Prediction of Outcome in Breast Cancer Patients Using Gene Expression Profiling

Daniel S. Oh

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Curriculum in Genetics and Molecular Biology.

Chapel Hill
2006

Approved by

_____
Advisor: Dr. Charles Perou

_____
Dr. Albert Baldwin

_____
Dr. Channing Der

_____
Dr. Mayetri Gupta

_____
Dr. Beverly Mitchell

ABSTRACT

Daniel S. Oh

Prediction of Outcome in Breast Cancer Patients Using Gene Expression Profiling

(Under the Direction of Charles M. Perou)

Breast cancer, the most common cancer diagnosed in women, is a complex and heterogeneous disease. In order to make the best treatment decision for a breast cancer patient, it is important to accurately determine that patient's risk of recurrence and the therapy to which that patient's tumor is most likely to respond. The prognostic and/or predictive factors currently accepted for use in primary breast cancer decision making (i.e. lymph node status, tumor size, nuclear grade, etc.) are not enough to accurately identify those patients who may require therapy and gives little information about what therapy they might best benefit from. Recent discoveries using gene expression profiling have greatly improved our understanding of the molecular pathogenesis of breast cancer. We believe that gene expression profiling may also improve the prognostication and/or prediction of breast cancer outcomes, and thus, the main objective of this work has been to develop and test gene expression-based predictors of outcome in breast cancer patients. First, we developed an expression-based predictor of outcome for Estrogen Receptor (ER) and/or Progesterone Receptor (PR)-positive breast cancer patients using biological differences among these tumors. Second, we developed a predictor for objectively classifying tumors into one of five intrinsic subtypes and validated this using multiple test

sets. Next, using a single patient dataset, we determined the concordance in outcome predictions made by several different gene expression profiles (developed on different platforms by different laboratories). Lastly, we developed gene expression-based predictors for response to neoadjuvant chemotherapy. In summary, this work shows that gene expression profiling holds great promise in being clinically useful in the treatment decision-making process for breast cancer patients.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Chuck Perou, for giving me the opportunity to learn science from one of the best. Without his support and mentoring, I could not have gotten to where I am today. He has been a great example to learn from, and I am privileged to have worked in his lab these past few years. I would like to thank the other members of the Perou Lab for their camaraderie, good spirit, and insightful advice. I thank the members of my thesis committee, Drs. Albert Baldwin, Channing Der, Mayetri Gupta, and Beverly Mitchell, for their invaluable support. I also want to thank the clinicians and others who are not members of the Perou lab, but are critical to the success of the lab's work. Without their expertise, the important discoveries made in our lab would not be possible. I give my final thanks to my family and friends who indirectly have had a hand in the work presented here.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF ABBREVIATIONS

CR	Clinical Complete Response
CV	Cross Validation
DSS	Disease Specific Survival
DWD	Distance Weighted Discrimination
EASE	Expression Analysis Systematic Explorer
ER+	Estrogen Receptor-positive
FDR	False Discovery Rate
GO	Gene Ontology
IFN	Interferon
k-NN	k-Nearest Neighbors
OS	Overall Survival
pCR	Pathologic Complete Response
PD	Progressive Disease
PR+	Progesterone Receptor-positive
PR	Clinical Partial Response
RFS	Relapse Free Survival
RS	Recurrence Score
SAM	Significance Analysis of Microarrays
SD	Stable Disease
SSP	Single Sample Predictor

CHAPTER 1: INTRODUCTION

Breast cancer, the most common cancer diagnosed and the second leading cause of cancer death among women in the US, is a complex and heterogeneous disease. It is critically important for clinicians to accurately determine which breast cancer patients are likely to show recurrence and what treatments they will best benefit from. Currently, the factors widely accepted for use in prognostication and/or prediction for breast cancer patients include axillary lymph node status, age, tumor size, estrogen and progesterone receptor status, and histologic grade[1,2]. Researchers have attempted to find other factors useful for predicting outcomes, but many of these have failed to become clinically useful. There is a need to improve on this set of factors currently used in breast cancer prognosis and prediction, as many patients are either being unnecessarily overtreated or being treated with toxic and expensive chemotherapeutics to which they will not show a response. For example, current guidelines recommend that most lymph node metastasis-negative patients should undergo systemic adjuvant chemotherapy; however, a majority of these node-negative patients are being unnecessarily overtreated because if left systemically untreated, approximately only 20% would develop recurrence[3].

Gene expression profiling using DNA microarrays has recently begun to be used to investigate the heterogeneous nature of breast cancer. It has allowed for a better understanding of the biological differences within breast cancer beyond the traditional methods of patient stratification (i.e. stratification by ER status)[4]. We believe it may also

provide a way to improve the outcome prediction and treatment decision-making for breast cancer patients. Gene expression profiling was used by Perou and colleagues to classify breast cancer into a luminal epithelial, basal-like, normal breast-like, and HER2+/ER- subtypes[4], which were shown to have significant differences in patient outcome[5,6] (Figure 1). This classification was done by hierarchical cluster analysis[7] of tumors using an 'intrinsic' gene set, which consisted of genes with significantly greater variation in expression between tumors than between paired samples from the same tumor. The subtypes identified by Perou et al.[4] were distinguished by extensive differences in expression of genes from the 'intrinsic' gene list (Figure 2). The luminal epithelial subtype of breast cancer was found to be mainly composed of ER+ and/or PR+ tumors and correlated with the high expression of a cluster of genes (luminal epithelial/ER+ gene cluster), which included ER and GATA3. In contrast, the basal-like subtype was found to have low expression of the ER+ gene cluster, but high expression of a cluster of genes that included cytokeratins 5, 6, and 17. The HER2+/ER- subtype also showed low expression of the ER+ gene cluster, but had high expression of HER2 and other genes in the ERBB2 (HER2) amplicon such as GRB7[4]. Thus, these analyses rediscovered important breast cancer biomarkers (i.e. ER and HER2) but also identified new "subtypes" of breast cancer.

Estrogen Receptor Biology and Tamoxifen Resistance. The estrogen receptor (ER) belongs to the superfamily of nuclear hormone receptors that includes the progesterone receptor (PR) and thyroid hormone receptor. ER is essential for mammary gland differentiation and morphogenesis as evidenced by the dramatic phenotypes seen in

ER-knock-out mice[8]. Upon binding to estrogen, ER undergoes a conformational change that results in dissociation from an inhibitory heat shock protein complex, followed by receptor dimerization and phosphorylation[9]. The activated ER then binds to target gene promoters at specific palindromic sequence motifs called estrogen response elements (EREs)[10], which leads to alterations in the transcription of these estrogen-regulated genes. Genes regulated by estrogen have diverse functions, including promotion of the cell cycle and proliferation, cell-cell interactions, angiogenesis, and inhibition of apoptosis[11-14]. In addition, it is postulated that estrogen-regulated genes include those involved in inhibition of tumor cell invasion and metastasis[8].

Positive regulation of estrogen-regulated genes is mediated by two different domains in ER, AF-1 and AF-2. AF-1 is located at the N-terminal of ER and is hormone-independent, while AF-2 is located in the ligand-binding domain of ER and is hormone-dependent[10]. Co-activator proteins such as SRC1, TIF2, and AIB1 form a complex with ER on the promoter to positively regulate transcription, partly through recruitment of histone-acetyltransferases to the promoter, which results in chromatin decondensation[9]. Negative regulation of transcription by ER is controversial and less well understood, and is thought to involve recruitment of corepressor proteins such as NCoR1 and NCoR2 which in turn recruit histone-deacetylase complexes, which results in chromatin condensation[15,16].

Modulation of gene expression by ER has also been shown to occur at non-ERE sites such as AP-1 and SP-1 regulatory sequences[9]. At these regulatory sequences, ER does not directly bind to DNA, but rather is attached to the promoter complex by interaction with other DNA-bound transcription factors (TFs) such as c-jun and c-fos. ER

modulates the activity of these TFs by stabilizing their DNA binding and/or recruiting additional coactivators to the complex[17]. In addition, it is believed that ER may also have non-genomic activity, in which ligand binding to membrane-bound or cytoplasmic ER causes it to interact with a variety of membrane-signaling molecules including key growth-factor receptors and/or growth-factor dependent kinases (i.e. IGF-1R, PI3K, MAPK, EGFR, and HER2)[18,19]. These kinases can then activate signaling cascades that may enhance the activity of nuclear ER and its coregulators via phosphorylation[20].

It is widely believed that estrogen, via the estrogen receptor, is involved in the pathogenesis of and sustains the growth of ER+ breast cancers. Thus, patients with ER+ breast cancer are given estrogen antagonists that block ER activity. The most widely used hormone antagonist is tamoxifen, a selective ER modulator (SERM) that binds ER and blocks its activity in the breast. Tamoxifen is the gold standard of treatment for ER+ breast cancer patients; however, up to 40% of these patients experience relapse despite receiving therapy[21-23].

Postulated tamoxifen resistance mechanisms include loss of ER expression/function and altered expression of a second ER (referred to as ERß) whose function in normal and malignant breast cells is poorly understood[20]. However, other postulated resistance mechanisms have received more attention and are believed to be more important: (1) alterations in co-regulatory proteins such as overexpression of the ER co-activator AIB1 and (2) the upregulation of growth factor signaling pathways such as those mediated by EGFR and HER2 that are believed to engage in cross-talk with ER signaling pathways as described earlier[19,24]. Clinical and laboratory studies suggest that high levels of AIB1 may enhance the partial agonistic effect of tamoxifen and therefore

contribute to tamoxifen resistance[25,26]. Clinical and experimental evidence indicate that overexpression of HER2 and/or EGFR is associated with and contributes to tamoxifen resistance[27,28]. It is hypothesized that one reason HER2 and/or EGFR overexpression contributes to tamoxifen resistance is that the subsequent increased activation of downstream kinases can activate and enhance ER activity by phosphorylating ER and its coregulators[24].

Outcome prediction in hormone-receptor positive breast cancer. Currently, the prognosis of a patient with ER+ and/or PR+ breast cancer can be highly variable. A subgroup of patients with this type of breast cancer (i.e. luminal) experiences disease relapse regardless of receiving therapy while other subsets do extremely well[21-23]. The "intrinsic" classification of Luminal A vs. Luminal B is predictive of outcomes in ER+ and/or PR+ patients; however, this distinction is correlated with grade and we believed it possible to develop a more biologically relevant predictor of outcomes for this patient subset. Not enough progress has been made in developing methods to predict which ER+ and/or PR+ patients are at risk or not at risk for experiencing relapse. Such a method of defining the prognosis of ER+ and/or PR+ breast cancer patients would be of significant value.

Pathways under the control of estrogen signaling (mediated by the estrogen receptor) are involved in the growth and differentiation of the mammary gland and are widely believed to be important in ER+ and/or PR+ breast cancer biology[8,9]. Thus, we speculated that gene expression profiling of the ER pathway using cell line models might be useful in developing a clinically robust outcome predictor for ER+ and/or PR+ breast

cancer and lead to a better understanding of the biological differences within this breast cancer tumor type. Because of the widely believed importance of estrogen responsiveness and signaling in ER+ and/or PR+ breast cancer biology, we hypothesized that estrogen-regulated genes might be useful in predicting outcome in this type of breast cancer, and therefore, we developed an ER-pathway based predictor for breast cancer patients that is described in Chapter 2.

Microarray platform-independent validation of the "intrinsic" breast tumor subtypes. A major challenge for microarray studies, especially those with clinical implications, is validation[29,30]. Due to the practical barriers of cost and access to large numbers of fresh frozen tumor samples with associated clinical information, very few microarray studies have analyzed enough samples to allow promising initial findings to be sufficiently validated to justify the major investment required for clinical testing. An efficient approach would be to use public gene expression data repositories as test sets; however, it has been difficult to compare and/or combine data sets from independent laboratories due to differences in sample preparation, experimental design, and microarray platforms. Fortunately, the multivariate analysis tool Distance Weighted Discrimination (DWD) has recently been shown to successfully overcome this difficulty in comparing and/or combining datasets from independent laboratories[31]. DWD identifies systemic biases present in separate microarray datasets and makes a global adjustment to these datasets to compensate for these biases. If one considers each separate dataset as a multi-dimensional cloud of data points, DWD works by taking two such clouds and shifting one such that it more optimally overlaps the other.

