Supplementary Methods

Central Pathology Review

Representative formalin fixed paraffin-embedded (FFPE) breast tumor tissue samples from patients participating in the MA.17 clinical trial were collected, stored and tissue sections cut at the NCIC CTG and/or the MGH Molecular Pathology Research Unit. All specimens underwent routine quality control evaluation for the presence of adequate invasive carcinoma. For IHC evaluation, two TMA blocks were constructed at the Yale Pathology Tissue Services using two 1.0-mm cores per patient tumor. Sections 5 μm in thickness from TMAs or whole tumor tissue sections (when blocks were not made available for TMA coring) were immunostained with the anti-ER antibody (clone SP1, Ventana) using the Ventana Benchmark System and with the anti-PR and -HER2 antibodies (clone PR 636 and clone CB11, dilution 1:50 and 1:250; Dako and Biocare Medical, respectively) using the Thermo Scientific LabVision 420 automated stainer. Central HER2 FISH was performed using the Vysis PathVysion test per the manufacturer's instructions. Positive and negative controls were used in each IHC staining and FISH run. Central ER, PR and HER2 immunohistochemical and HER2 FISH assays were independently reviewed and scored visually according to the ASCO/CAP guidelines[1,2] by two pathologists (D.C.S. and A.B.), blinded to all clinico-pathological and gene expression data. ER, PR and HER2 IHC status used in the statistical analysis were based on central laboratory testing and complemented by local testing results when tumor tissue was not available. For both ER and PR, 238 cases had central testing available, 11 used local testing. For HER2, 234 cases had central HER2 testing available, and in 15 cases HER2 IHC / FISH testing was not available because of lack of sufficient tissue in the TMA blocks. For these 15 cases, HER2 status was determined by RT-PCR measurements using a cutoff (>= -0.05 and < 0.05 for positive and negative, respectively) that produced the greatest accuracy (97% concordance) in the samples for which central IHC/FISH testing results were available (N = 234).

Case-Control Selection

A nested case-control design with a planned ratio of 1:2 recurrences versus non-recurrences, was implemented to study the ability of H/I to predict late recurrence and the benefit of extended therapy with letrozole. Consent to collect primary tumors for analysis was approved by the North American Correlative Science Committee and by each institution's local IRB. Patients were sampled from 16 strata defined by

age (< 60 yr vs. ≥ 60yr), T stage (T1+T2 vs. T3+T4+unknown), node status (positive vs. negative) and prior chemotherapy (yes vs. no). Cases were not matched based on study treatment in order to directly assess whether H/I could identify patients who were likely to respond to letrozole therapy. Cases included patients with local, regional or distant recurrence for whom primary tumor tissue blocks were available. For each case, two matched controls were randomly selected within the corresponding strata from patients who had been recurrence-free for a period of time longer than the case. Patients with contralateral breast cancer or unknown disease, insufficient invasive carcinoma as assessed by pathological evaluation or a weak RT-PCR signal (average cycle threshold for normalizing genes, > 26.5) were excluded from the study. Cases of recurrence from women on letrozole who developed the recurrence after enrolling in the MA.17R trial for additional 5-years of letrozole treatment, or from women on placebo who had recurrence after electing to switch to letrozole therapy after unblinding, were excluded. Similarly, controls were selected neither from those post-unblinding patients in the placebo arm who switched to letrozole, nor from those in the letrozole arm post-unblinding who enrolled in MA.17R, before the corresponding cases recurred.

Estimation of Absolute Risk of Recurrence

Estimation of absolute risk of breast cancer recurrence from the nested case-control study was done using the methods described in Borgan, Goldstein and Langholz [3] and Langholz and Borgan [4]. Two controls were randomly sampled from those matching potential controls for each recruited case in this study. With n is the total number at risk at the relapse time and $\tilde{R}$ denote the case-control set, the sampling weights due to the control selection are given by (Borgan, Goldstein and Langholz [3]; equation 4.2)

$$n \times \frac{1}{\sum_{j \in \tilde{R}} 1} = n/3.$$

In addition, features of the nested case-control sampling in the present study that are relevant to computing sampling weights were that 100 (31%) of the 319 recurrences in the cohort study were recruited into this case-control study, which requires the sampling weights to be adjusted by the proportion (Langholz and Borgan [4]). The reduction in case numbers due to exclusion of 17 cases for contralateral/unknown recurrences is accounted for by the reduced number of hazard increments estimated; i.e., the hazard increments for the 17 excluded cases are zero. With x_0 an indicator vector identifying the H/I-treatment group of interest, the hazard increment for at the time of each case's relapse is estimated by

$$\hat{h}_k(x_0) = 1/\sum_{j \in R_k} \exp[\hat{\beta}(X_j - x_0) + \log(.31 \times n/3)]$$

where X_j is H/I-treatment group indictor vector for person *j* in the case-control set, and $\hat{\beta}$ is the vector of rate ratios for the H/I-treatment groups. The estimated cumulative hazard is then given by $\hat{H}(t_l, x_0) = \sum_{k: t_k \le t_l} \hat{h}_k(x_0)$, along with 95% confidence intervals, were estimated using the SAS Cox regression procedure PHREG with log(.31 × n/3) specified as an *offset* in the model. Risk of recurrence up to time *t* (with 95% confidence intervals) was then estimated using the standard expression from survival analysis $1 - \exp[-\hat{H}(t_l, x_0)]$.

References:

1. Hammond ME, Hayes DF, Dowsett M, et al: American Society of Clinical Oncology/College Of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. *J Clin Oncol* 2010; **28**:2784-95.

2. Wolff AC, Hammond ME, Schwartz JN, et al: American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. *Arch Pathol Lab Med* 2007; **131**:18-43.

3. Borgan O, Goldstein L, Langholz B. Methods for the analysis of sampled cohort data in the Cox proportional hazards model. *Annals of Statistics* 1995; **23**:1749-1778.

4. Langholz B, Borgan O. Estimation of absolute risk from nested case-control data. *Biometrics*. 1997;**53**:767-774.

Supplementary Table 1. Comparison of patient clinical and pathological characteristics between The case-control study and the overall MA.17 trial

	Case-control (n=249)	MA.17 overall (n=5157)	P value*
Age at diagnosis			
<50 yr	9 (4%)	415 (8%)	0.08
50-59 yr	83 (33%)	1730 (33%)	
60-69 yr	82 (33%)	1684 (33%)	
≥70 yr	75 (30%)	1321 (26%)	
Unknown	0	7 (<1%)	
Node status			
Negative	94 (38%)	2581 (50%)	<0.0001
Positive	146 (59%)	2370 (46%)	
Unknown	9 (3%)	206 (4%)	
Hormonal receptor status			
Positive	247 (99%)	5048 (98%)	0.23
Negative	2 (1%)	109 (2%)	
Type of surgery			
Lumpectomy	152 (61%)	2917 (57%)	0.18
Mastectomy	129 (52%)	2587 (50%)	0.66
Axillary-node dissection	237 (95%)	4875 (95%)	0.23
Prior adjuvant radiation therapy			
No	150 (60%)	2059 (40%)	<0.0001
Yes	99 (40%)	3078 (60%)	
Unknown	0	20 (<1%)	
Prior adjuvant chemotherapy			
No	148 (59%)	2797 (54%)	0.13
Yes	101 (41%)	2352 (46%)	
Treatment			
Letrozole	122 (49%)	2575 (50%)	0.82
Placebo	127 (51%)	2582 (50%)	

* P values were calculated using Chi-square test, except for Type of surgery where binomial proportional test was used.

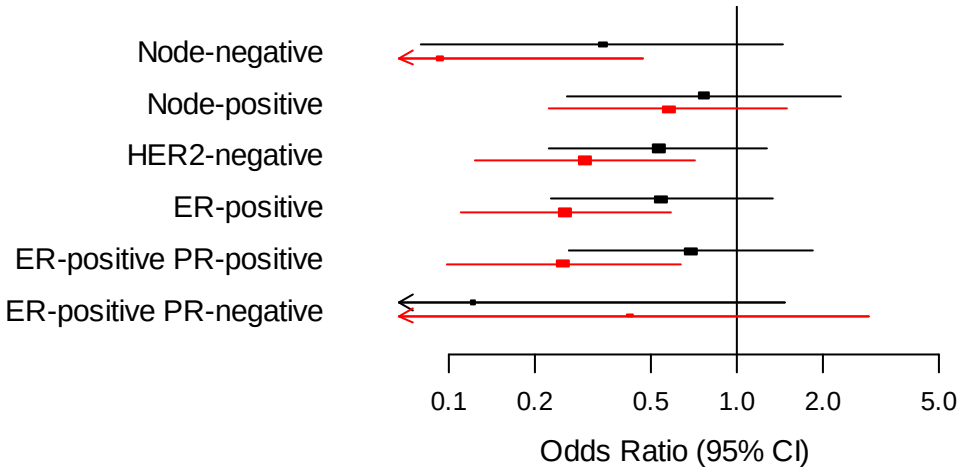
Supplementary Table 2. Unadjusted and Adjusted analyses of prognosis of late recurrence by H/I groups in each treatment arm

Variable	Unadjusted OR [95% CI]	*P* value	Adjusted OR [95% CI]	*P* value
Age (Post vs. Pre)	0.25 [0.02-2.76]	0.2583	0.13 [0.01-1.60]	0.1097
Tumor size (T2+T3 vs. T1)	1.00 [0.23-4.35]	1.0000	1.13 [0.21-6.00]	0.8832
Grade (3 vs. 1-2)	1.56 [0.82-2.98]	0.1753	1.23 [0.58-2.60]	0.5949
ER status (pos vs. neg)	0.67 [0.15-2.98]	0.5955	0.83 [0.15-4.72]	0.8349
PR status (pos vs. neg)	1.05 [0.53-2.09]	0.8802	1.33 [0.62-2.86]	0.4604
HER2 (pos vs. neg)	1.32 [0.55-3.18]	0.5382	0.99 [0.35-2.78]	0.9823
Node status (pos vs. neg)	1.00 [0.06-15.99]	1.0000	1.93 [0.11-33.77]	0.6519
Prognosis (H/I-high vs. H/I-low)				
Placebo	2.24 [1.09-4.61]	0.0282	2.15 [1.00-4.64]	0.0506
Letrozole	1.15 [0.53-2.50]	0.7190	1.22 [0.54-2.76]	0.6256

Supplementary Table 3. Adjusted analyses of treatment benefit by H/I groups in clinical relevant patient subgroups

Patient Groups	**N**	**H/I Group**	**Adjusted OR [95% CI]**	***P* value**
Node-negative	94	H/I-low	0.34 [0.08-1.45]	0.1465
		H/I-high	0.09 [0.02-0.47]	0.0040
Node-positive	146	H/I-low	0.77 [0.26-2.30]	0.6368
		H/I-high	0.58 [0.22-1.50]	0.2611
HER2-negative	226	H/I-low	0.53 [0.22-1.28]	0.1587
		H/I-high	0.30 [0.12-0.71]	0.0065
ER-positive	242	H/I-low	0.55 [0.22-1.33]	0.1849
		H/I-high	0.25 [0.11-0.59]	0.0014
ER-positive & PR-positive	202	H/I-low	0.69 [0.26-1.83]	0.4605
		H/I-high	0.25 [0.10-0.64]	0.0036
ER-positive & PR-negative	40	H/I-low	0.12 [0.01-1.47]	0.0972
		H/I-high	0.42 [0.06-2.87]	0.3790

Supplementary Figure 1. Forest plots showing odds ratio (OR) for recurrence due to extended letrozole treatment for each H/I group in clinically relevant patient subgroups. Red color indicates H/I-high patients, black color for H/I-low patients.



