase chain reaction assay. *Am J Pathol* 2004;164:35–42.