The "intrinsic" subtypes identified by Perou and colleagues (described above) has not yet been validated as a predictor/prognosticator of breast cancer patient outcomes to the extent that other gene expression-based predictive/prognostic profiles have[32,33]. Thus, in Chapter 3 we used DWD to (1) validate the intrinsic subtypes on independent test datasets generated on differing microarray platforms and (2) constructed a method to objectively assign any given tumor to an intrinsic tumor subtype.

Concordance of several different gene expression-based predictors. Many different laboratories have examined the prognostic and/or predictive utility of gene expression profiling for breast cancer. These predictors include the discovery of the "intrinsic" subtypes using supervised hierarchical clustering analysis as done by Perou and colleagues as described above. Independently, van't Veer and colleagues identified a 70-gene prognostic signature that classifies tumors into either a good or poor prognosis group[34]. This signature was obtained through a supervised analysis in which genes were selected according to correlation with patient outcome for a set of 78 node-negative breast cancer patients all less than 55 years of age at time of diagnosis. Genes involved in cell cycle, invasion and metastasis, angiogenesis, and signal transduction were found to be significantly upregulated in the poor prognosis signature[34].

Chang and colleagues[35,36] characterized the transcriptional response of normal fibroblasts to serum and showed that this "wound-response signature" can predict outcomes in breast cancer patients. Patients whose tumors expressed the wound-response signature ("activated" tumors) had markedly poor overall survival and distant metastasis-free survival compared to patients with tumors that did not express this signature

("quiescent" tumors). The wound-response signature included induction of genes involved in the cell cycle, cell motility, extracellular matrix remodeling, and cell-cell signaling[35]. Lastly, two laboratories independently derived expression-based predictors of outcome in ER+ tamoxifen-treated patients. First, Paik and colleagues developed a qRT-PCR-based 21-gene prognostic score (referred to as the "Recurrence score") for node-negative tamoxifen-treated breast cancers, and used it to successfully categorize patients as having a low, intermediate, or high risk of recurrence[37]. Second, Ma and colleagues determined that a 2-gene ratio (HOXB13:IL17BR) could predict disease-free survival in tamoxifen-treated patients: a high ratio indicated a poor clinical outcome compared to a low ratio[38].

These different gene expression-based predictors were developed by different laboratories, and a comparison of the gene lists from some of these predictors showed that they overlapped each other by only a modest amount, if at all. This lower than expected gene similarity between lists might be explained by differences in microarray platforms, cohort biases due to different patient selection criteria and sample size, and differences in statistical methods used to develop the predictors. An important and unanswered question is whether the lack in gene overlap between these predictors reflects an actual discordance in outcome predictions for the individual patient (i.e. do these different predictors actually disagree or agree concerning outcome predictions for the individual patient). To answer this question, in Chapter 4 we describe our studies where we performed an analysis of a single patient dataset on which the five prognostic/predictive gene expression-based predictors described above (intrinsic subtypes[4], van't Veer et al's 70-gene predictor[34], Chang et al's Wound-Response

predictor[35], the Recurrence score predictor[37], and Ma et al's 2-gene ratio predictor[38]) were simultaneously compared.

Prediction of neoadjuvant chemotherapy response. The identification of biomarkers that predict chemotherapy response in breast cancer is an area of intense research, in large part because no such predictor currently exists. In some cases, cell proliferation index or TP53 somatic mutation status has been correlated with response, but these markers are not in standard use for patient stratification[39-42]. Recently, response studies have focused on using DNA microarrays to identify gene expression patterns predictive of chemotherapy response. Neoadjuvant chemotherapy (treatment before primary surgery) facilitates these response studies because it allows for the direct and timely observation of tumor response to treatment and allows access to primary tumor samples before and during treatment to develop and assess markers that might be predictive of response[3].

An additional reason why we may want to study neoadjuvant chemotherapy response is that it has been shown to be correlated with, and can be used as a surrogate marker, for improved long-term disease-free and overall survival[43-49]. For example, in a study of 372 patients given neoadjuvant chemotherapy, Kuerer et al.[44] reported that the 5-year overall and disease-free survival rates were significantly higher in those achieving Pathological Complete Response (pCR) compared to those who did not. Likewise, in a study of 158 patients, Chang et al.[47] reported that good clinical response (defined by Chang et al. as clinical complete response or minimal residual disease) to neoadjuvant chemotherapy was significantly associated with decreased risk of relapse and death.

Thus, encouraged by these neoadjuvant studies, in Chapter 5 we determined whether gene expression patterns from pre-treatment tumor samples could predict response to neoadjuvant chemotherapy.

In summary, this work builds on that of Perou and others regarding the heterogeneity of breast cancer. The overall aim of this work has been to use this knowledge about the heterogeneity of breast cancer to develop/analyze gene expression-based predictors of outcome in breast cancer patients. In all of our aims, we were successful in developing expression-based predictors that provided added value, and thus, we believe that the results of our analyses will improve the outcomes of future breast cancer patients.

REFERENCES

1. Goldhirsch A, Wood WC, Gelber RD, et al: Meeting highlights: updated international expert consensus on the primary therapy of early breast cancer. J Clin Oncol 21:3357-65, 2003

2. Clark GM: Interpreting and integrating risk factors for patients with primary breast cancer. J Natl Cancer Inst Monogr:17-21, 2001

3. Chang JC, Hilsenbeck SG, Fuqua SA: Genomic approaches in the management and treatment of breast cancer. Br J Cancer 92:618-24, 2005

4. Perou CM, Sorlie T, Eisen MB, et al: Molecular portraits of human breast tumours. Nature 406:747-52, 2000

5. Sorlie T, Perou CM, Tibshirani R, et al: Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 98:10869-74, 2001

6. Sorlie T, Tibshirani R, Parker J, et al: Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A 100:8418-23, 2003

7. Eisen MB, Spellman PT, Brown PO, et al: Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 95:14863-8, 1998

8. Platet N, Cathiard AM, Gleizes M, et al: Estrogens and their receptors in breast cancer progression: a dual role in cancer proliferation and invasion. Crit Rev Oncol Hematol 51:55-67, 2004

9. McDonnell DP, Norris JD: Connections and regulation of the human estrogen receptor. Science 296:1642-4, 2002

10. Sanchez R, Nguyen D, Rocha W, et al: Diversity in the mechanisms of gene regulation by estrogen receptors. Bioessays 24:244-54, 2002

11. Lobenhofer EK, Bennett L, Cable PL, et al: Regulation of DNA replication fork genes by 17beta-estradiol. Mol Endocrinol 16:1215-29, 2002

12. Cicatiello L, Scafoglio C, Altucci L, et al: A genomic view of estrogen actions in human breast cancer cells by expression profiling of the hormone-responsive transcriptome. J Mol Endocrinol 32:719-75, 2004

13. Soulez M, Parker MG: Identification of novel oestrogen receptor target genes in human ZR75-1 breast cancer cells by expression profiling. J Mol Endocrinol 27:259-74, 2001

14. Frasor J, Danes JM, Komm B, et al: Profiling of estrogen up- and down-regulated gene expression in human breast cancer cells: insights into gene networks and pathways underlying estrogenic control of proliferation and cell phenotype. Endocrinology 144:4562-74, 2003

15. Nagy L, Kao HY, Chakravarti D, et al: Nuclear receptor repression mediated by a complex containing SMRT, mSin3A, and histone deacetylase. Cell 89:373-80, 1997

16. Dobrzycka KM, Townson SM, Jiang S, et al: Estrogen receptor corepressors -- a role in human breast cancer? Endocr Relat Cancer 10:517-36, 2003

17. DeNardo DG, Kim HT, Hilsenbeck S, et al: Global gene expression analysis of estrogen receptor transcription factor cross talk in breast cancer: identification of estrogen-induced/activator protein-1-dependent genes. Mol Endocrinol 19:362-78, 2005

18. Kelly MJ, Levin ER: Rapid actions of plasma membrane estrogen receptors. Trends Endocrinol Metab 12:152-6, 2001

19. Normanno N, Di Maio M, De Maio E, et al: Mechanisms of endocrine resistance and novel therapeutic strategies in breast cancer. Endocr Relat Cancer 12:721-47, 2005

20. Ring A, Dowsett M: Mechanisms of tamoxifen resistance. Endocr Relat Cancer 11:643-58, 2004

21. Group EBCTC: Tamoxifen for early breast cancer: an overview of the randomised trials. Lancet 351:1451-67, 1998

22. Osborne CK: Tamoxifen in the treatment of breast cancer. N Engl J Med 339:1609-18, 1998

23. Bardou VJ, Arpino G, Elledge RM, et al: 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 21:1973-9, 2003

24. Osborne CK, Schiff R: Estrogen-receptor biology: continuing progress and therapeutic implications. J Clin Oncol 23:1616-22, 2005

25. Webb P, Nguyen P, Shinsako J, et al: Estrogen receptor activation function 1 works by binding p160 coactivator proteins. Mol Endocrinol 12:1605-18, 1998

26. Osborne CK, Bardou V, Hopp TA, et al: Role of the estrogen receptor coactivator AIB1 (SRC-3) and HER-2/neu in tamoxifen resistance in breast cancer. J Natl Cancer Inst 95:353-61, 2003

27. Shou J, Massarweh S, Osborne CK, et al: Mechanisms of tamoxifen resistance: increased estrogen receptor-HER2/neu cross-talk in ER/HER2-positive breast cancer. J Natl Cancer Inst 96:926-35, 2004

28. Ellis MJ, Coop A, Singh B, et al: 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 19:3808-16, 2001

29. Michiels S, Koscielny S, Hill C: Prediction of cancer outcome with microarrays: a multiple random validation strategy. Lancet 365:488-92, 2005

30. Jenssen TK, Hovig E: Gene-expression profiling in breast cancer. Lancet 365:634-5, 2005

31. Benito M, Parker J, Du Q, et al: Adjustment of systematic microarray data biases. Bioinformatics 20:105-114, 2004

32. Paik S, Shak S, Tang G, et al: A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 351:2817-26, 2004

33. van de Vijver MJ, He YD, van't Veer LJ, et al: A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 347:1999-2009, 2002

34. van 't Veer LJ, Dai H, van de Vijver MJ, et al: Gene expression profiling predicts clinical outcome of breast cancer. Nature 415:530-6, 2002

35. Chang HY, Sneddon JB, Alizadeh AA, et al: Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. PLoS Biol 2:E7, 2004

36. Chang HY, Nuyten DS, Sneddon JB, et al: Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. Proc Natl Acad Sci U S A 102:3738-43, 2005

37. Paik S, Shak S, Tang G, et al: Expression of the 21 genes in the Recurrance Score assay and prediction benefits from tamoxifen in NSABP study B-14 and chemotherapy in NSABP study B-20, San Antionio Breast Cancer Symposium. San Antionio, TX, on line abstract, 2004

38. Ma XJ, Wang Z, Ryan PD, et al: A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen. Cancer Cell 5:607-16, 2004

39. Aas T, Borresen AL, Geisler S, et al: Specific P53 mutations are associated with de novo resistance to doxorubicin in breast cancer patients. Nat Med 2:811-4, 1996

40. Aas T, Geisler S, Eide GE, et al: Predictive value of tumour cell proliferation in locally advanced breast cancer treated with neoadjuvant chemotherapy. Eur J Cancer 39:438-46, 2003

41. Geisler S, Lonning PE, Aas T, et al: Influence of TP53 gene alterations and c-erbB-2 expression on the response to treatment with doxorubicin in locally advanced breast cancer. Cancer Res 61:2505-12., 2001

42. Geisler S, Borresen-Dale AL, Johnsen H, et al: TP53 gene mutations predict the response to neoadjuvant treatment with 5-fluorouracil and mitomycin in locally advanced breast cancer. Clin Cancer Res 9:5582-8, 2003