CORRELATIVE SCIENCE PROTOCOL

Revision (2) 12/12/2005.
Proposal for use of specimens of the Breast Cancer Intergroup of North America (TBCI) in correlative scientific studies

INSTRUCTIONS FOR PROPOSING INVESTIGATORS:

Please type directly into this electronic form. The completed form should not exceed 10 pages in length (excluding references and appendices). You may, however, attach appendices to provide further details.

Proposals should adhere to the guidelines set forth in the Instructions found on the TBCI Correlative Sciences Committee Web site at http://ctep.cancer.gov/resources/tbci/correlative_studies.html

Proposals should be emailed to Daniel Hayes, M.D., TBCI Correlative Sciences Committee Chair, at HayesDF@umich.edu, with carbon copy to Rebecca Enos, MPH, TBCI Coordinator at REnos@emmes.com. The committee accepts only electronic submissions.

Submission deadlines are February 1, May 1, August 1, and November 1. Under special circumstances, proposals may be reviewed at non-standard times.

Date submitted: December 12, 2005 (Rev 2)

Investigator information

Name of applicant: Paul Goss M.D. Ph.D, Dennis Sgroi M.D.

Cooperative Group affiliation: NCIC CTG, CALGB

Mailing address: Massachusetts General Hospital, 32 Fruit St. Boston, MA 02114

Email: pgoss@partners.org; dsgroi@partners.org

Phone: 617-724-3118 Fax: 617-724-3166

Co-investigators

Name: Lois Shepherd M.D. Group affiliation (if any): NCIC CTG
Institution: NCIC CTG Queens University, Kingston ON Email: lshepherd@ctg.queensu.ca

Name: Dr. James Ingle M.D. Group affiliation (if any): NCCTG
Institution: Mayo Clinic Email: ingle.james@mayo.edu

Name: Sridhar Ramaswamy M.D. Group affiliation (if any): none
Institution: Harvard University Email: sridhar@mgh.harvard.edu

Name: Paula Ryan M.D. Ph.D. Group affiliation (if any): CALGB
Institution: Harvard University Email: pryan@partners.org

Name: Dr.David Rimm M.D. PhD. Group affiliation (if any): none
Institution: Yale University School of Medicine Email: david.rimm@yale.edu

Name: Dianne Finkelstein PhD. Group affiliation (if any): none

Institution: Harvard University Email: dfinkelstein@partners.org

Name: Dongsheng Tu PhD. Group affiliation (if any): NCIC CTG
Institution: NCIC CTG Queens University Email: DTu@ctg.queensu.ca

Name: Dr. Hyman Muss M.D. Group affiliation (if any): CALGB
Institution: University of Vermont Email: Hyman.Muss@vtmednet.org

Name: Hironobu Sasano, M.D., Ph.D. Group affiliation (if any): None
Institution: Tohoku University School of Medicine Email: hsasano@patholo2.med.tohoku.ac.jp

Peggy Porter, M.D. Group affiliation (if any): SWOG
Institution: University of Michigan Email: pporter@fhcrc.org

Name: Dr. Michael Pins M.D. Group affiliation (if any): ECOG
Institution: Northwestern University Email: m-pins@northwestern.edu

Name: Soonmyung Paik M.D. Group affiliation (if any): NSABP
Institution: University of Pittsburgh Email: soon.paik@NSABP.org

Name: Ari Ristimaki M.D. PhD. Group affiliation (if any): None
Institution: University of Helsinki Email: ari.ristimaki@helsinki.fi

Title of proposed correlative study: Quantitative Protein and Gene Expression Biomarkers of Tamoxifen and Letrozole Recurrence in the NCIC CTG MA.17 Cohort.

Study number: MA.17ICSC
(You may assign a number specific to your own Group or institution's numbering system.)

From which treatment trial(s) are you requesting specimens?

Protocol number: **JMA-17**

Protocol title: A PHASE III RANDOMIZED DOUBLE BLIND STUDY OF LETROZOLE VERSUS PLACEBO IN WOMEN WITH PRIMARY BREAST CANCER COMPLETING FIVE OR MORE YEARS OF ADJUVANT TAMOXIFEN (NCIC CTG MA.17/JMA.-17)

Coordinating Group: **NCIC CTG**

Tissue type

What tissue types are you requesting? *(e.g., FFPE malignant primary)*: 957 primary tumors (319 from recurrences and 638 matched non-recurrences) from consenting study participants (prior to unblinding of the study) in the MA.17 clinical trial.
Required number and thickness of sections:
We plan to keep this to a minimum, conserving approximately 85% of the material for future studies. Our revised requirements are:

a) Six 5 μm sections for the MGH 2 Gene Signature
b) Six 5 μm sections for the GHI Oncotype DX assay
c) One 5 μm section for centralized histology review and tumor grading
d) Eight 5 μm sections for microarray-based gene discovery

e) Six 5 μm TMA sections for standard IHC (one per marker; ER, PgR, Her1, Her2, Cox2, aromatase)
f) Eight 5 μm TMA sections for Aqua (one per marker; ER, PgR, Her1, Her2, Cox2, Aromatase, NAT1, GATA3)

Hypotheses:

What are your hypotheses? We hypothesize that:

1. The recently described gene expression signatures sub-classify patients from the MA.17 trial into improved prognostic and treatment-predictive subgroups.
2. A new technique, quantitative immunofluorescence, will be a superior prognostic and treatment-predictive tool as compared with standard immunohistochemistry when applied to tumors from patients treated in MA.17.
3. We can identify (through DNA microarray-based discovery using RNA extracted from formalin fixed paraffin embedded tissues) novel gene expression profiles that will be predictive of outcome and predictive of responsiveness to letrozole.

Objectives:

What are your objectives?

1. To conduct a case control study to evaluate the prognostic utility of the MGH two-gene and the GHI 21-gene expression signatures in the MA.17 sub-cohort that received tamoxifen followed by placebo (Tam-Plac cohort) and in the MA.17 sub-cohort that received tamoxifen followed by letrozole (Tam-Let cohort versus Tam-plac cohort).
2. To evaluate the ability of the MGH two-gene and the GHI 21-gene expression signatures to predict responsiveness to letrozole.
3. To evaluate the prognostic utility of quantitative immunofluorescence of ER, PR, Her-2, tumor aromatase, COX-2, GATA3 and Nat1 in the TAM-PLACEBO and the TAM-LETROZOLE cohorts and compare these results with that derived by standard immunohistochemistry.
4. To evaluate the ability of quantitative immunofluorescence and standard immunohistochemistry of the aforementioned proteins (stated in objective # 3) to predict responsiveness to letrozole (Tam-Let cohort versus Tam- Plac cohort).
5. To undertake gene discovery from FFPE tumor specimens to identify novel gene expression profiles which may predict for outcome in the MA.17 cohort (Tam-Let and Tam-Plac cohorts) and predict responsiveness to letrozole (Tam-Let cohort versus Tam-Plac cohort).

Facilities & personnel

Who will be doing the work? Please include the name(s) of investigator(s) responsible for the assay(s): Formalin fixed paraffin-embedded breast tumor tissue samples from patients participating in the MA.17 clinical trial will be collected through an initiative led by Dr. Goss (MGH) and Dr. Lois Shepherd at the NCIC CTG (Queens University, Kingston, Ontario). All collected FFPE tumor blocks will be stored, and tissue sections cut and TMAs made, at the NCIC CTG tumor banking facility at Queens University under Dr Shepherd's supervision. The tissue sections from each case will be forwarded to the MGH Molecular Pathology Research Unit and distributed accordingly (see below). Tumor collection is underway at the NCIC CTG central repository where 176 tumors have already been banked. Central Collections are also underway through other group offices. Central office funding is available for this collection of samples at the NCIC CTG.

For the MGH 2 gene signature gene expression aspect of the proposal, RNA will be extracted from 6 (5μm) tissue sections and subjected to real-time quantitative PCR (RTQ-PCR) in the MGH Molecular Pathology Research Unit and the data will be analyzed by Dr Dianne Finkelstein (MGH Biostatistics Center). For the GHI 21-gene signature Genomics Health Inc. will perform real-time quantitative PCR (RTQ-PCR), and subsequent analysis in collaboration with Dr Dianne Finkelstein (MGH Biostatistics Center).

For routine ER, PR, Her1, Her2 immunohistochemistry (IHC) 6 5μm tissues sections will be sent to Dr. Sgroi and Dr. Goss at the MGH Immunohistochemistry DF/HCC SPORE Core Facility within the Molecular Pathology Unit. This MGH facility is well equipped, and the staff has extensive experience with immunohistochemistry. IHC of COX-2 will be performed by Dr Ari Ristimaki (see letter of collaboration attached). IHC of aromatase will be conducted by Professor Hironobu Sasano.

For the advanced quantitative immunofluorescence aspect of the project, TMAs will be sent to the Rimm laboratory at Yale University. The immunofluorescence assay development and quantitative analysis will be performed under the direction of Dr David Rimm. Dr Rimm has developed a unique imaging methodology that accurately and reproducibly quantifies immunofluorescent staining of tissue sections (see below) [1]. Dr. Rimm will collaborate with Professor Sasano and Dr Ristimaki regarding the quality control and analysis of tumoral aromatase and COX-2, respectively. Immunofluorescence data analysis will be performed by Dr Dianne Finkelstein in collaboration with Dr David Rimm.

For the microarray novel gene discovery aspect of the proposal, 5μm tissue sections will be sent to Dr Sgroi's lab, where RNA will be extracted, amplified and labeled. The microarray hybridization will be performed at the MGH Cancer Center Affymetrix Core Facility and Dr Sridhar Ramaswamy will perform the gene expression data analysis using methodologies previously described [2].