43. Kaufmann M, von Minckwitz G, Smith R, et al: International expert panel on the use of primary (preoperative) systemic treatment of operable breast cancer: review and recommendations. J Clin Oncol 21:2600-8, 2003

44. Kuerer HM, Newman LA, Smith TL, et al: Clinical course of breast cancer patients with complete pathologic primary tumor and axillary lymph node response to doxorubicin-based neoadjuvant chemotherapy. J Clin Oncol 17:460-9, 1999

45. Fisher B, Bryant J, Wolmark N, et al: Effect of preoperative chemotherapy on the outcome of women with operable breast cancer. J Clin Oncol 16:2672-85, 1998

46. Ferriere JP, Assier I, Cure H, et al: Primary chemotherapy in breast cancer: correlation between tumor response and patient outcome. Am J Clin Oncol 21:117-20, 1998

47. Chang J, Powles TJ, Allred DC, et al: Biologic markers as predictors of clinical outcome from systemic therapy for primary operable breast cancer. J Clin Oncol 17:3058-63, 1999

48. Pierga JY, Mouret E, Laurence V, et al: Prognostic factors for survival after neoadjuvant chemotherapy in operable breast cancer. the role of clinical response. Eur J Cancer 39:1089-96, 2003

49. Cleator SJ, Makris A, Ashley SE, et al: Good clinical response of breast cancers to neoadjuvant chemoendocrine therapy is associated with improved overall survival. Ann Oncol 16:267-72, 2005

FIGURES

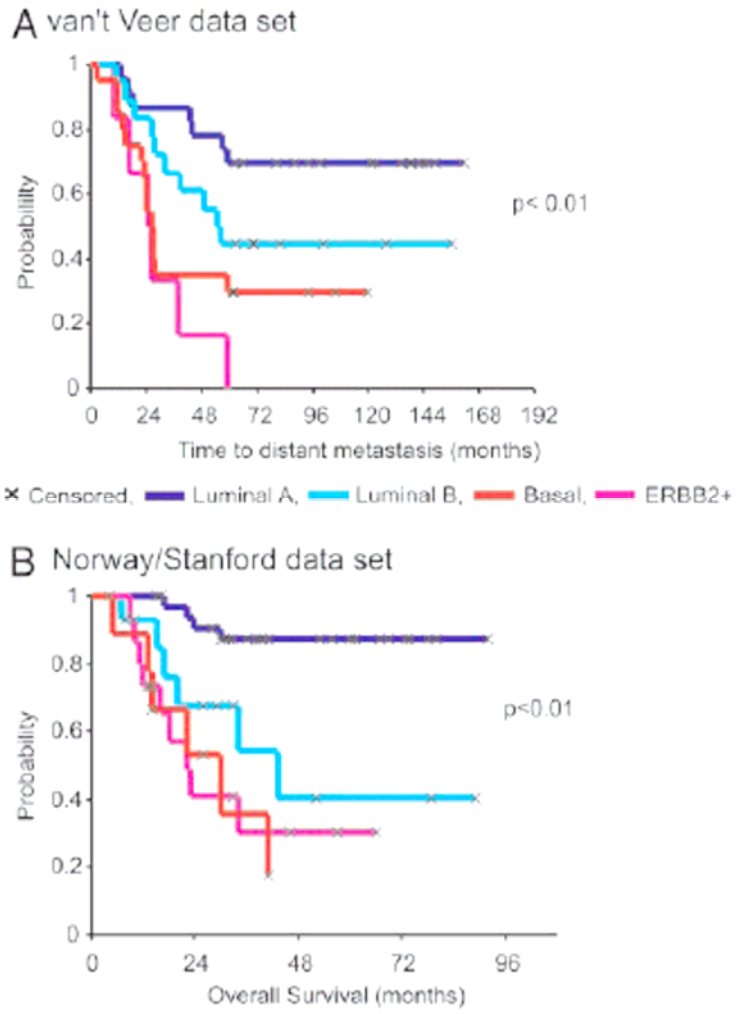


Figure 1.1. Kaplan-Meier analysis of disease outcome in two patient cohorts stratified by intrinsic subtype. **A.** Time to development of distant metastasis in the 97 sporadic cases from van't Veer et al. **B.** Overall survival for 72 patients with locally advanced breast cancer in the Norway cohort. Figure adapted from Sorlie et al.[6]

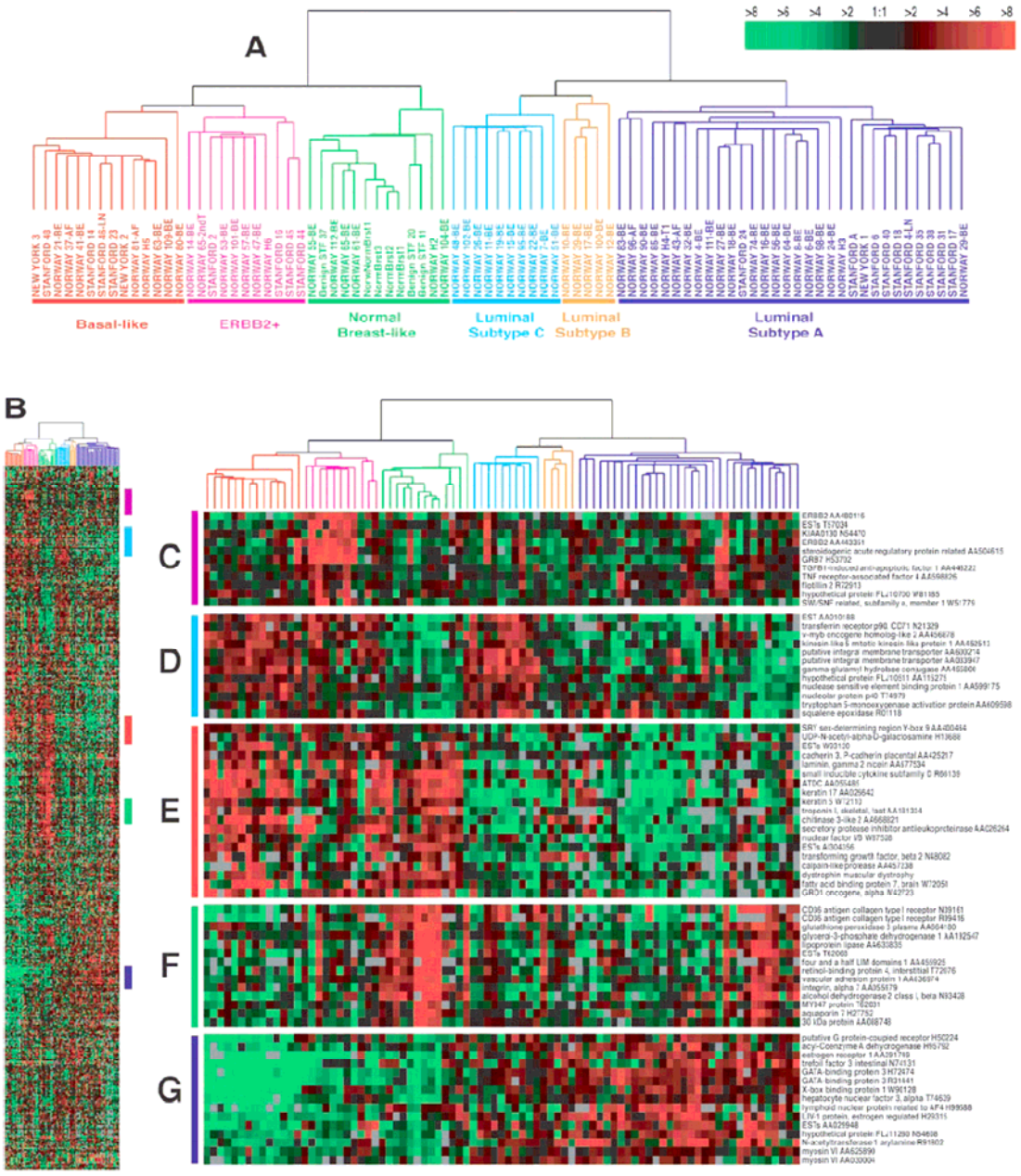


Figure 1.2. Hierarchical cluster analysis using the "intrinsic" gene set. Gene expression patterns of 85 experimental samples (78 carcinomas, 3 benign tumors, and four normal tissues) were analyzed by hierarchical clustering using the 476 cDNA intrinsic clone set. **A.** The tumor specimens were divided into subtypes based on differences in gene expression. The cluster dendrogram showing the subtypes of tumors are colored as: luminal subtype A, dark blue; luminal subtype B, yellow; luminal subtype C, light blue; normal breast-like, green; basal-like, red; and ERBB2+, pink. **B.** Scaled-down representation of the complete cluster diagram. **C.** ERBB2 amplicon cluster. **D.** Novel unknown cluster. **E.** Basal epithelial cell-enriched cluster. **F.** Normal breast-like cluster. **G.** Luminal epithelial gene cluster containing ER. Figure adapted from Sorlie et al.[5]

CHAPTER 2: Estrogen-regulated genes predict survival in hormone receptor-positive breast cancers

Daniel S. Oh, Melissa A. Troester, Jerry Usary, Zhiyuan Hu, Xiaping He, Cheng Fan, Junyuan Wu, Lisa A. Carey and Charles M. Perou

This work was published in the Journal of Clinical Oncology. 2006 Apr 10; 24:1656-64

ABSTRACT

Purpose

The prognosis of a patient with Estrogen Receptor (ER) and/or Progesterone Receptor (PR)-positive breast cancer can be highly variable. Therefore, we developed a gene expression-based outcome predictor for ER+ and/or PR+ (*i.e.* Luminal) breast cancer patients using biological differences among these tumors.

Materials and Methods

The ER+ MCF-7 breast cancer cell line was treated with 17β-estradiol to identify estrogen-regulated genes. These genes were used to develop an outcome predictor on a training set of 65 luminal epithelial primary breast carcinomas. The outcome predictor was then validated on three independent published datasets.

Results

The estrogen-induced gene set identified in MCF-7 cells was used to hierarchically cluster a 65 tumor training set into 2 groups, which showed significant differences in

survival (p=0.0004). Supervised analyses identified 822 genes that optimally defined these two groups, with the poor prognosis Group IIE showing high expression of cell proliferation and anti-apoptosis genes. The good prognosis Group IE showed high expression of estrogen and GATA3-regulated genes. Mean expression profiles (*i.e.* centroids) created for each group were applied to ER+ and/or PR+ tumors from three published datasets. For all datasets, Kaplan-Meier survival analyses showed significant differences in Relapse-Free and Overall Survival between Group IE and IIE tumors. Multivariate Cox analysis of the largest test dataset showed that this predictor added significant prognostic information independent of standard clinical predictors and other gene expression-based predictors.

Conclusion

This study provides new biological information concerning differences within hormone receptor-positive breast cancers and a means of predicting long-term outcomes in tamoxifen-treated patients.

INTRODUCTION

Breast cancers are traditionally stratified into hormone receptor-positive and negative groups to guide patient management. This is because almost all hormone antagonist (*i.e.* tamoxifen)-responsive breast cancers are Estrogen Receptor (ER) and/or Progesterone Receptor (PR)-positive[1]. However, a subgroup of these patients experience disease relapse irrespective of standard therapy with tamoxifen[2-4]. In many cases the resistance mechanism(s) are unknown[5-7]. A method for identifying those ER+ and/or PR+ patients that do well in the presence of tamoxifen versus those that do poorly would be of significant value.

Gene expression profiling is a powerful method for breast cancer prognostication. Using gene expression profiling, breast tumors can be classified into five molecular subtypes (Basal-like, HER2+/ER-, Normal Breast-like, and Luminal A and B) with significant differences in patient outcome[8,9]. The two Luminal subtypes are almost entirely composed of ER+ and/or PR+ tumors and are defined by the high expression of a gene set (luminal epithelial/ER+ set) that includes *ER* and *GATA3*. Compared to Luminal A tumors, Luminal B tumors have poor outcomes despite being clinically ER+[8,9].