In which facilities will the work be performed? The MGH Molecular Pathology Research Unit possesses all the necessary equipment and molecular biological skills to extract and amplify RNA from FFPE tissue samples. This lab is equipped with an ABI 7900 HT (Applied Biosystems) real-time quantitative PCR system that is necessary for the gene expression analysis aspect of the study. For the analysis of the GHI 21-gene signature, RNA will be extracted and real-time quantitative PCR analysis will be performed at Genomic Health Inc. using their proprietary technologies. Routine immunohistochemistry will be performed at the MGH Immuno-histochemistry Core Facility and within Dr Goss's laboratory within molecular pathology at MGH. IHC of COX-2 will be performed by Dr Ari Ristimaki at the Department of Pathology, Molecular and Cancer Biology Research Program, Helsinki University Central Hospital, Finland. IHC of aromatase will be conducted by Professor Hironobu Sasano, in the department of pathology at Tohoku University School of Medicine in Japan. Dr. Rimm's laboratory operates the quantitative immuno-fluorescence and tissue microarray facility at Yale Medical School (YCCMF). Dr Rimm's lab has two PM-1000 image stations, and the AQUA™ (for Automated Quantitative Analysis) imaging software that was designed and created by Drs. Camp and Rimm. For the microarray novel gene discovery aspect of the proposal RNA will be extracted, amplified and labeled in MGH Molecular Pathology Unit and hybridizations will be performed at the MGH Cancer Center Affymetrix Core Facility.

If this study is approved, who would be the person(s) at your tissue bank responsible for coordinating the receipt of specimens from other Groups/sites? FFPE tumor blocks will be

received and archived with coded identifiers in the NCIC CTG tumor banking facility at Queens University (Dr Shepherd), Kingston, Ontario, Canada. Drs. Paul Goss and Dennis Sgroi will be responsible for interactions between NCIC CTG, Yale University, Genomics Health Inc, Dr Ristimaki and Prof Sasano and Massachusetts General Hospital regarding sample transfer. When clinical information is requested, Dr Goss, Dr Sgroi and Dr Finkelstein will interact with Dr Dongsheng Tu within the NCIC CTG biostatistical center. Dr Dongsheng Tu will undertake the matching of cases and controls using the MA.17 clinical database in Kingston, Ontario, Canada.

Brief overview of proposed study

Please briefly summarize the study you are proposing:

Background and rationale

Clinicopathological breast cancer biomarkers

The classic clinicopathological parameters that include patient age, tumor size, tumor grade and lymph node status fail to accurately predict clinical outcome in the setting of tamoxifen therapy. The best predictors of tamoxifen responsiveness in the clinical setting are the absolute levels of ER and progesterone receptor (PR, an indicator of functional ER pathway) expression [3,4]. However, 25% of ER+/PR+ tumors, 66% of ER+/PR- cases and 55% of ER-/PR+ cases fail to respond or develop early resistance to tamoxifen, through mechanisms that remain unclear [4, 5,6,7]. Currently, no reliable means exist that allow the identification of these non-responders. In these patients, the use of alternative hormonal therapies, such as the aromatase inhibitors letrozole and anastrozole [8-10], chemotherapeutic agents [11], or inhibitors of other signaling pathways, such as trastuzmab, gefitinib and others [5,6,12-14] might offer the possibility of improving clinical outcome. Discovery of a biomarker that identifies those patients who will become resistant to tamoxifen therapy would assist in selecting optimal alternative therapeutic strategies.

NCIC CTG MA.17 Trial

Results from the NCIC CTG -sponsored international MA.17 trial have demonstrated that the addition of 5 years of an aromatase inhibitor, letrozole, after the completion of standard tamoxifen treatment significantly improves disease free, distant disease free and overall survival (in node positive patients) in women with breast cancer [8]. These results have had a considerable effect on the treatment of early-stage breast cancer. The MA.17 trial consists of a unique cohort of patients and as such, it provides an unprecedented opportunity to assess biomarkers that may predict which women with ER+ breast cancer who have received 5 years of tamoxifen may benefit from continued letrozole treatment and which women can be safely monitored without further therapy. Importantly, identification of prognostic and predictive biomarkers of tamoxifen and letrozole therapy may provide for personalized (or individualized) treatment regimens that ultimately result in an increased overall absolute benefit for women with hormone receptor positive breast cancer. In summary many women receive unnecessary treatment because we currently have no means of identifying who will and who will not respond. Additionally some women have recurrence of their cancer despite all existing therapies. Our project will help to identify these women and to develop novel patient-tailored therapeutic approaches.

Microarray-based gene expression biomarkers

Contemporary advances in expression profile analysis using microarrays are revolutionizing the classification of most human cancers by providing a comprehensive molecular analysis of all expressed genes to identify highly sophisticated biomarkers of clinical behavior. Early studies in lymphomas, leukemias, as well as breast cancer, have shown that morphologically similar cancers

can in fact be comprised of distinct subtypes at the molecular level [15-22]. Furthermore, we, and others have shown that the pattern of genes expressed in specific subsets of cancer is likely to be representative of the pathways driving malignancy [21,23]. Several retrospective studies in human breast cancer have demonstrated that expression profile analysis can identify prognostic categories that may guide treatment choices. Recent studies linking gene expression profiles to clinical outcome have demonstrated that the potential for distant metastasis and overall survival probability may be predictable through biological characteristics of the primary breast tumor. In particular, a 70-gene expression signature has been identified from a cohort of untreated breast cancer patients and was shown to be a strong prognostic factor, out-performing all known clinicopathological parameters [24,25]. Furthermore, Chang et al recently identified a 92-gene expression signature that predicted responsiveness to neoadjuvant docetaxel [26]. These studies highlight the robust nature of microarray-based biomarker strategies and led to the possibility of a broad application of these technologies as a means to identify clinically useful prognostic categories. Currently, our greatest need is to identify breast cancer biomarkers that might guide treatment choices in the clinical setting and that provide insight into the mechanisms associated with the development and progression of this disease.

MGH 2-gene and GHI 21-gene signature biomarkers

Recently, we have identified a novel HOXB13:IL17BR gene expression ratio (signature) that predicts outcome in breast cancer patients treated with adjuvant tamoxifen monotherapy. This two-gene expression signature (MGH 2-gene signature) outperforms all known positive (ER and PR) and negative (Her-1 and Her-2) predictors of responsiveness to tamoxifen in node negative women with a trend in the same direction for node positive women [27]. Importantly, this signature is predictive of early recurrence, i.e. within 5 years of initiation of adjuvant tamoxifen monotherapy, and has been validated with real-time quantitative PCR using routine formalin fixed paraffin-embedded (FFPE) clinical pathology breast cancer samples from two independent cohorts of women treated with adjuvant tamoxifen (one cohort from Massachusetts General Hospital and a second cohort in collaboration with Dr. James Ingle from the Mayo clinic). In addition, Genomic Health Inc in collaboration with Dr Soon Paik of the NSABP has also recently identified a 21-gene expression signature (GHI 21-gene signature) that predicts outcome to adjuvant tamoxifen monotherapy in lymph node negative patients [28]. This GHI 21-gene signature has been validated in a large randomized clinical trial of adjuvant tamoxifen (NSABP-B14 trial). The GHI 21-gene signature also appears independent of the known clinicopathological prognosticators. A head to head comparison of these two signatures in such a setting has not been performed to date. The potential of these signatures to aid in the treatment stratification process supports the justification for objectives #1 and #2.

Immunofluorescence- versus immunohistochemical-based breast cancer biomarkers

Many studies have evaluated the clinical relevance of measuring ER and PR by immunohistochemistry (IHC), and the large majority reported statistically significant relationships with clinical outcome. Despite the positive prognostic and predictive value of ER and PR IHC, there are significant limitations associated with the use of IHC, as 30-40% of hormone receptor positive breast cancer patients experience recurrence of their disease in the setting of adjuvant hormonal therapy. It is believed that a more objective continuous system of protein quantitative analysis of ER and PR would improve prognostic and predictive accuracy in breast cancer patients with hormone receptor positive disease. The prognostic and predictive utility of standard IHC has yet to be ascertained in a cohort of patients that have been treated with tamoxifen followed by letrozole. The lack of such information supports justification for objective #3. The recent development and successful application of quantitative immunofluorescence, a

technique that overcomes the aforementioned technical limitations of IHC (please see below), supports the justification for objective #4.

Novel gene expression signature discovery

Both the MGH two-gene signature and the GHI 21-gene signature have independent prognostic value for women taking adjuvant tamoxifen for early stage breast cancer. However, neither is in any way perfect at doing so, with both false positive and false negative results. In addition each of these profiles is prognostic but there is no overlap between the gene profiles. This suggests that there are additional gene expression patterns which have not been determined which may be of predictive and prognostic value. Two other features of the MA.17 cohort make a search for novel expression profiles of importance. Firstly unlike in the two gene and genomic health cohorts, relapses on this study are all "late recurrences" by definition and this may carry a specific prognostic profile. Secondly half of the MA.17 cohort received initial tamoxifen followed by letrozole and a distinct gene expression profile may carry specific prognostic and/or predictive profile for women relapsing on this treatment strategy.

Statistical design

Please provide below a description of your proposed statistical design.
Include power justification for the number of specimens requested; endpoints; monthly accrual rate; and description of planned analyses, including relevant cutoffs.

Although it is recommended that this section be developed in consultation with a biostatistician from the Group that led the clinical trial, the exact nature and extent of the collaboration is left to the investigator to define.

NCIC CTG MA.17 randomized 5137 women, disease free after five years of adjuvant tamoxifen, to five years of letrozole or placebo. As of February 2005, 319 women have experienced a recurrence of cancer (109 in the letrozole treated group and 210 in the placebo control group). We propose to conduct and analyze this study using a nested case-control design (as discussed in Langholz and Goldstein, "Risk set sampling in epidemiologic cohort studies", Statistical Science 1996, p 35-53.) [29]. For each recurrent patient, 2 matched non-recurrent control patients will be chosen from among patients who have been followed for at least as long as the time of the case's recurrence. The matching criteria will be based on age, tumor size, lymph node status, and prior chemotherapy. Time since diagnosis for the controls will be at least as long as that of their matched case. In addition, we will include all stratification factors defined in the MA 17 clinical protocol.