To date, several laboratories have developed gene expression-based predictors for ER+ and/or PR+ patients. All used supervised analyses based upon patient outcomes/tumor response[10-13]. Here we developed a gene expression-based predictor using an approach based solely upon biological characteristics of the tumors. Because of the importance of estrogen signaling in breast epithelial cell biology, we hypothesized that differential expression of estrogen-regulated genes would be useful in predicting outcome.

MATERIALS AND METHODS

Cell culture and collection of mRNA. MCF-7 cells were maintained as described previously[14]. Cells were plated in 150mm dishes and grown until 50% confluence. Media was changed and cells maintained for 48 hours in estrogen-free medium (phenol red-free RPMI-1640 with 10% charcoal-dextran-stripped FBS) before treating for 2,4,8,or 24 hours with 10^{-6} M 17β-estradiol (Sigma). Cells were harvested, and mRNA isolated using a Micro-FastTrack kit (Invitrogen). A reference mRNA sample was harvested from cells maintained for 48 hours in estrogen-free medium (*i.e.* estrogen-starved cells).

Microarray Experiments. Agilent Human whole-genome microarrays were hybridized according to manufacturer's protocol with Cy3-CTP labeled cRNA from estrogen-starved cells (2μg/sample) and Cy5-CTP labeled cRNA from 17β-estradiol-treated cells (2μg/sample), with dye-flip replicates for each time point. Microarrays were scanned and image files analyzed as described previously[15]. All primary microarray data are available via the UNC Microarray Database (https://genome.unc.edu/) and the GEO with series number GSE2740 (GSM52882-GSM52909, GSM34423-GSM34568).

Analysis of microarray data to identify GATA3 and estrogen-regulated genes. Data from microarray experiments were calculated as described[15]. Genes were excluded from data analysis if they did not have signal intensity ≥30 in both channels for ≥70% of the experiments. To identify estrogen-regulated genes, we used one-class Significance

Analysis of Microarrays (SAM) to identify genes that changed in all estrogen-treated time points (as a single class) relative to the estrogen-starved cells[16] (Note: in our SAM analyses we did not use the fold-change cutoff option to avoid the fold-change associated complications/pitfalls described by Larsson et al.[17]). Using a false discovery rate (FDR) of 0.04%, SAM identified 383 estrogen-induced and 574 estrogen-repressed genes; for subsequent "estrogen-SAM" analyses, only the 383 induced genes were used. Average linkage hierarchical cluster analysis was conducted and the results visualized in Treeview[18].

GATA3-induced genes were identified by microarray experiments on 293T cells transfected with GATA3 gene constructs, as detailed in Usary *et al.*[19]. One-class SAM analysis (0.58% FDR) identified 407 genes that were induced in the GATA3 samples (as a single class) relative to empty vector controls.

Analysis of primary breast tumor data using the estrogen-induced gene set. The primary breast tumor samples (collected with patient consent and UNC-CH Human Investigations Review Committee approval) used in the training dataset are described in Hu *et al.*[20], except for 14 new tumor samples. A total of 118 fresh frozen breast tumor and 9 normal breast samples represented by 160 microarray experiments were analyzed using the 1300-gene "breast intrinsic" gene set developed by Hu *et al.*[20], which identified 65 tumors as belonging to the "Luminal subtype". These "Luminal" tumors included 61 ER+ and/or PR+ tumors according to immunohistochemistry, 3 ER- and PR-, and one not determined.

The 383-gene MCF-7 estrogen-induced gene list was used to hierarchically cluster the 65 Luminal tumors resulting in two groups, which we called Groups I and II.

We used a two-class, unpaired SAM analysis (with 1% FDR) to identify 822 genes (referred to as the estrogen-SAM list) that optimally differentiated Group I versus Group II tumors[21,22]. The 65 Luminal tumors were then clustered using the 822 genes and two groups (Groups IE and IIE) were evident.

By matching Unigene identifiers, data for as many as possible of the 822 estrogen-SAM genes was obtained for ER+ and/or PR+ tumors (classified as provided in the primary publications) from three independent test datasets[8,9,12,23]. The Ma *et al.*, Sorlie *et al*., and Chang *et al.* datasets consisted of 60, 90, and 250 ER+ and/or PR+ tumors respectively. Ma *et al*. tumors were uniformly treated with adjuvant tamoxifen alone. Sorlie *et al.* tumors received adjuvant tamoxifen, with some also receiving neoadjuvant chemotherapy. Chang *et al.* tumors were heterogeneously treated (http://www.pnas.org/cgi/content/abstract/0409462102v1); 24 of the 250 tumors we used for this dataset were published earlier[24].

To remove microarray platform/source systematic biases, we applied Distance Weighted Discrimination/DWD[25] to the training and test datasets. From the DWD standardized Luminal tumor training dataset, centroids were created consisting of the average expression of the 822 estrogen-SAM genes for Groups IE and IIE. We then classified each ER+ and/or PR+ tumor in the test datasets as Group IE or IIE according to each sample's nearest centroid as determined by Spearman correlation.

Survival Analyses. Kaplan-Meier survival plots were compared using the Cox-Mantel log-rank test in Winstat for Excel (R. Fitch Software, Staufen, Germany). Two-way contingency table analysis was done using Winstat for Excel. For the Chang *et al.*

dataset, multivariate Cox proportional hazards analysis was performed using SAS (Cary, NC).

RESULTS

Identification of estrogen-regulated genes. To identify estrogen-regulated genes, we used the ER+ human breast tumor-derived MCF-7 cell line as a model system. A "one-class" SAM supervised analysis[16] with an FDR of 0.04% identified 383 induced and 574 repressed genes in microarray experiments on MCF-7 cells treated with 17β-estradiol for 2, 4, 8, or 24 hours (hierarchical clustering of these genes is shown in Figure 1). Many genes identified were previously known to be ER-regulated including *CCND1*, *PR*, *RERG*, *CTSD*, and *PDZK1*[26-29]. Using the program EASE[30], the Gene Ontology (GO) categories "sterol metabolism/biosynthesis", "ribosome biogenesis/assembly", and "cytoskeleton structural constituent" were over-represented relative to chance in the set of 383 estrogen-induced genes.

Estrogen and GATA3-regulated genes are present in the Luminal/ER+ gene cluster. The Luminal/ER+ expression cluster is a gene set identified in many breast tumor profiling studies[24,31-33], includes *GATA3* and *ER*, and is expressed in the Luminal A and B tumor subtypes[8,9]. To determine whether estrogen-regulated genes are present in the Luminal/ER+ gene set, we first clustered 118 primary breast tumors using a 1300-gene "breast intrinsic" gene set developed by Hu *et al.*[20] (Figure 2). Figure 2B shows that many of the estrogen-regulated genes from our *in vitro* MCF-7 experiments were present in the tumor defined Luminal/ER+ gene cluster. To further define relationships among genes in this cluster, we also ascertained the presence of genes regulated by *GATA3*, a

transcription factor with an important role in ER+ breast cancer biology[19]. Of the 407 genes induced by GATA3 *in vitro*, many were present in the Luminal/ER+ gene cluster. Thus, genes identified as being regulated by ER and/or GATA3 *in vitro* also cluster near these transcription factors *in vivo* and help to define an expression pattern seen in many studies[24,31,33,34].

Analysis of Luminal tumors using estrogen-induced genes. We hypothesized that expression differences of estrogen-induced genes may define clinically relevant subgroups within clinically defined ER+ and/or PR+ tumors. To test this hypothesis, we clustered the 65 tumors identified as "Luminal" in Figure 2 (blue dendrogram branch) using the 383 MCF-7 estrogen-induced genes. Two main groups resulted (Groups I and II). Group I had higher expression of *XBP1*, *PR*, and *TFF*, which are all known ER targets. Group II had higher expression of a cluster of estrogen-induced genes that included *CTPS*, *E2F6*, and *FANCA*. Kaplan-Meier survival analysis showed that Group I patients had significantly better Relapse-Free Survival (RFS) outcomes than Group II ($p=0.0004$).

To further characterize the differences between Group I and II tumors, we performed a supervised analysis (2-class SAM with 1% FDR) using the major dendrogram branch division that separated Group I and II tumors to define the two supervising groups. This analysis identified 822 genes for which Group I and II tumors showed significant differential expression. This gene set, called the "estrogen-SAM" list, was then used to hierarchically cluster the 65 Luminal tumors (Figure 3), which as expected, resulted in a very similar grouping of samples when compared to that using the

383 estrogen-induced genes. Kaplan-Meier analysis showed that using the estrogen-SAM list grouped the tumors into two groups (referred to as Group IE and IIE) that predicted RFS (p=0.019, Figure 4A).

Group IE tumors showed high expression of *XBP1*, *FOXA1*, *PR* and many ribosomal genes (Figure 3B-E). According to EASE, the GO categories "transcriptional regulation", "DNA binding", and "extracellular" were over-represented relative to chance in Group IE tumors. Group IIE tumors showed the high expression of a prominent proliferation signature[35,36] including *Ki-67*, *MYBL2*, *Survivin*, *STK6*, and *CCNB2* (Figure 3G); these first four genes plus *CCNB1* form the basis for the proliferation portion of the Paik *et al.* "Recurrence Score" predictor[13], which is a gene expression-based outcome predictor for ER+/node-negative, tamoxifen-treated patients. Recently Dai *et al.* performed a supervised analysis for genes that correlated with outcomes in patients with high ER expression relative to age and identified this same proliferation signature as the main determinant for predicting patient outcomes[10]; however, they identified few genes associated with good outcomes.

Group IIE tumors also showed high expression of a cluster of MAGE-A genes (Figure 3F), which have been associated with an increased recurrence risk[37] and poor tumor differentiation[38]. Figure 3H shows Group IIE tumors have high expression of genes with functions in the Interferon-pathway and apoptosis such as FLIP/CFLAR, which is an inhibitor of TNFR-mediated apoptosis[39]. Several anti-apoptosis genes including *FLIP*, *AVEN*, *Survivin* and *BCL2A1* showed high expression in Group IIE, suggesting an impaired ability to undergo cell death. Recent reports have shown that high expression of FLIP[40] or BCL2A1[41,42] can directly contribute to chemoresistance,

suggesting that functional inhibition of these proteins may provide a therapeutic target for Group IIE patients. According to EASE, the GO categories "cell cycle/mitosis", "anti-apoptosis", and "MHC-I" were over-represented relative to chance in Group IIE.

Group IE-IIE classification predicts outcome in ER+ and/or PR+ tumors. To test the Group IE-IIE classification as a clinically relevant outcome predictor, we analyzed ER+ and/or PR+ tumors from 3 published breast tumor microarray datasets[9,12,23]. We used a Single Sample Prediction algorithm to classify tumors in each test dataset, which involved creating Group IE and IIE centroids/average profiles from the training dataset (see Methods). Kaplan-Meier analysis (Figure 4B-D) showed that Group IE tumors had significantly better RFS in all test datasets. Figure 5 shows that the Group IE-IIE classification was also a significant predictor of Overall Survival (OS) for the test datasets in which OS data was available[9,23]. Furthermore, by decreasing the FDR in SAM, we were able to define Groups IE and IIE using a reduced estrogen-SAM list of 113 genes without any loss of predictive ability (Table 1).

Multivariate analysis. Multivariate Cox proportional hazards analysis was performed on the Chang *et al.* dataset (Table 2). Using RFS and OS as the endpoints, multivariate analysis showed that classifying tumors as Group IE or IIE provided significant prognostic power independent of standard clinical factors ($p<0.0001$ using RFS, $p=0.001$ using OS). The Group IE-IIE designation had the strongest association of all variables with RFS and OS.

In multivariate analyses that included Chang *et al.*'s[23] "wound-response" signature and van't Veer *et al.*'s[24] "70-gene signature" along with the clinical variables, the Group IE-IIE classification continued to provide significant prognostic power independent of other variables in the model (p=0.014 using RFS, p=0.042 using OS, Table 3). The performance of the 70-gene and wound-response signatures in this multivariate analysis may be optimistically high because a subset of the patients in the Chang *et al.* dataset was used to train/optimize these two signatures; therefore, the ability of the Group IE-IIE classification to show independent prognostic power in a model containing these two predictors indicates its usefulness in predicting outcomes.