The primary objective of this study is to assess the predictive value of the gene expression signature (ratio HOXB13:IL17BR) for women experiencing a recurrence of their cancer versus a matched cohort of those remaining disease free on, firstly, the Tam-Plac arm of the trial and secondly, on the Tam-Let arm. Other objectives of this study include: to determine the value of the GHI 21-gene signature "recurrence score" in women experiencing recurrence of cancer versus those remaining disease free in the two arms; to compare differences in the predictive value of protein expression in recurrent versus non-recurrent invasive tumors in the two arms determined by standard immunohistochemistry versus quantitative immunofluorescence; to determine differences in novel gene expression signatures between women experiencing relapse or not of their cancer (in Tam-Plac alone and Tam-Let cohorts alone and comparing Tam-Let to Tam-Plac). The analysis of the GHI signature excluded women with DCIS/LCIS lesions. Four of 25 women in MA17 had a local recurrence which was DCIS/LCIS and 6/40 contralateral primaries

were DCIS/LCIS lesions. Thus for the purposes of comparing the MGH 2 gene signature to the GHI assay we will exclude cases in which recurrence was DCIS/LCIS.

Gene Expression (RTQ-PCR):
As performed in our previously published work, expression ratios of HOXB13:IL17BR will be determined as ΔC_T values (ΔC_T is the difference in PCR cycle threshold) and normalized across all patients.

For our other objective of evaluating the GHI 21-gene signature, a recurrence score will be calculated as described in (REFERENCE GHI PAPER HERE). We will assess the predictive performance of the MGH two- and GHI 21- gene signatures through assessing the area under the curve values (AUC) of the receiver operator characteristic (ROC) calculated for each signature.

Sample Size Considerations:

.

Our study will have sufficient power for an analysis of the odds ratio associated with being positive on MA17. Although this analysis will be done using a regression model that accounts for the matching criteria, we calculate power on the basis of a test of the stratified 2X2 tables with case/control on the rows and signature positive/negative on the columns. The table below indicates the odds ratio that can be detected for each n from the matched case-control study, given that the probability of MA17 positive among sampled control patients is 0.20 and the correlation coefficient for exposure between matched case and control patents is 0.20000. A sample of 319 case patients is obtained. For each case patient, a matching sample of 2 control patient(s) is also obtained. Thus, for example a sample of 957 patients achieves 80% power to detect an odds ratio of 1.64 versus the alternative of equal odds using a Chi-Square test with a 0.05000 significance level. This was calculated assuming the matched case-control design using PASS software.

Total n (cases + controls in 1:2 ratio)	Detectable OR
957	1.64
450	2.05
225	2.65
102	4.13

Therefore, on the basis of our 957 subjects, we will have a sufficient sample size to analyze the signature within subgroups such as node negative and node positive patients and so the MA.17 placebo cohort of over 2400 should be ample to provide the matched subjects to perform this study. Similarly power will hold in the aromatase arm. Estimates and confidence bounds for the odds ratio and classification accuracy will be calculated.

Gene Expression Profiling: Minimum sample size determination for the identification of a gene expression signature in a cohort such as that proposed herein has been, and continues to be, a challenging area of experimental design with both controversy and recent innovation in the literature. While there are at least several salient methodologies (i.e. Cochran-Armitage trend test, multi-factor dimensionality reduction, and Fisher Discriminant analysis, to name a few) for determining minimum sample size in the context of large microarray dataset size, we see Fisher Discriminant analysis as perhaps the best currently available numerical solution, as it concurrently provides balanced statistical power prediction and economical compactness of cohort set size. Given our current hypothesis of unbiased random distribution of gene expression levels and directions, we exclude any intrinsic benefit stemming from the use of asymmetric control pairing. Assuming that a trend is embedded in the case-matched data, the non-directed

query facilitated by Fisher Discriminant analysis will provide the appropriate subset, with use of exhaustive subset search of the overall set space. The computational methodology and requirements for this test are well established and computationally within realistic bounds [30]. Therefore, based on this methodology [30] and the results obtained by several predictive gene expression studies [24,25], we conservatively estimate the number of training set cases needed to identify a discriminatory gene expression signature to be no more than 80 patient samples (40 recurrences and 40 non-recurrences).

Once a predictive gene expression signature is converged upon, we will perform subsequent validation with an untested group of 80 patients; this is a group of samples that is independent of the training set. The identified gene signatures will be categorized as present or absent in each of the 40 subjects from the favorable and in each of the 40 subjects from the poor outcome (progression) groups. If we observe 38/40 subjects correctly categorized as having the gene signature for favorable (unfavorable) outcome, then we can be 80% certain that the sensitivity for detecting favorable (unfavorable) outcome is at least 80%. If we observe 36/40 correctly categorized, we can be 80% certain that the sensitivity is at least 80%.

These power calculations are based upon traditional statistical assumptions of independence. It is well appreciated that gene expression values in given samples do not behave as completely independent variables. This fact makes precise power calculations difficult to perform in a highly accurate manner. Therefore, our calculations are a best possible estimate given these limitations.

Immunofluorescence Analysis: Immunohistochemical and Immunofluorescence analysis for these secondary endpoints will compare protein expression between breast cancers in women experiencing recurrent and non-recurrent disease. The laboratory and statistical studies will be performed at Yale University under the direction of Dr. Rimm. In each of 957 samples in each of these two patient groups, AQUA measured expression will be obtained on each of the eight proteins: ER, PR, Her-1 and –2, aromatase, GATA3, NAT1, and COX-2. Our experience with the AQUA assay has been to either examine each protein expression in terms of their raw AQUA values, their ranked values (*i.e.*, transformed to uniform scale), or else dichotomized as above or below their marginal median, depending upon their empirical distribution. In prior studies of AQUA data we have found each of these transformations useful in different experiments. A risk profile will be constructed from each of these sets of transformed AQUA values using logistic and probit regression models, computed in standard statistical software packages. The risk profile, or linear discriminant function will be used to distinguish between the protein expressions in the recurrent and non-recurrent tumors. Statistical significance for inclusion in the risk profile will be determined using a Bonferroni correction. Specifically, we will declare statistical significance at the $\alpha = .05$ level if the multivariate significance level reported in the software achieves a *p*-value of .05/8= .00625 or smaller for each of these eight protein indicators. At significance level $\alpha = .05$, in a sample of size 957, a single binary valued protein marker will have power greater than .95 to detect an odds ratio of 1.7 using a chi-squared test in a 2x2 table. If we include the Bonferroni correction, each of the eight markers will have power more than .90 when the underlying odds ratio is 1.8.

We will construct ROC curves to express the sensitivity and specificity of linear models containing each of two, three, or more or these eight AQUA markers. Models will be compared on the basis of these criteria along with their simplicity of mathematical expression. When the final models are decided upon, these will be further examined using a cross validation with a delete-one jackknife to determine their true predictive quality. In a separate analysis we will also examine the eight binary coded AQUA values plus the binary variable indicating recurrence

status using log-linear model methodology. These analyses will be used to determine multivariate correlations within the various markers and with the recurrence status.

Analysis

The analysis of our MGH-2 marker will be based on a regression analysis for the association of the signature with the risk of recurrence. For our objective of evaluating the GHI 21-gene signature, a recurrence score will be calculated. The markers will be compared on the basis of how well they predict recurrence. This will be assessed on the basis of the effect on the likelihood of inclusion of the covariate indicating the marker. We will consider both binary covariates (based on whether a marker is above or below a pre-determined cut-off) as well as continuous valued covariates which reflect raw value of the marker or a calculated recurrence score (as appropriate).

We will perform a standard multivariate analysis of the risk of recurrence using these markers which we measure both for standard IHC and quantitative immunofluorescence (IFL). We will do this analysis using a conditional logistic regression model, which is stratified by time since diagnosis. Marker values will be included as covariates in the model, as well as whether IFL or IHC was used, what treatment the patient received, and an interaction between treatment and marker.

We will calculate the AUC for the ROC curves from the GHI signature and the MGH-two gene signature utilizing the methods described in Chapters 8 and 9 of the book by M Pepe, "The Statistical Evaluation of Medical Tests for Classification and Prediction" (Oxford Press 2003) which handle the time-dependent nature of the data as well as issues of subject dependence [31].

Marker and assay information

Please provide background information on the marker, including its designation, biology, and how it is detected (the type of assay, reagents, etc.).

MGH 2-gene signature

Background: The biological significance of HOXB13 and IL17BR as it relates to the biology and pathogenesis of human breast cancer is poorly understood and is under active investigation. IL17BR is expressed in both lymphoid and epithelial cells and cells lines and evidence suggests that it may mediate pro-inflammatory cytokine signaling. HOXB13, on the other hand, is a member of the homeoprotein family of transcription factors that are considered to be master regulators of pattern formation during embryogenesis. HOXB13 is expressed in several different normal and neoplastic tissues including breast and prostate cancer [32,33]. We have demonstrated HOXB13 expression in the normal breast epithelial cells of the terminal duct lobular unit as well as various stages of human breast cancer progression. In the MCF-10A cell line, a non-transformed mammary epithelial cell, ectopic expression of HOXB13 has been demonstrated to enhance EGF- and Her-2-mediated cell motility and invasion. Furthermore, ectopic expression of HOXB13 in the highly estrogen sensitive MCF-BUS cell line results in enhanced ER-mediated transcriptional activation as measured in vitro using a luciferase reporter plasmid under the control of a tandem consensus estrogen receptor element (ERE) motif. Recent evidence in prostate cancer cell lines suggests that HOXB13 functions as a cell growth suppressor [33]. This observation in conjunction with our unpublished data suggests that the HOXB13-mediated

functional activity may differ significantly depending upon the cell-specific context in which it is expressed [27,33].

The Assay and Cutoff: All of the proposed RTQ-PCR methodologies have been previously demonstrated to perform well in FFPE tissues that are up to 20 years of age [27]. RNA extraction, amplification and real-time quantitative PCR will be performed as described [27]. Briefly, RNA from FFPE tissue sections will be extracted and subjected to one round of T7-based RNA amplification using the Paradise Reagent System (Arcturus Biosciences, Mountain View, CA). RNA will be converted to cDNA and subjected to real-time quantitative PCR with an ABI 7900HT using previously described primers and fluorogenic probes to HOXB13 and IL17BR [27]. The original MGH two-gene RTQ-PCR assay required RNA greater than 130bps. Recent interrogation of FFPE blocks from different institutions demonstrates variable RNA length. In fact, RNA from older (>18 years) FFPE blocks averages ~100bps. Thus, in order to improve the sensitivity of the MGH two-gene assay, we have recently re-designed RT-QPCR primer probe sets such that extremely fragmented (< 80bps) RNA can be successfully and reproducibly quantified by RTQ-PCR [33]. Given that the MA.17 FFPE blocks are less than 15 years old, we do not anticipate any technical issues with either the MGH or GHI assays. The HOXB13:IL17BR gene expression ratio will be compared between recurrences and non-recurrences in the MA.17 cohorts using the expression ratio cutoff identified in our previous studies. More specifically we will use a predetermined cutoff (this cutoff was generated in our initial training set and tested in an independent cohort of patients- please see [27] Ma et al. Cancer Cell 2004) to dichotomize the continuous HOXB13:IL17BR ratio variable.

The GHI 21-Gene Expression Signature.

Background: In addition to our signature, Dr Soon Paik in collaboration with Dr Steve Shak at Genomic Health Inc. (GHI) has also recently identified a 21-gene expression signature (GHI 21-gene signature) that predicts outcome to adjuvant tamoxifen monotherapy, and most recently it has been determined to predict benefit from chemotherapy [28]. The GHI 21-gene signature has been validated in a large randomized clinical trial of adjuvant tamoxifen (NSABP-14 trial) and appears independent of the known clinicopathological prognosticators. Like the MGH signature, the GHI signature is a quantitative real-time PCR-based assay that was optimized for use in routine FFPE tissue sections. The GHI signature consists of 5 normalizing genes and 16 other genes that divided into proliferation-, invasion-, Her-2- and estrogen-related gene categories. Several of these genes (ER, PR, Her-2 survivin) have been associated with breast cancer prognosis and/or prediction of responsiveness to tamoxifen therapy, while others have been putatively implicated in playing a role in the pathogenesis and progression of the disease.

The Assay and Cutoff: The GHI 21-gene signature will be conducted at GH using their proprietary methodologies. RNA will be extracted and non-amplified RNA subjected to quantitative real-time PCR. Using proprietary algorithms a "recurrence score" will be generated from the gene expression data from all 21 different genes and use cutoffs as previously described [28].

Immunohistochemistry-based and immunofluorescence-based assays

Background: Immunohistochemistry is routinely applied to FFPE tissues worldwide. Conventional pathologist-based analysis of protein quantity in tissue sections suffers from a number of weaknesses. Perhaps the most significant is the lack of a continuous system of analysis. Most pathologists use an ordinal scale of 0-3 to estimate staining intensity. Unfortunately, the human eye is not well suited to assess intensity. People are not able to

reproducibly subdivide intensity levels in immunostained specimens, whether they are conventional histologic sections or tissue microarray histospots. Over the past few years, the Rimm laboratory has developed a method of analysis for tissue microarrays that circumvents this problem [1]. First a series of images are collected by the PM-1000, a custom microscope platform assembled in our lab from "off the shelf components". Then each image set is analyzed by a system called AQUA (for Automated Quantitative Analysis). It is a set of algorithms that provides a reproducible, automated, quantitative analysis of expression of a given marker within a user defined sub-compartment (or subcellular locale) in a histospot. This method allows measurements of protein expression within subcellular compartments that result in a number directly proportional to the number of molecules expressed per unit area. Details of out-of-focus light subtraction imaging methods required to make this analysis work are described in detail in Camp et al [1] and have been utilized in a number of published studies [34-35].

ER, PR HER1 and HER2 IHC

Assay and Cutoff: Immunohistochemistry is routinely applied to FFPE tissues worldwide. ER, PR HER1 and HER2 IHC are all markers that are used in daily practice in breast pathology. The justification for these markers is more for validation of the TMA and the cohort than for true discovery. The antibodies used, method of analysis, scoring system, and cutoffs closely mirror the currently accepted clinical practices and are summarized in the table below:

Antigen and Antibody source	**Ab name and Method of analysis**	**Scoring system (pathologist-based assessment)**	**Cut point**
Estrogen Receptor (ER) from Dako	clone ID5, 1:50 dilution for 1 hour	Percent nuclei positive	Any case with greater than or equal to 10% is considered positive
Progesterone Receptor (PR) from Dako	clone 636, 1:50 dilution for 1 hour	Same as ER	Same as ER
HER2, from Dako	The Herceptest ™ antibody, 1:8000 dilution for 1 hour	Scored using a 4 point ordinal scale	2+ and 3+ considered positive
EGFR(HER1), from Dako	clone H11, 1:50 dilution for 1 hour	Scored using a 4 point ordinal scale	0 considered negative, any evidence of staining considered positive.

ER, PR HER1 and HER2 Immunofluorescence (AQUA)

Assay and Cutoff: The methods and primary reagents used for ER, PR HER2 and HER1 will be the same for the AQUA-based studies as the IHC studies. The secondary detection reagents will be fluorescence-based as required for the AQUA technology. The analysis method results in scores that are directly proportional to protein concentration. The scores will be normalized by simultaneous analysis of standard cell lines from which a standard curve will be constructed; more specifically, a standard set of 8-10 breast cancer cell lines are included on each tissue or array slide to control for slide to slide variability. Optimal cut points will be determined by different methods depending on the issue in question. For example to evaluate outcome, optimal cutpoints will be derived from historical cohorts from previous publications using AQUA [1,36].

This allows use of the entire cohort for analysis. For response to therapy, the cohort will be divided into a training set and a validation set using X-tile software to find, and then test the optimal cut-point [35].

COX-2 by IHC and AQUA

Background: Cox-2 is over expressed in a proportion of breast cancers [30,37-40]. Prostaglandins, which are synthesized by the action of phospholipase A2 and cyclo-oxygenases, are synthesized within breast tumours and are involved in tumour invasion [40]. COX-2 expression has been associated with decreased survival of breast cancer patients [41-43]. This association was especially strong in patients who had estrogen receptor (ER) positive tumors. The fact that elevated expression of COX-2 is associated with poor survival in ER-positive tumors is of particular interest, since COX-2-derived prostanoids have been shown to be potent local inducers of stromal cell aromatase expression [44,45]. It is thus possible that elevated COX-2 expression in ER-positive cancers could enhance a growth-promoting microenvironment for the tumor cells by inducing estrogen production via the aromatase pathway in the stromal cells.

Assay and Cutoff: Ari Ristimaki will perform IHC and the subsequent analysis using methodology and scoring system as previously described [41]. A tissue microarray slide will be subjected to microwave antigen retrieval followed by immunostaining with a COX-2-specific antihuman mouse monoclonal antibody [41]. Evaluation of COX-2 immunostaining of tumor cells will be performed using a combined staining intensity and proportional scoring system that was employed to demonstrate its prognostic significance in human breast cancer [41]. Dr. Rimm will perform COX-2 immunofluorescence with the identical COX-2 antibody as that used by Dr. Ristimaki and will quantitate COX-2 expression in the tumor cells using AQUA. To evaluate outcome and response to therapy, the cohort will be divided into a training set and a validation set using X-tile software to find, and then test, the optimal cut-point [43]. Our cohort size should be sufficient as evidenced by our prior work in this area [43].

Aromatase by IHC and AQUA

Background: Aromatase is expressed in both epithelial breast cancer cells and also in peri-tumoral stromal cells [46,47]. Aromatase is controlled by tissue specific promoters allowing differences in estrogen synthesis locally in different tissues. Tumors are able to “de-silence” silencers [48-51], in tumoral and peri-tumoral cells which then allows aromatase to be upregulated, synthesize estrogen and potentially promote tumor growth in an intracrine and paracrine fashion. It is known that intra-tumoral estrogen levels are suppressed by letrozole given in the neoadjuvant setting [9,52]. Results from studies correlating intra-tumoral aromatase to patient outcome have been mixed and hampered by the limitations of inadequate antibodies as indicated in our references. A panel of antibodies has now been raised against aromatase through an international effort led by Dr Sasano, an international consortium of collaborators and Novartis to determine the optimal antibody. Dr. Sasano has recently published preliminary data for aromatase IHC methodologies and a scoring system for aromatase immunoreactivity in formalin fixed paraffin-embedded tissues and correlated the IHC results with aromatase biochemical activity (please see attached reference [53]. Briefly, these methodologies combine standard biotin-streptavidin detection with a scoring system that consists of IHC staining proportion and intensity; this scoring system is that devised by Dr. Craig Allred for routine evaluation of steroid receptors [54]. The results of Dr. Sasano's work demonstrate that aromatase immunoreactivity can be detected in different tissue compartments (normal and malignant epithelium, normal and tumor-associated stroma) but a significant positive correlation was detected only in malignant epithelium.

Assay and cutoff: Dr. Sasano will perform the IHC staining and score aromatase immunoreactivity for both the normal and malignant epithelial and stromal compartments of each tissue section using the aforementioned published methodology [53]. Dr. Rimm will perform aromatase immunofluorescence with the monoclonal antibody (aromatase antibody 677) provided by Dr. Sasano and quantitate aromatase staining in the different tissue compartments using AQUA. To evaluate outcome and response to therapy, the cohort will be divided into a training set and a validation set using X-tile software to find, and then test the optimal cut-point [35]. Our cohort size should be sufficient as evidenced by our prior work in this area [35].

GATA3 and NAT1 by IHC and AQUA

Background: GATA3 and NAT1 are the products of a TMA-based discovery study that seems highly promising and can be used, in combination with ER to predict recurrence, and possibly response to hormonal therapy. Dr Rimm's lab has recently completed a series of studies toward the goal of defining a minimal set of multiplexed immunofluorescence-based assays that could be used to predict responsiveness to hormonal therapies. Preliminary data on GATA3 and NAT1, is as follows (See Appendix , Figs 1 and 2). A series of 45 markers were selected from classification studies in the literature and a training cohort of breast cancer cases (collected between 1962 and 1982) was analyzed using the AQUA™ method of in situ quantitative analysis of protein expression. Using the unsupervised clustering algorithm (TreeView) we defined a subset of markers, most closely related to estrogen receptor, including NAT1, GATA3, ACADSB, TFF3, XBP1, KRT8 and KRT18. Since the ultimate goal was to perform this assay on a single slide, we selected ER, NAT1 and GATA3 since these 3 were the most informative of this group. Neither NAT1 nor GATA3 are classically associated with hormonal response. GATA3 (Gata binding protein 3) is a transcription factor with a key role in T-cell lineage development and differentiation of naïve CD4+ T-cells into Th2 effector cells and NAT1 (N-acetyltransferase 1) is an enzyme responsible for the N-acetylation of certain arylamine drugs. We then did a clustering analysis of our full cohort (n=545) using only these 3 markers. This analysis, shown in figure 1 defines 4 expression groups. Cluster 1 represents the cases with low levels of ER while the cases with higher levels of ER are sub-divided into 3 groups. Figure 2 shows the K-M survival curves of each of these groups. It is interesting to note that the conventional ER negative cases are predominantly in group 1, but this group also contains some cases that were called ER positive by the pathologist. Groups 2-4 were predominantly called ER positive by pathologists, but this analysis divides that group into 3 distinct groups with different 5 year survivals. Our confidential manuscript (currently under review) describing the aforementioned work in greater detail is attached [55] Importantly, our work shows how genetic algorithms can be used in combination with multiplexed, AQUA based assessment of protein expression to define patient subsets by outcome. The optimal solution defined by a historical training set finds a subset of patients with 96% 5 year survival in an independent validation set. It has prognostic value similar to the Oncotype Dx and HoxB13/IL17 ratio tests, and thus is ideal to include in this study in order to compare the value of each test to one another.