Group IE-IIE associations with clinical and biological parameters. To examine the hypothesis that Group IE may be more differentiated than Group IIE tumors, we determined whether an association existed between this classification and histological grade. Two-way contingency table analysis showed significant association between grade and Group IE-IIE class (Table 4), with grade 1 and 3 tumors more likely to be classified as Group IE and IIE, respectively. Cramer's V statistic, which measures the strength of association between two variables in a contingency table, indicated a substantial association (Cramer's V>0.36) between grade and Group IE-IIE class for all datasets. For the Sorlie *et al.* dataset, p53 mutation data was available and a two-way contingency table analysis showed a significant association between p53 status and Group IE-IIE class, with Group IIE more likely to be p53 mutant (p=0.0019; Cramer's V=0.44).

Comparison of the Group IE-IIE classification to the Luminal A-B classification. We compared the Group IE-IIE classification to the Luminal A-B classification[8,9]. To identify Luminal A and B tumors in the three test datasets, we used the Single Sample Predictor developed in Hu *et al.*[20], which employs centroids for each of the five breast tumor "intrinsic subtypes". We then reclassified Luminal A and B tumors from each dataset as Group IE or IIE. Kaplan-Meier analyses showed that the Group IE-IIE classification did equally well or slightly better than the Luminal A-B classification in separating Luminal tumors into two groups with different survival outcomes (Table 5).

DISCUSSION

The search for markers that predict long-term outcomes in hormone receptor-positive tamoxifen-treated patients has been an intense area of study. Genomic analyses have contributed to this area, with the development of several predictive gene sets and assays based upon the selection of genes that directly correlate with patient/tumor outcomes[10-13,24]. We took a different approach and selected genes using no knowledge of outcomes and instead, selected genes based upon regulation by estrogen and their natural patterns of expression in primary tumors. The 822-gene estrogen-SAM list identified many genes that may help explain the outcome differences seen in ER+ and/or PR+ patients. Good outcome group IE tumors tended to be more differentiated and highly expressed a subset of estrogen and *GATA3*-regulated genes. Conversely, poor outcome Group IIE tumors were more likely to be poorly differentiated. Association of the Group IE-IIE profile with grade is expected because grade includes a measure of proliferation, which is an important determinant of outcomes in ER+ and/or PR+ patients[8,10,13,19].

However, because the Group IE-IIE distinction was significant in a multivariate analysis with grade included, this distinction adds prognostic information beyond what grade provides.

We used three published datasets as test sets and confirmed that the Group IE-IIE classification was a significant predictor in ER+ and/or PR+ patients. We note, however, that the relapse rates differed between datasets and that the Group IE tumors showed 7-40% relapse rates depending on the dataset (Figure 4). This underscores the fact that relapse rates are dependent upon the characteristics of the patient set used. For example, comparing relapse rates in the Chang *et al.* dataset to those observed in Paik *et al.* may not be valid because the Paik *et al.* dataset was comprised of tamoxifen-treated node-negative patients, while the majority of the Chang *et al.* patients received no adjuvant therapy and many were node-positive. However, the multivariate analysis we performed on the Chang *et al.* dataset indicated that our predictor had significant prognostic value independent of standard clinical factors and other gene expression-based predictors, and a hazard ratio of 2.90 for Group IIE vs. IE indicates our predictor has potential clinical utility. By limiting the Chang *et al.* dataset to those patients who received adjuvant therapy and were Stage I+II, we observed a relapse rate for Group IE patients of 12% (Figure 5D, p=0.007) and significance for overall survival outcomes (data not shown). This indicates that given a patient population similar to Paik *et al.*, our predictor's "good group" can achieve outcomes similar to the Paik *et al.* "Low Risk" group.

An important unanswered question is whether the Group IE-IIE distinction predicts pure prognosis, responsiveness to endocrine therapy, or both. From analyses of patient subsets in the Chang *et al.* dataset, it is clear that the Group IE-IIE distinction

predicts outcome in ER+ and/or PR+ patient subsets either receiving or not receiving adjuvant hormone therapy (data not shown). Paik *et al.* observed similar results for their predictor[13]. This is not surprising because half (8/16) of the Paik *et al.* genes were present in the "estrogen-SAM" gene set. However, an advantage of our analysis is that it provides additional biological information (*e.g.* anti-apoptosis genes) that the Paik *et al.* and other predictors did not. Paik *et al.*'s finding that their predictor also predicts benefit of chemotherapy[43] may also apply to ours. Thus, the most pressing questions remaining regarding the Group IE-IIE classification are (1) whether Group IE and IIE gain similar benefits from chemotherapy, and (2) because Group IIE tumors do poorly in the presence of tamoxifen, might they do better if given alternative endocrine therapies.

Acknowledgements

This work was supported by funds from the NCI Breast SPORE program to UNC-CH (P50-CA58223-09A1), National Institutes of Health (M01RR00046), and by the National Institute of Environmental Health Sciences (U19-ES11391-03). M.A.T. was supported by the UNC Lineberger Cancer Control Education Program (R25 CA57726). We thank Drs. Phil Bernard, Juan Palazzo and Funmi Olopade for providing tumor materials and advice, and Lynda Sawyer and Amy Drobish for collection of patient associated data.

REFERENCES

1. Osborne CK, Yochmowitz MG, Knight WAd, et al: The value of estrogen and progesterone receptors in the treatment of breast cancer. Cancer 46:2884-8, 1980

2. Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. Lancet 365:1687-717, 2005

3. Bardou VJ, Arpino G, Elledge RM, et al: 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 21:1973-9, 2003

4. Osborne CK: Tamoxifen in the treatment of breast cancer. N Engl J Med 339:1609-18, 1998

5. Ellis MJ, Coop A, Singh B, et al: 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 19:3808-16, 2001

6. Osborne CK, Bardou V, Hopp TA, et al: Role of the estrogen receptor coactivator AIB1 (SRC-3) and HER-2/neu in tamoxifen resistance in breast cancer. J Natl Cancer Inst 95:353-61, 2003

7. Shou J, Massarweh S, Osborne CK, et al: Mechanisms of tamoxifen resistance: increased estrogen receptor-HER2/neu cross-talk in ER/HER2-positive breast cancer. J Natl Cancer Inst 96:926-35, 2004

8. Sorlie T, Perou CM, Tibshirani R, et al: Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 98:10869-74., 2001

9. Sorlie T, Tibshirani R, Parker J, et al: Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A 100:8418-23, 2003

10. Dai H, van 't Veer LJ, Lamb J, et al: A Cell Proliferation Signature is a Marker of Extremely Poor Outcome in a Subpopulation of Breast Cancer Patients. Proc Natl Acad Sci U S A 65:4059-4066, 2005

11. Jansen MP, Foekens JA, van Staveren IL, et al: Molecular classification of tamoxifen-resistant breast carcinomas by gene expression profiling. J Clin Oncol 23:732-40, 2005

12. Ma XJ, Wang Z, Ryan PD, et al: A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen. Cancer Cell 5:607-16, 2004

13. Paik S, Shak S, Tang G, et al: A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 351:2817-26, 2004

14. Troester MA, Hoadley KA, Sorlie T, et al: Cell-type-specific responses to chemotherapeutics in breast cancer. Cancer Res 64:4218-26, 2004

15. Hu Z, Troester M, Perou CM: High reproducibility using sodium hydroxide-stripped long oligonucleotide DNA microarrays. Biotechniques 38:121-4, 2005

16. Tusher VG, Tibshirani R, Chu G: Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A 98:5116-21, 2001

17. Larsson O, Wahlestedt C, Timmons JA: Considerations when using the significance analysis of microarrays (SAM) algorithm. BMC Bioinformatics 6:129, 2005

18. Eisen MB, Spellman PT, Brown PO, et al: Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 95:14863-8, 1998

19. Usary J, Llaca V, Karaca G, et al: Mutation of GATA3 in human breast tumors. Oncogene 23:7669-78, 2004

20. Hu Z, Fan C, Oh DS, et al: A New Breast Tumor Intrinsic Gene List Identifies Novel Characteristics that are Conserved Across Microarray Platforms. Submitted, 2006

21. Bair E, Tibshirani R: Semi-supervised methods to predict patient survival from gene expression data. PLoS Biol 2:E108, 2004

22. Bullinger L, Dohner K, Bair E, et al: Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. N Engl J Med 350:1605-16, 2004

23. Chang HY, Nuyten DS, Sneddon JB, et al: Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. Proc Natl Acad Sci U S A 102:3738-43, 2005

24. van 't Veer LJ, Dai H, van de Vijver MJ, et al: Gene expression profiling predicts clinical outcome of breast cancer. Nature 415:530-6, 2002

25. Benito M, Parker J, Du Q, et al: Adjustment of systematic microarray data biases. Bioinformatics 20:105-114, 2004

26. Altucci L, Addeo R, Cicatiello L, et al: 17beta-Estradiol induces cyclin D1 gene transcription, p36D1-p34cdk4 complex activation and p105Rb phosphorylation during mitogenic stimulation of G(1)-arrested human breast cancer cells. Oncogene 12:2315-24, 1996

27. Finlin BS, Gau CL, Murphy GA, et al: Rerg is a novel ras-related, estrogen-regulated and growth-inhibitory gene in breast cancer. J Biol Chem 31:31, 2001

28. Ghosh MG, Thompson DA, Weigel RJ: PDZK1 and GREB1 are estrogen-regulated genes expressed in hormone-responsive breast cancer. Cancer Res 60:6367-75, 2000

29. Jakowlew SB, Breathnach R, Jeltsch JM, et al: Sequence of the pS2 mRNA induced by estrogen in the human breast cancer cell line MCF-7. Nucleic Acids Res 12:2861-78, 1984

30. Hosack DA, Dennis G, Jr., Sherman BT, et al: Identifying biological themes within lists of genes with EASE. Genome Biol 4:R70, 2003

31. Gruvberger S, Ringner M, Chen Y, et al: Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns. Cancer Res 61:5979-84., 2001

32. Hoch RV, Thompson DA, Baker RJ, et al: GATA-3 is expressed in association with estrogen receptor in breast cancer. Int J Cancer 84:122-8, 1999

33. Sotiriou C, Neo SY, McShane LM, et al: Breast cancer classification and prognosis based on gene expression profiles from a population-based study. Proc Natl Acad Sci U S A 100:10393-8, 2003

34. West M, Blanchette C, Dressman H, et al: Predicting the clinical status of human breast cancer by using gene expression profiles. Proc Natl Acad Sci U S A 98:11462-7., 2001

35. Perou CM, Jeffrey SS, van de Rijn M, et al: Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. Proc Natl Acad Sci U S A 96:9212-9217, 1999

36. Whitfield ML, Sherlock G, Saldanha AJ, et al: Identification of genes periodically expressed in the human cell cycle and their expression in tumors. Mol Biol Cell 13:1977-2000, 2002

37. Otte M, Zafrakas M, Riethdorf L, et al: MAGE-A gene expression pattern in primary breast cancer. Cancer Res 61:6682-7, 2001

38. Kavalar R, Sarcevic B, Spagnoli GC, et al: Expression of MAGE tumour-associated antigens is inversely correlated with tumour differentiation in invasive ductal breast cancers: an immunohistochemical study. Virchows Arch 439:127-31, 2001

39. Srinivasula SM, Ahmad M, Ottilie S, et al: FLAME-1, a novel FADD-like anti-apoptotic molecule that regulates Fas/TNFR1-induced apoptosis. J Biol Chem 272:18542-5, 1997

40. Abedini MR, Qiu Q, Yan X, et al: Possible role of FLICE-like inhibitory protein (FLIP) in chemoresistant ovarian cancer cells in vitro. Oncogene 23:6997-7004, 2004

41. Wang CY, Guttridge DC, Mayo MW, et al: NF-kappaB induces expression of the Bcl-2 homologue A1/Bfl-1 to preferentially suppress chemotherapy-induced apoptosis. Mol Cell Biol 19:5923-9, 1999

42. Cheng Q, Lee HH, Li Y, et al: Upregulation of Bcl-x and Bfl-1 as a potential mechanism of chemoresistance, which can be overcome by NF-kappaB inhibition. Oncogene 19:4936-40, 2000

43. Paik S, Shak S, Tang G, et al: Expression of the 21 genes in the Recurrance Score assay and prediction benefits from tamoxifen in NSABP study B-14 and chemotherapy in NSABP study B-20, San Antionio Breast Cancer Symposium. San Antionio, TX, on line abstract, 2004

TABLES

Table 2.1. Summary of Kaplan-Meier Relapse-Free Survival analyses performed for each tumor dataset using the estrogen-induced or estrogen-SAM gene lists.