Assay and cutoff: NAT1 and GATA3 will be assessed using routine immunohisto-chemistry methods as described above for ER, PR, Her1 and Her2. AQUA will be performed as described in the attached confidential manuscript (1). To evaluate outcome, the cutpoint for NAT1 and GATA3 will be derived from our work using AQUA as described in the attached appendix (1). This cutoff allows for the use of the entire cohort for analysis. To evaluate response to therapy, the cohort will be divided into a training set and a validation set using X-tile software to find, and then test the optimal cut-point [35].

Novel gene expression signature discovery

Background: The current application of gene expression technologies to clinical samples has been limited to RNA derived from fresh tissue or cells, or from archival fresh frozen tissues. Given this technical limitation and the fact that large archival frozen tissue repositories with long term clinical outcome data are scarce worldwide, the rapid and widespread use of gene expression profiling technologies to address many clinically and biologically relevant issues in human breast cancer has been significantly hampered. One potential solution to this problem is the use of formalin fixed, paraffin-embedded (FFPE) resected breast tissue samples. FFPE breast tissues have been collected worldwide throughout decades of routine histopathological examination and are a potentially invaluable resource for investigative gene expression profile studies. Given the wide availability of FFPE breast tissue blocks along with linked long-term clinical follow-up data, clinical and biological paradigms can be rapidly studied retrospectively. Therefore, the use of FFPE breast cancer tissues corresponding to well-defined clinical cohorts, such as the MA.17 patients, with long term follow-up data provides a major leap in our ability to expedite the comprehensive evaluation of gene expression technologies in the clinical setting of breast cancer treatment.

Although previous studies have demonstrated that nucleic acids can be retrieved from FFPE tissue samples, demonstration of reliable quantitation of gene expression from such specimens has been limited to the use of real-time quantitative Taqman reverse-transcriptase polymerase chain reaction (RTQ-PCR) technologies. Recently, we have explored the possibility of generating gene expression profiles using RNA derived from FFPE tissues samples. Our preliminary data (see below) indicate that such profiles can be obtained and that relevant gene expression signatures can be extracted from such material.

The Assay: The assay for gene expression profiling of FFPE tumor tissues is as follows. RNA will be extracted from FFPE tissue sections and subjected to T7-based RNA amplification using the Paradise Reagent System (Arcturus Biosciences Inc.) as described [27], labeled and hybridized to the Affymetrix X3P GeneChips. Since the average length of the oligo-dT primed aRNA derived from FFPE tissues is consistently 100 bps, the X3P GeneChips are the preferred microarray as these GeneChips were designed to maximize broad hybridization coverage in this setting. More specifically, X3P GeneChip was designed such that greater 95% of the 25-mers represented sequences located less than 300bps, from the polyA tail of the represented gene. The resulting hybridization data will be subjected to analysis using the GeneMaths and S-PLUS software similar to that described [27].

The generation of gene expression profiles from FFPE tissues is an evolving method. We have significant preliminary data that suggest this approach is valid. Our preliminary data to date is as follows:

1. We have assessed the quality of RNA derived from FFPE breast tissue blocks that had been routinely processed and stored over a fourteen year period. We isolated and amplified RNA derived from 300 FFPE samples that were selected from the MGH and Mayo Clinic. From these 300 samples, we were successful in generating high quality aRNA (90 bps fragments or greater) in ~85% of all cases.
2. Importantly, gene expression data generated with RNA derived from duplicate microdissected FFPE tissue samples is highly reproducible (Fig 3 Appendix). The correlation coefficient (r) of microarray data generated from multiple pairs of duplicate microdissected FFPE samples is consistently above 0.95.

3.We have demonstrated that the gene expression profiles from matched frozen and FFPE breast cancer samples correlate with each other. More specifically, in 6 frozen and FFPE matched pairs the profiles from the FFPE and frozen samples from the same breast tumor had greater similarity to each other than to profiles generated from tumor samples from different patients (Figure 4, Appendix). Inspection of tumor matched FFPE and frozen tumor-derived profiles demonstrate similar gene expression patterns. Some differences do exist, however the high degree of similarity supports the notion that RNA from FFPE samples can be used for gene expression profiling. Comparing gene expression profiles derived from greater than 400 estrogen receptor (ER) positive and estrogen receptor negative frozen breast tumor samples, a robust estrogen receptor gene expression signature consisting of 488 genes has been identified. We have recently performed gene expression profiling of 14 different archived (from 1990 to 2003) FFPE breast cancer samples and demonstrate that 475 out of 488 (97.3%) of the ER signature genes clustered correctly in these FFPE samples (Figure 5, Appendix). This indicates that a gene expression signature derived from frozen tissue samples can be identified from gene expression profiles derived from FFPE tissue samples of varying age.

We acknowledge that gene expression profiling using FFPE tissue is not as robust as that performed with frozen tissue. Thus, there is a possibility that a novel predictive signature may not be identified. However, we believe that our aforementioned findings (see section above and data in appendix) demonstrate sufficient feasibility to allow for microarray-based interrogation of the MA.17 cohort.

Histopathological review.

In conjunction with the NCIC CTG/NCCTG /CALGB/SWOG/ECOG, the formalin-fixed paraffin embedded (FFPE) tumor tissue blocks and slides will be retrieved and submitted to a central repository. For each case, a single slide will be stained with hematoxylin and eosin and subjected to centralized standard histopathological assessment that will include evaluation of tumor type (ductal, lobular, other) and tumor grade as described [28,21]. Three breast pathologists at Harvard Medical School (Dr Dennis Sgroi, Dr Frederick Koerner and Dr Melinda Fan) will independently evaluate the tumors. For cases in which there is a discrepancy in subtyping or grading, all three pathologists will arrive at a consensus through simultaneous review using a multi-headed microscope.

Preliminary results

Briefly summarize preliminary results from prior correlative studies that suggest what utility this marker might have (for example, predicts resistance to taxanes).

NOTE: Please see appendix and enclosed PDF files of recent publications demonstrating proof-of-principle for the RTQ-PCR and based-based assays.

Choice of trial

Briefly summarize the clinical trial from which you wish to receive samples, and why this marker should be studied using these samples.

It is increasingly recognized that hormone receptor positive breast cancer carries a prolonged risk of recurrence beyond 5-years of diagnosis. The MA.17 trial cohort provides an unprecedented opportunity to assess the predictive value of our aforementioned two-gene signature in patients who develop late breast cancer recurrence, i.e. after 5 years of successful disease control with adjuvant tamoxifen. Using the MA.17 cohort, we propose to assess the prognostic and predictive

utility of the MGH two-gene expression signature and the GHI 21-gene expression to identify women who are at risk for late tumor recurrence and to assess the ability of this signature to predict for responsiveness or lack of responsiveness to letrozole. Furthermore, we propose to assess the predictive utility of ER, PR, Her-2, Her-1, aromatase and Cox-2 protein expression levels through the use of standard immunohistochemical and advanced quantitative immunofluorescence technologies. Depending upon the results of these studies, we will have the tools to identify new gene and protein expression biomarkers that may predict the risk of late relapse following tamoxifen therapy, or the risk of relapse following both tamoxifen and letrozole therapy.

Funding

Please list any proposed funding sources.

Massachusetts General Hospital/ Avon Foundation Partnership Grant.

NCIC CTG Core Tumor Banking Funds

References:

1. Camp RL, Chung GG, Rimm DL. Automated subcellular localization and quantification of protein expression in tissue microarrays. Nat Med 2002;8:1323-7.
2. Ramaswamy S, Ross KN, Lander ES, Golub TR. A molecular signature of metastasis in primary solid tumors. Nat Genet. 2003;33:49-54.
3. Group EBCTC. Tamoxifen for Early Breast Cancer. Cochrane Database Syst Rev, CD000486.
4. Bardou VJ, Arpino G, Elledge RM, Osborne CK, Clark GM. Progesterone Receptor Status Significantly Improves Outcome Prediction Over Estrogen Receptor Status Alone for Adjuvant Endocrine Therapy in Two Large Breast Cancer Databases. J Clin Oncol 2003;21:1973-9.
5. Nicholson RI, Gee JM, Knowlden J, McClelland R, Madden TA, Barrow D, Hutcheson I. The Biology of Antihormone Failure in Breast Cancer. Breast Cancer Res Treat 2003;80 Suppl 1:S29-S34.
6. Osborne CK Schiff R. Growth Factor Receptor Cross-Talk With Estrogen Receptor As a Mechanism for Tamoxifen Resistance in Breast Cancer. Breast 2003;12:362-7.
7. Clarke R, Liu MC, Bouker KB, Gu Z, Lee RY, Zhu Y, Skaar TC, Gomez B, O'Brien K, Wang Y, Hilakivi-Clarke LA. Antiestrogen Resistance in Breast Cancer and the Role of Estrogen Receptor Signaling. Oncogene 2003;22:7316-39.
8. Goss PE, Ingle JN, Martino S, Robert NJ, Muss HB, Piccart MJ, Castiglione M, Tu D, Shepherd LE, Pritchard KI, Livingston RB, Davidson NE, Norton L, Perez EA, Abrams JS, Therasse P, Palmer MJ, Pater JL. A Randomized Trial of Letrozole in Postmenopausal Women After Five Years of Tamoxifen Therapy for Early-Stage Breast Cancer. N Engl J Med 2003;349:1793-1802.
9. Ellis MJ, Coop A, Singh B, Mauriac L, Llombert-Cussac A, Janicke F, Miller WR, Evans DB, Dugan M, Brady C, Quebe-Fehling E, Borgs M. Letrozole Is More Effective Neoadjuvant Endocrine Therapy Than Tamoxifen for ErbB-1- and/or ErbB-2-Positive, Estrogen Receptor-Positive Primary Breast Cancer: Evidence From a Phase III Randomized Trial. J Clin Oncol 2001;19:3808-16.
10. Buzdar AU. Anastrozole: a New Addition to the Armamentarium Against Advanced Breast Cancer. Am J Clin Oncol 1998;21:161-6.
11. Colleoni M, Curigliano G, Minchella I, Peruzzotti G, Nole F, Mazzarol G, Renne G, Orlando L, Rocca A, Veronesi P, Intra M, Viale G, Sandri MT, Severi G, Goldhirsch A. Preoperative and Perioperative Chemotherapy With 5-Fluorouracil As Continuous Infusion in Operable