	Training	**Testing**		
Gene list	**Luminals** (65 tumors, 10 events)	**Ma et al. ER+ and/or PR+** (60 tumors, 28 events)	**Sorlie et al. ER+ and/or PR+** (90 tumors, 45 events)	**Chang et al. ER+ and/or PR+** (250 tumors, 86 events)
383 estrogen-induced gene list	p=**0.0004**	**0.044**	**0.0008**	**8.1e-5**
822 gene estrogen-SAM list	**0.019**	**0.0006**	**0.0007**	**1.3e-5**
113 gene reduced estrogen-SAM list	**0.007**	**0.008**	**0.001**	**6.6e-6**

Each cell in the table contains the p-value calculated using the log-rank test for the Kaplan-Meier relapse-free survival curves produced for the two tumor groups identified by each gene list in each tumor dataset. p-values <0.05 are in bold.

Table 2.2. Multivariate Cox proportional hazards analysis of various prognostic factors in relation to Relapse-Free Survival and Overall Survival for ER+ and/or PR+ tumors in the Chang et al. (2005) dataset.

	Relapse-Free survival		**Overall survival**	
Variable	**Hazard Ratio (95% CI)**	**p-value**	**Hazard Ratio (95% CI)**	**p-value**
Group IIE vs. IE	**2.90** (1.71-4.92)	**<0.0001**	**3.64** (1.67-7.95)	**0.001**
Age, per decade	**0.48** (0.31-0.74)	**0.001**	**0.53** (0.30-0.93)	**0.028**
Size	**1.59** (1.01-2.48)	**0.044**	1.45 (0.80-2.64)	0.22
Tumor grade 2,3 vs. 1	1.80 (0.99-3.3)	0.056	**3.57** (1.24-10.23)	**0.02**
Node status	**2.11** (1.08-4.11)	**0.028**	1.85 (0.74-4.61)	0.19
Hormonal or chemotherapy vs. no adjuvant therapy	**0.36** (0.18-0.71)	**0.003**	0.47 (0.19-1.19)	0.11

Size was a binary variable (0= diameter of 2cm or less, 1= greater than 2cm); node status was a binary variable (0= no positive nodes, 1= one or more positive nodes); age was a continuous variable formatted as decade-years. Tumors were classified as Group IE or IIE using the estrogen-SAM derived list. Variables found to be significant ($p<0.05$) in the Cox proportional hazards model are shown in bold.

Table 2.3. Multivariate Cox proportional hazards analysis for ER+ and/or PR+ tumors in the Chang et al. (2005) dataset using various prognostic factors including the Group IE-IIE classification, the van't Veer et al. (2002) 70-gene signature, and the Chang et al. (2005) Wound-Response signature.

	Relapse-Free survival		**Overall survival**	
Variable	**Hazard Ratio (95% CI)**	**p-value**	**Hazard Ratio (95% CI)**	**p-value**
Group IIE vs. IE	**2.01** (1.15-3.49)	**0.014**	**2.31** (1.03-5.19)	**0.042**
70-gene signature (poor vs. good)	**2.76** (1.50-5.06)	**0.001**	**4.17** (1.62-10.73)	**0.003**
Wound-response signature (activated vs. quiescent)	**2.30** (1.09-4.85)	**0.028**	2.80 (0.82-9.55)	0.10
Age, per decade	**0.56** (0.36-0.87)	**0.010**	0.64 (0.36-1.14)	0.13
Size	1.45 (0.93-2.28)	0.10	1.34 (0.74-2.45)	0.34
Tumor grade 2,3 vs. 1	0.93 (0.47-1.82)	0.82	1.62 (0.53-4.93)	0.40
Node status	1.72 (0.89-3.33)	0.11	1.51 (0.62-3.70)	0.36
Hormonal or chemotherapy vs. no adjuvant therapy	**0.37** (0.19-0.74)	**0.005**	0.46 (0.18-1.14)	0.095

Tumor size, node status, age, and Group IE-IIE were defined as in Table 1. The 70-gene signature and the wound-response signature classifications were taken exactly as calculated in Chang et al. (2005), and their performances in multivariate analysis may be optimistic. Variables found to be significant ($p<0.05$) in the Cox proportional hazards model are shown in bold.

Table 2.4. Association between tumor grade and the Group IE-IIE classification (estrogen-SAM derived list).

	65 luminal tumors from training set		**Ma et al. ER+/PR+ tumors**		**Sorlie et al. ER+/PR+ tumors**		**Chang et al. ER+/PR+ tumors**	
Two-way contingency table	**Group**		**Group**		**Group**		**Group**	
	IE	**IIE**	**IE**	**IIE**	**IE**	**IIE**	**IE**	**IIE**
Histologic grade	(# of patients)		(# of patients)		(# of patients)		(# of patients)	
1 (well)	11	0	3	0	10	0	50	24
2 (intermediate)	16	12	29	10	32	11	46	49
3 (poor)	9	14	6	12	16	20	13	64
Statistics for two-way contingency table analysis								
p-value†	0.003		0.005		0.001		1.7e-9	
Cramer's V††	0.43		0.42		0.39		0.40	

† p-value calculated from Chi-square test on contingency table. †† Cramer's V statistic (value can range from 0 to 1) measures the strength of association between the two variables analyzed in the contingency table, with 1 indicating perfect association and 0 indicating no association.

Table 2.5. Comparison of the Group IE-IIE classification to the Luminal A-B classification in predicting RFS for three datasets of Luminal epithelial tumors.

Classification method	Survival analysis statistic	**Ma et al.** (43 tumors, 20 events)	**Sorlie et al. †** (57 tumors, 20 events)	**Chang et al.** (194 tumors, 62 events)
Luminal A-B classification	p-value	p=**0.011**	0.14	**1.1e-9**
	Hazard ratio of B vs. A (95% CI)	HR=3.1 (0.9- 10.6)	1.9 (0.7- 5.0)	4.2 (2.4-7.5)
Group I-II classification	p-value	**0.0002**	0.076	**1.3e-6**
	Hazard ratio of II vs. I (95% CI)	4.9 (1.8- 13.0)	2.2 (0.8-5.8)	4.0 (2.4-6.6)

†When the Kaplan-Meier disease-specific survival curves were compared, the Group IE-IIE classification scheme showed significantly different survival curves (p=**0.031**) whereas the Luminal A-B classification scheme did not (p=0.17).

FIGURES

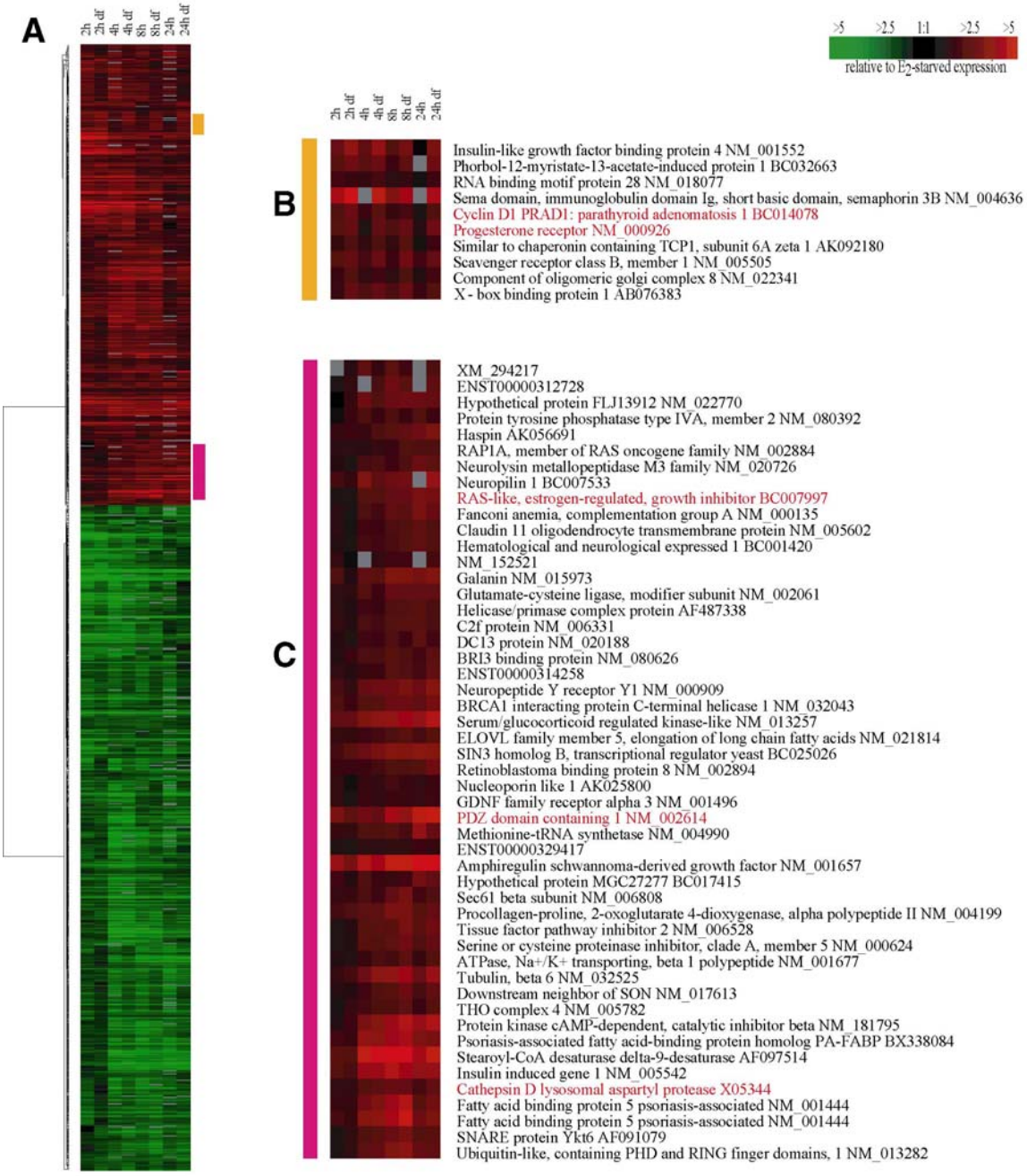


Figure 2.1. Hierarchical clustering analysis of the genes determined by 1-class SAM to be estrogen-induced (383 genes) or estrogen-repressed (574 genes). **A.** Scaled-down representation of the complete cluster diagram. **B, C.** Subsets of estrogen-induced genes with known estrogen-induced genes highlighted in red. For each gene, mRNA levels at the indicated time point are relative to the estrogen (E_2)-starved control sample.

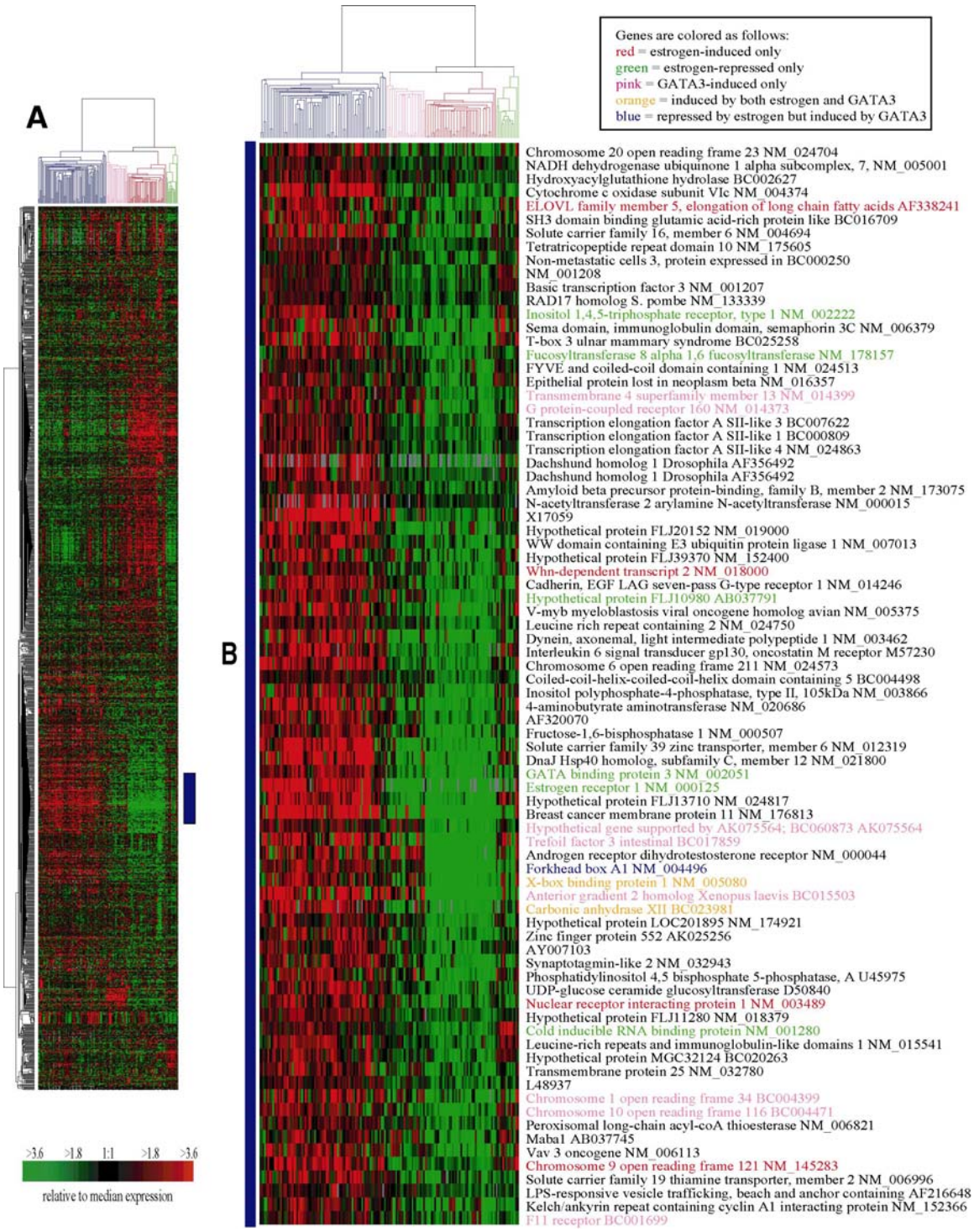


Figure 2.2. Genes regulated by estrogen and/or GATA3 *in vitro* are present in the primary tumor Luminal epithelial/ER+ gene cluster. **A.** Scaled-down representation of 118 tumors hierarchically clustered using the 1300-gene intrinsic list developed by Hu *et al*.[20] **B.** Luminal/ER+ gene cluster. The tumor sample dendrogram is colored as: blue=Luminal epithelial subtype, pink=HER2+/ER-, red=Basal-like, and green=Normal Breast-like.

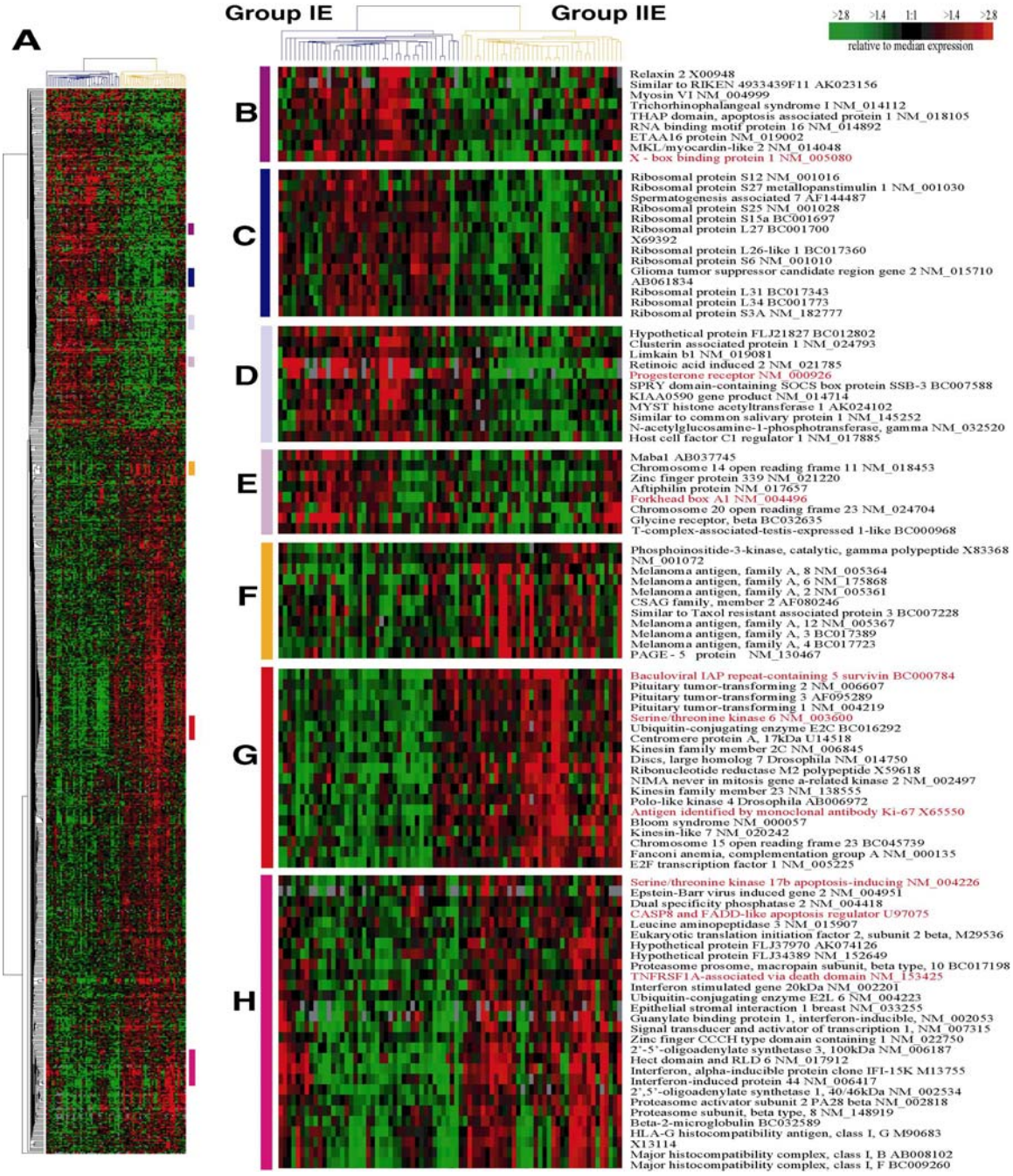


Figure 2.3. Hierarchical cluster analysis of the 65 Luminal tumors (identified in Figure 2.1) using the 822-gene estrogen-SAM derived list. **A.** Scaled-down representation of the complete cluster diagram. Group IE and IIE tumors are indicated by blue and orange, respectively. Gene clusters containing **B.** XBP1, **C.** Ribosomal genes, **D.** Progesterone receptor, **E.** FOXA1, **F.** MAGE genes, **G.** Proliferation signature, and **H.** Apoptosis and interferon-response genes.

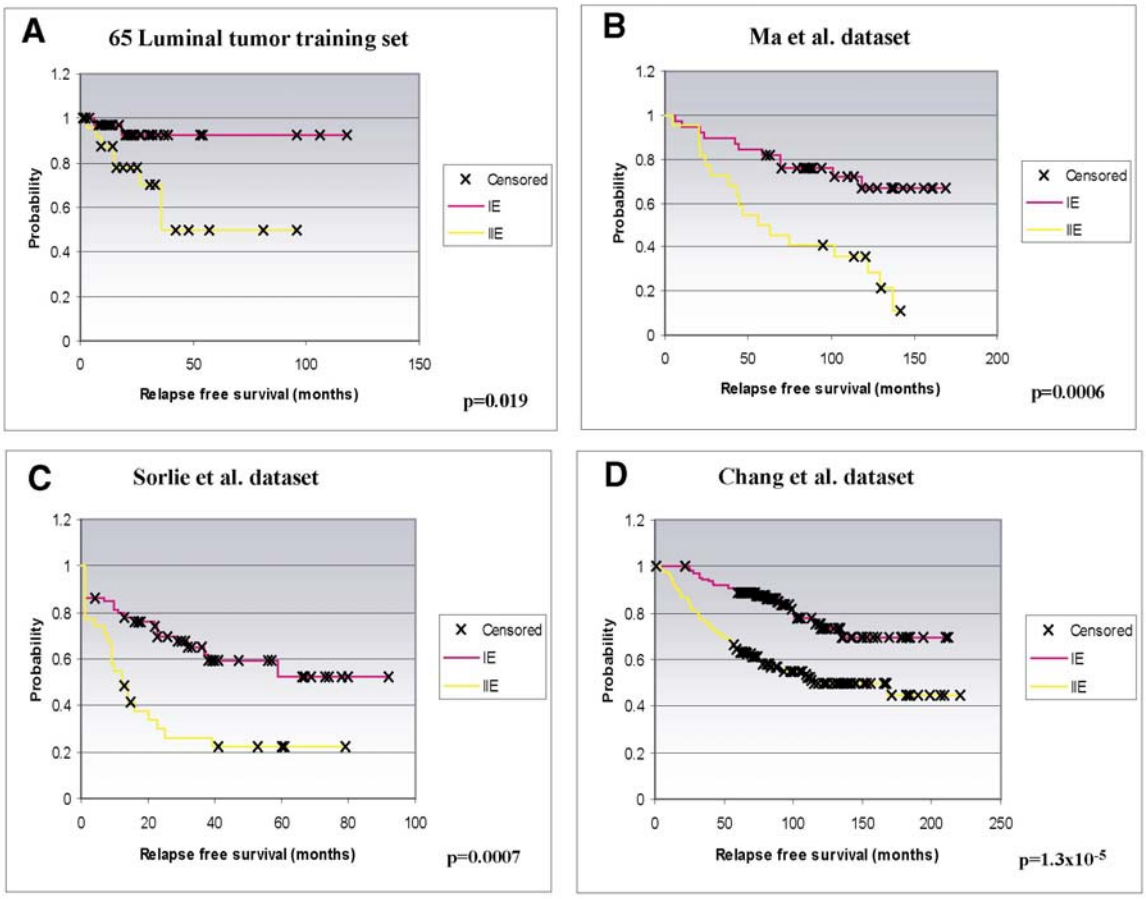


Figure 2.4. Kaplan-Meier survival curves of ER+ and/or PR+ tumors classified as Groups IE or IIE using the 822-gene estrogen-SAM derived list. Survival curves are shown for **A.** the 65 Luminal epithelial tumor training dataset, **B.** the Ma *et al*., **C.** Sorlie *et al*., and **D.** Chang *et al.* datasets. p-values calculated using the log-rank test.