Breast Cancer Expressing a High Proliferation Fraction: Cytotoxic Treatment During the Surgical Phase. Ann Oncol 2003;14:1477-83.
12. Morris C. The Role of EGFR-Directed Therapy in the Treatment of Breast Cancer. Breast Cancer Res Treat 2002;75 Suppl 1:S51-S55.
13. Lichtner RB. Estrogen/EGF Receptor Interactions in Breast Cancer: Rationale for New Therapeutic Combination Strategies. Biomed Pharmacother 2003;57:447-51.
14. Atalay G, Cardoso F, Awada A, Piccart MJ. Novel Therapeutic Strategies Targeting the Epidermal Growth Factor Receptor (EGFR) Family and Its Downstream Effectors in Breast Cancer. Ann Oncol 2003;14:1346-63.
15. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson J Jr., Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Levy R, Wilson W, Grever MR, Byrd JC, Botstein D, Brown PO, Staudt LM. Distinct Types of Diffuse Large B-Cell Lymphoma Identified by Gene Expression Profiling. Nature 2000;403:503-11.
16. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES. Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring. Science 1999;286:531-7.
17. Nutt CL, Mani DR, Betensky RA, Tamayo P, Cairncross JG, Ladd C, Pohl U, Hartmann C, McLaughlin ME, Batchelor TT, Black PM, von Deimling A, Pomeroy SL, Golub TR, Louis DN. Gene Expression-Based Classification of Malignant Gliomas Correlates Better With Survival Than Histological Classification. Cancer Res 2003;63:1602-7.
18. Bittner M, Meltzer P, Chen Y, Jiang Y, Seftor E, Hendrix M, Radmacher M, Simon R, Yakhini Z, Ben Dor A, Sampas N, Dougherty E, Wang E, Marincola F, Gooden C, Lueders J, Glatfelter A, Pollock P, Carpten J, Gillanders E, Leja D, Dietrich K, Beaudry C, Berens M, Alberts D, Sondak V. Molecular Classification of Cutaneous Malignant Melanoma by Gene Expression Profiling. Nature 2000;406:536-40.
19. Perou CM, Sorlie T, Eisen MB, van de RM, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO, Botstein D. Molecular Portraits of Human Breast Tumours. Nature 2000;406:747-52.
20. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de RM, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Eystein LP, Borresen-Dale AL. Gene Expression Patterns of Breast Carcinomas Distinguish Tumor Subclasses With Clinical Implications. Proc Natl Acad Sci U S A 2001;98:10869-74.
21. Ma XJ, Salunga R, Tuggle JT, Gaudet J, Enright E, McQuary P, Payette T, Pistone M, Stecker K, Zhang BM, Zhou YX, Varnholt H, Smith B, Gadd M, Chatfield E, Kessler J, Baer TM, Erlander MG, Sgroi DC. Gene Expression Profiles of Human Breast Cancer Progression. Proc Natl Acad Sci U S A 2003;100:5974-9.
22. Hedenfalk I, Duggan D, Chen Y, Radmacher M, Bittner M, Simon R, Meltzer P, Gusterson B, Esteller M, Kallioniemi OP, Wilfond B, Borg A, Trent J. Gene-Expression Profiles in Hereditary Breast Cancer. N Engl J Med 2001;344:539-48.
23. Sgroi DC, Teng S, Robinson G, LeVangie R, Hudson JR Jr, Elkahloun AG. In Vivo Gene Expression Profile Analysis of Human Breast Cancer Progression. Cancer Res 1999;59:5656-61.
24. van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der KK, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH. Gene Expression Profiling Predicts Clinical Outcome of Breast Cancer. Nature 2002;415:530-6.

25. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, Parrish M, Atsma D, Witteveen A, Glas A, Delahaye L, van d V, Bartelink H, Rodenhuis S, Rutgers ET, Friend SH, Bernards R. A Gene-Expression Signature As a Predictor of Survival in Breast Cancer. N Engl J Med 2002;347:1999-2009.
26. Chang JC, Wooten EC, Tsimelzon A, Hilsenbeck SG, Gutierrez MC, Elledge R, Mohsin S, Osborne CK, Chamness GC, Allred DC, O'Connell P. Gene Expression Profiling for the Prediction of Therapeutic Response to Docetaxel in Patients With Breast Cancer. Lancet 2003;362:362-9.
27. Ma XJ, Wang Z, Ryan PD, Isakoff SJ, Barmettler A, Fuller A, Muir B, Mohapatra G, Salunga R, Tuggle JT, Tran Y, Tran D, Tassin A, Amon P, Wang W, Wang W, Enright E, Stecker K, Estepa-Sabal E, Smith B, Younger J, Balis U, Michaelson J, Bhan A, Habin K, Baer TM, Brugge J, Haber DA, Erlander MG and Sgroi DC. A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen. Cancer Cell 2004;5:607-16.
28. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, Baehner FL, Walker MG, Watson D, Park T, Hiller W, Fisher ER, Wickerham DL, Bryant J, Wolmark N. A Multigene Assay to Predict Recurrence of Tamoxifen-Treated, Node-Negative Breast Cancer. N Engl J Med 2004 Dec 10; [Epub ahead of print].
29. Langholz, Goldstein. Risk set sampling in epidemiologic cohort studies. Statistical Science 1996, p.35-53.
30. Hwang D, Schmitt WA, Stephanopoulos G, Stephanopoulos G. Determination of minimum sample size and discriminatory expression patterns in microarray data. Bioinformatics 18:1184-93, 2002.
31. M Pepe. The Statistical Evaluation of Medical Tests for Classification and Prediction, page 259 (Oxford Press 2003).
32. Cantile M, Pettinato G, Procino A, Feliciello I, Cindolo L, Cillo C. In Vivo Expression of the Whole HOX Gene Network in Human Breast Cancer. Eur J Cancer 2003;39:257-64.
33. Jung, C, Kim R, Lee S, Wang C, Jeng M. HOXB13 homeoprotein suppresses the growth of prostate cancer cells by negative regulation of T-cell factor 4. Cancer Research 2004;64: 3046-51.
34. Dolled-Filhart M, Camp RL, Kowalski DP, Smith BL, Rimm DL. Tissue microarray analysis of signal transducers and activators of transcription 3 (Stat3) and phospho-Stat3 (Tyr705) in node-negative breast cancer shows nuclear localization is associated with a better prognosis. Clin Cancer Res. 2003;9:594-600.
35. Camp RL, Dolled-Filhart M, Rimm DL. X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. Clin Cancer Res. 2004;10:7252-9.
36. Camp RL, Dolled-Filhart M, King BL, Rimm DL. Quantitative Analysis of Breast Cancer Tissue Microarrays Shows That Both High and Normal Levels of HER2 Expression Are Associated with Poor Outcome. Cancer Res 2003;63:1445-48.
37. O'Neill GP and Ford-Hutchinson AW. Expression of mRNA for cyclooxygenase-1 and cyclooxygenase-2 in human tissues. FEBS Lett 1993;330(2):156-60.
38. Parrett ML, Harris RE, Joarder FS, Ross MS, Clausen KP, Robertson FM. Cyclooxygenase-2 expression in human breast cancer. Int J Oncol 1997;10:503–7.
39. Hwang D, Scollard D, Byrne J, Levine E. Expression of cyclooxygenase-1 and cyclooxygenase-2 in human breast cancer. Journal of the National Cancer Institute 1998;90(6):455-60.
40. Takahashi Y, Kawahara F, Noguchi M, Miwa K, Sato H, Seiki M, Inoue H, Tanabe T, Yoshimoto T. Activation of matrix metalloproteinase- 2 in human breast cancer cells overexpressing cyclooxygenase- 1 or -2. FEBS Lett 1999;460:145–48.

41. Ristimaki A, Sivula A, Lundin J, Lundin M, Salminen T, Haglund C, Joensuu H, Isola J. Prognostic significance of elevated cyclooxygenase-2 expression in breast cancer. Cancer Res 2002;62:632–35.
42. Denkert C, Winzer KJ, Muller B, Weichert W, Pest S, Kobel M, Kristiansen G, Reles A, Siegert A, Guski H, Hauptmann S. Elevated expression of cyclooxygenase-2 is a negative prognostic factor for disease free survival and overall survival in patients with breast carcinoma. Cancer 2003;97:2978–87.
43. Spizzo G, Gastl G, Wolf D, Gunsilius E, Steurer M, Fong D, Amberger A, Margreiter R, Obrist P: Correlation of COX-2 and Ep-CAM overexpression in human invasive breast cancer and its impact on survival. Br J Cancer 2003;88:574–78.
44. Brueggemeier RW, Quinn AL, Parrett ML, Joarder FS, Harris RE, Robertson FM. Correlation of aromatase and cyclooxygenase gene expression in human breast cancer specimens. Cancer Lett 1999;140:27–35.
45. Sasano H, Frost AR, Saitoh R, Harada N, Poutanen M, Vihko R, Bulun SE, Silverberg SG, Nagura H. Aromatase and 17 beta-hydroxysteroid dehydrogenase type 1 in human breast carcinoma. J Clin Endocrinol Metab 1996;8(11):4042-6.
46. Miller WR, O'Neil J. The importance of local synthesis of estrogen with the breast, Steroids 1987;50:537–48.
47. Sasano H, Harada N. Intratumoral aromatase in human breast endometrial, and ovarian malignancies. Endocr Rev 1998;19:593–607.
48. Jin T, Zhang X, Li H, Goss PE. Characterization of a novel silencer element in the human aromatase gene PII promoter. Breast Cancer Res Treat 2000;62(2):151-9.
49. Zhou J, Gurates B, Yang S, Sebastian S, Bulun SE. Malignant breast epithelial cells stimulate aromatase expression via promoter II in human adipose fibroblasts: an epithelial-stromal interaction in breast tumors mediated by CCAAT/enhancer binding protein beta Cancer Res 2001;61(5):2328-34.
50. Sebastian S, Takayama K, Shozu M, Bulun SE. Cloning and characterization of a novel endothelial promoter of the human CYP19 (aromatase P450) gene that is up-regulated in breast cancer tissue. Cancer Res 2001;61(5):2328-34.
51. Agarwal VR, Bulun SE, Leitch M, Rohrich R, Simpson ER. Use of alternative promoters to express the aromatase cytochrome P450 (CYP19) gene in breast adipose tissues of cancer-free and breast cancer patients. J Clin Endocrinol Metab 1996;81(11):3843-9.
52. Harper-Wynne CL, Sacks NP, Shenton K, MacNeill FA, Sauven P, Laidlaw IJ, Rayter Z, Miall S, Howes A, Salter J, Hills MJ, Lowe FM, A'Hern R, Nasiri N, Doody D, Iqbal J, Dowsett M. Comparison of the systemic and intratumoral effects of tamoxifen and the aromatase inhibitor vorozole in postmenopausal patients with primary breast cancer. J of Clin Oncol 2002;20(4):1026-35.
53. Sasano H, Andersen TJ, Silverberg SG, Evans DB, Krause A, Edwards DP, Santen RJ, Bhatnager AS and Miller WR. J Steroid Biochem Mol Biol 2005;95(1-5):35-9.
54. Allred DC, Harvey JM, Berardo M, Clark GM. Prognostic and predictive factors in breast cancer by immunohistochemical analysis. Modern Pathology 1998;11(2):155-68.
55. Dolled-Filhart M, et al. 2005 *in submission*

* * *

Appendices

You may provide further detail on your proposed correlative study in appendices to this form.

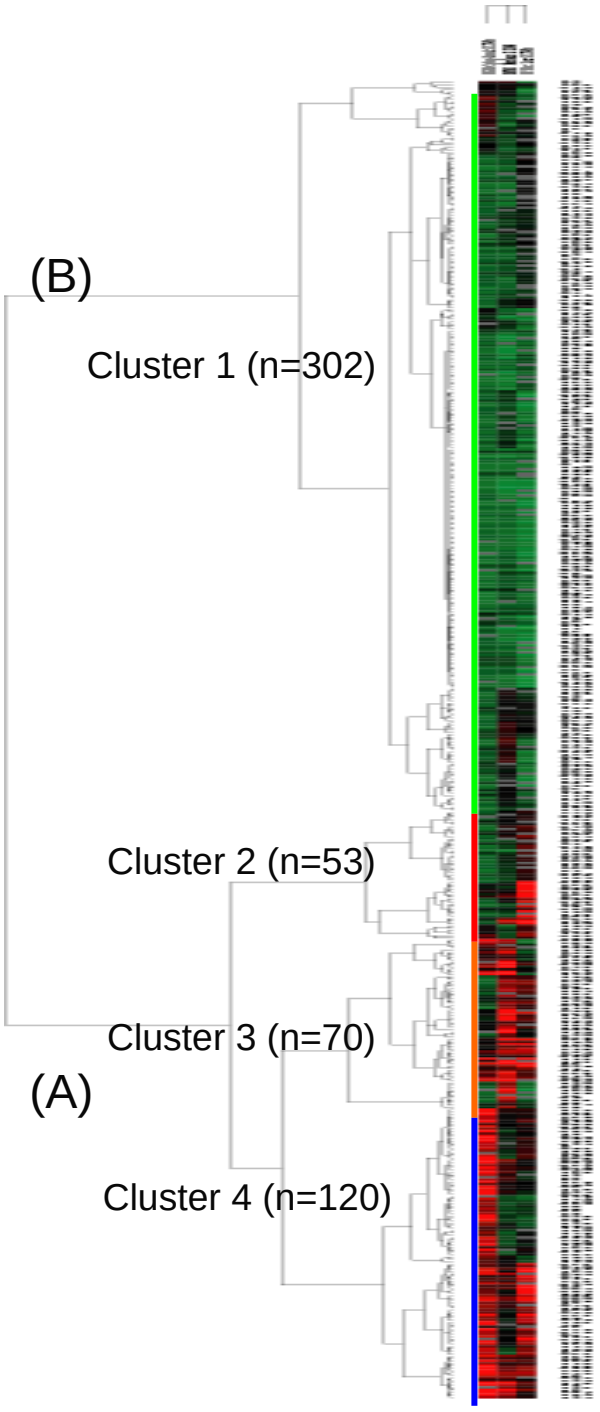


Figure 1

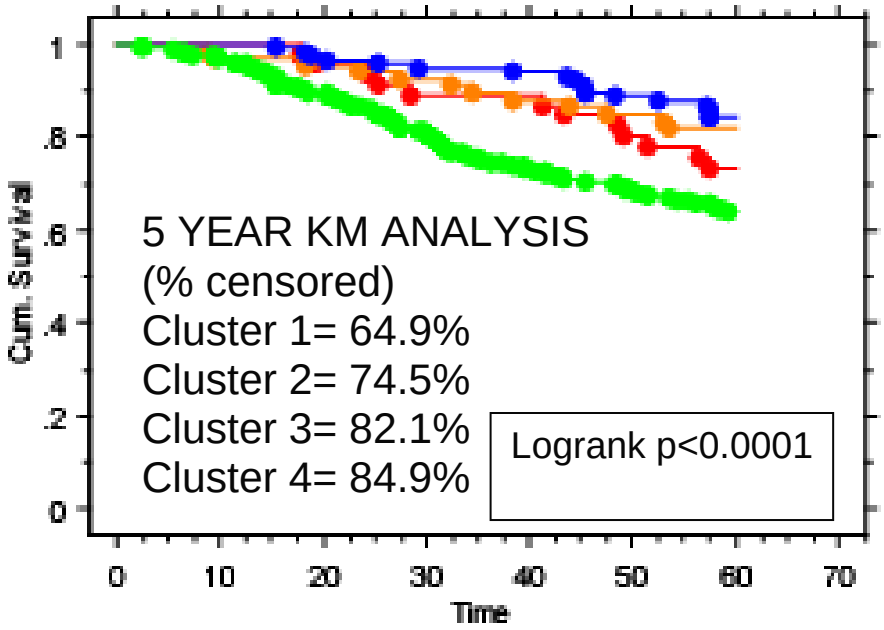


Figure 2

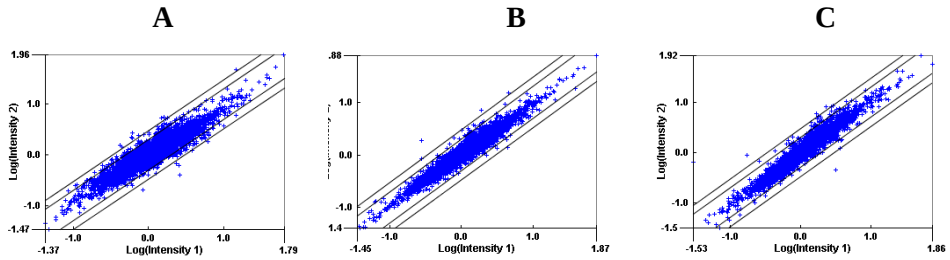


Figure 3. Scatter plots of microarray data for duplicate hybridations of amplified RNA from three different FFPE breast samples to an Affymetrix X3P GeneChip. The correlation (r values) for duplicates in FFPE sample A, r = 0.930; in FFPE sample B, r = 0.971; and for FFPE sample C, r = .968.

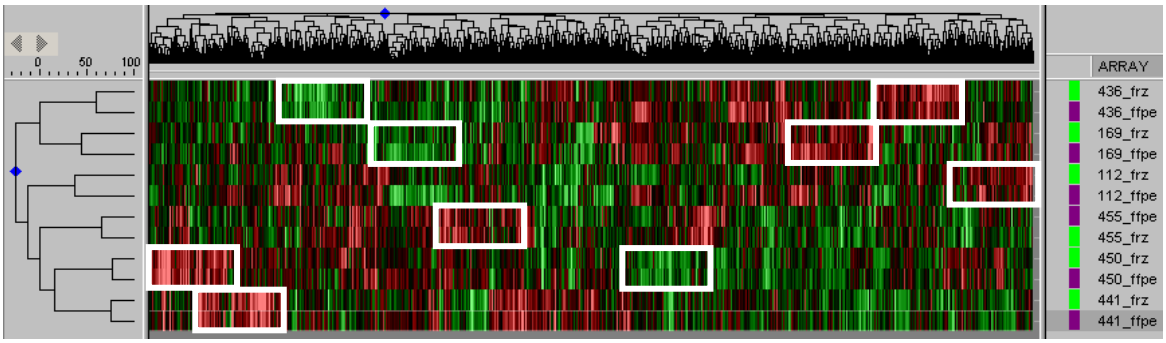


Figure 4. Heat map of the overall expression profile of 11255 genes among 6 matched FFPE and frozen samples (from 6 patients). Matched tissue samples (frz and ffpe represent frozen and formalin fixed paraffin embedded tissues, respectively; numbers represent different tumor samples) are on the Y-axis and the 11255 genes are along the X-axis. The FFPE and frozen samples from the same breast cancer clustered together (having the highest similarity to each other). Representative examples of clusters of highly similar gene expression within a FFPE-Frozen tissue pair are depict as white rectangles.

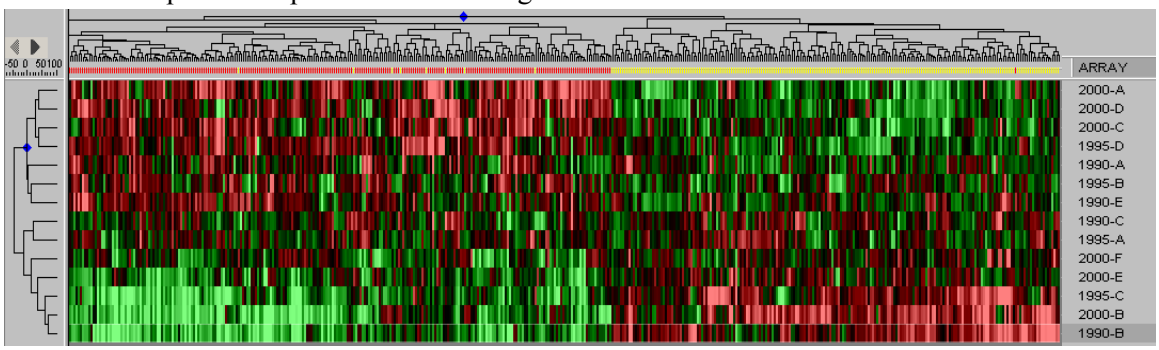


Figure 5. Heat map of the expression of 488 ER genes in 14 FFPE breast cancers. The 14 individual tumor samples are on the Y-axis (number depicts archival year for FFPE specimen) and the 488ER genes are on the Y-axis. The 488 ER signature genes were identified from frozen breast cancers, with those genes expressed higher in ER+ cases labeled red in columns and those higher in ER- yellow. 475 out of the 488 (97.3%) ER signature genes clustered correctly in these FFPE samples.