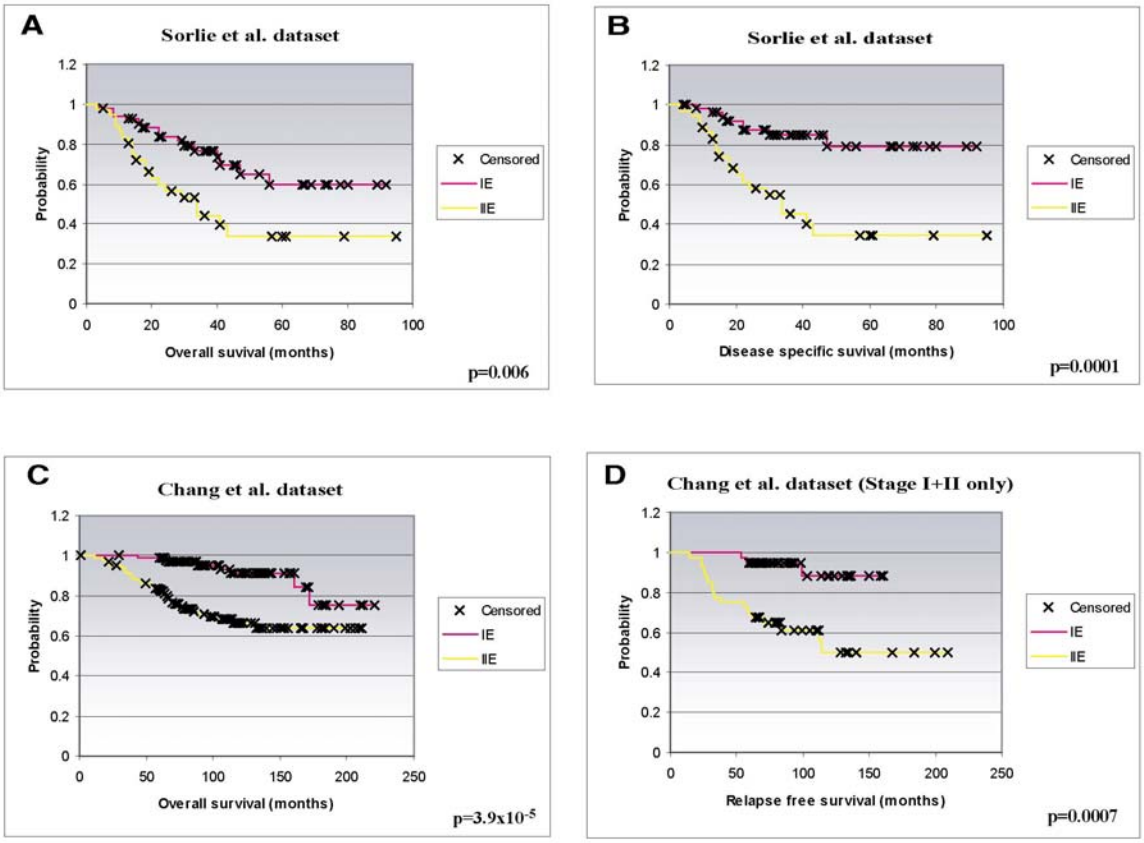


Figure 2.5. Kaplan-Meier Overall Survival curves of tumors classified into Groups IE and IIE using the 113-gene estrogen-SAM derived list. Overall survival curves are shown for **A.** the Sorlie *et al.* and **C.** Chang *et al.* data sets. Disease specific survival curves are shown for **B.** the Sorlie *et al.* dataset. **D.** Relapse-free survival curves for Chang *et al.* Stage I + II patients only. p-values were calculated using the log-rank test.

CHAPTER 3: The molecular portraits of breast tumors are conserved across microarray platforms

Z. Hu, C. Fan, Daniel S. Oh, J.S. Marron, X. He, B.F. Qaqish, G. Karaca, C. Livasy, L.A. Carey, E. Reynolds, L. Dressler, A. Nobel, J. Parker, M.G. Ewend, L.R. Sawyer, D. Xiang, J. Wu, Y. Liu, M. Karaca, R. Nanda, M. Tretiakova, A.R. Orrico, D. Dreher, J.P. Palazzo, L. Perreard, E. Nelson, M. Mone, H. Hansen, M. Mullins, J.F. Quackenbush, M.J. Ellis, O.I. Olopade, Philip S. Bernard and C.M. Perou.

This work was published in BMC Genomics. 2006 Apr 27; 7:96.

ABSTRACT

Background

Validation of a novel gene expression signature in independent data sets is a critical step in the development of a clinically useful test for cancer patient risk-stratification. However, validation is often unconvincing because the size of the test set is typically small. To overcome this problem we used publicly available breast cancer gene expression data sets and a novel approach to data fusion, in order to validate a new breast tumor intrinsic list.

Results

A 105-tumor training set containing 26 sample pairs was used to derive a new breast tumor intrinsic gene list. This intrinsic list contained 1300 genes and a proliferation signature that was not present in previous breast intrinsic gene sets. We tested this list as

a survival predictor on a data set of 311 tumors compiled from three independent microarray studies that were fused into a single data set using Distance Weighted Discrimination. When the new intrinsic gene set was used to hierarchically cluster this combined test set, tumors were grouped into LumA, LumB, Basal-like, HER2+/ER-, and Normal Breast-like tumor subtypes that we demonstrated in previous datasets. These subtypes were associated with significant differences in Relapse-Free and Overall Survival. Multivariate Cox analysis of the combined test set showed that the intrinsic subtype classifications added significant prognostic information that was independent of standard clinical predictors. From the combined test set, we developed an objective and unchanging classifier based upon five intrinsic subtype mean expression profiles (i.e. centroids), which is designed for single sample predictions (SSP). The SSP approach was applied to two additional independent data sets and consistently predicted survival in both systemically treated and untreated patient groups.

Conclusions

This study validates the "breast tumor intrinsic" subtype classification as an objective means of tumor classification that should be translated into a clinical assay for further retrospective and prospective validation. In addition, our method of combining existing data sets can be used to robustly validate the potential clinical value of any new gene expression profile.

INTRODUCTION

The classification of human tumors using microarray data has been an area of intense research, but it remains a daunting task to validate a new profile and generate a clinically useful test. Many different gene expression-based predictors have been developed for breast cancer[1-9], and two different gene expression predictors have reached the final step of prospective clinical trial testing[10,11]. Using cDNA microarrays, we previously identified five distinct subtypes of breast tumors arising from at least two distinct cell types (basal-like and luminal epithelial cells)[1-3]. This molecular taxonomy was based upon an "intrinsic" gene set, which was identified using a supervised analysis to select genes that showed little variance within repeated samplings of the same tumor, but which showed high variance across tumors[1]. We showed that an intrinsic gene set reflects the stable biological properties of tumors and typically identifies distinct tumor subtypes that have prognostic significance, even though no knowledge of outcome was used to derive this gene set[3,12-14].

A major challenge for microarray studies, especially those with clinical implications, is validation[15,16]. Due to the practical barriers of cost and access to large numbers of fresh frozen tumor samples with associated clinical information, very few microarray studies have analyzed enough samples to allow promising initial findings to be sufficiently validated to justify the major investment required for clinical testing. An efficient approach would be to use public gene expression data repositories as test sets; however, it has been difficult to compare and/or combine data sets from independent laboratories due to differences in sample preparation, experimental design, and microarray platforms. An accepted method for validation is to derive a

prognostic/predictive gene set from a "training set" and then apply it to a completely independent "test set"[17]. The "purest" test sets are comprised of samples not generated by the primary investigators to remove any possibility of bias[18]. In this study, we illustrate the successful application of these principles by (1) deriving a new breast tumor intrinsic gene list that identifies the "intrinsic" biological features of breast tumors and (2) validating this predictor using a combined test set of 311 breast tumor samples compiled from the public domain. These analyses show that the breast tumor intrinsic subtypes are significant predictors of outcome when correcting for standard clinical parameters, and that common patterns of expression and outcome predictions can be identified in data sets generated by independent labs.

MATERIALS AND METHODS

Sample collection, RNA isolation and microarray hybridization. 105 fresh frozen breast tumor samples and 9 normal breast samples were obtained using IRB-approved protocols at 4 institutions: the University of North Carolina at Chapel Hill (UNC-CH), The University of Utah, Thomas Jefferson University, and the University of Chicago. This sample set represents an ethnically and geographically diverse cohort. Patients were heterogeneously treated according to the standard of care dictated by disease stage, ER and HER2 status.

Total RNA was purified from each sample using the Qiagen RNAeasy Kit. RNA integrity was determined using the RNA 6000 Nano LabChip Kit and Agilent 2100 Bioanalyzer. Total RNA amplification and labeling were done as previously described[35]. Microarray hybridizations were performed using Agilent Human oligonucleotide (1Av1,

1Av2 and custom designed 1Av1-based) microarrays using 2μg of Cy3-labeled common reference sample that is a modified version of the Stratagene Human Universal Reference[36], and 2μg of Cy5-labeled experimental sample. Microarrays were hybridized overnight, washed, dried, and scanned as described[35]. The image files were analyzed with GenePix Pro 4.1 and loaded into the UNC-CH Microarray Database (https://genome.unc.edu/) where a Lowess normalization procedure was performed to adjust the Cy3 and Cy5 channels[37]. All primary microarray data associated with this study are available at https://genome.unc.edu/pubsup/breastTumor/ and in the GEO (http://www.ncbi.nlm.nih.gov/geo/) under the accession number of GSE1992, series GSM34424-GSM34568.

Identification of the Intrinsic gene set. We derived a new breast tumor intrinsic gene set, referred to as the "Intrinsic/UNC" list, using a training set composed of the 105 tumor samples described above, 9 normal breast samples, and 26 sample pairs (in total, represented by 146 microarrays). 15, 9, and 2 of the 26 sample pairs were different physical pieces of the same tumor (taken at the same time point), tumor-metastasis pairs and normal sample pairs, respectively. The background subtracted, Lowess normalized $\log_2$ ratio of Cy5 to Cy3 intensity values were first filtered to select genes that had a signal intensity of at least 30 units above background in both the Cy5 and Cy3 channels. Only genes that met these criteria in at least 70% of the 146 microarrays were included for subsequent analysis. Next, we performed an "intrinsic" analysis as described previously[3] using the 26 sample pairs and 86 additional microarrays. An intrinsic analysis identifies genes showing low variability in expression within paired samples but high

variability in expression across different tumors; for each gene a ratio of "within-pair variance" to "between-subject variance" is computed. Genes with ratios below one standard deviation of the mean ratio were defined as "intrinsic". This analysis resulted in 1410 microarray elements representing 1300 genes being identified as "intrinsic". In order to obtain an estimate of the number of false-positive intrinsic genes, we permuted the sample labels to generate 26 random pairs and 86 non-paired samples. This permutation was performed 100 times and the intrinsic scores were calculated for each. These permuted scores were used to determine a threshold on the intrinsic score corresponding to a false discovery rate (FDR) less than 1%. The selected threshold resulted in 1410 microarray features being called significant with a median FDR=0.3% and 90th percentile FDR=0.5%. (See Tusher *et al.* for a complete description of this calculation[38]).

Creation and analyses of the combined test set. The independent test set was a 315-sample "combined test set" consisting of three DNA microarray datasets (Sorlie *et al*. 2001 and 2003[2,3], van't Veer *et al*. 2002[5] and Sotiriou *et al.* 2003[19]). To combine these datasets obtained from different microarray platforms, we performed the following pre-processing methods. First, the R/G ratios in each dataset were $\log_2$ transformed and Lowess normalized[37]. Next, missing values were k-NN imputed[39]. Gene annotations from each dataset were converted into UniGene Cluster IDs (UCIDs, Build 161) using the SOURCE database[40], and multiple occurrences of a UCID were collapsed by taking the median value for that ID within each experiment and platform, which resulted in ~2800 genes having expression data in all three datasets. Next, Distance Weighted

Discrimination[20] was performed in a pair-wise fashion by first combining the Sorlie et al. and Sotiriou et al. datasets, and then combining this with the van't Veer et al. dataset to make a single dataset. In the final pre-processing step, each microarray experiment was normalized such that each column/experimental sample was standardized to N(0,1), and each row/gene was median centered. 306 of the 1300 Intrinsic/UNC genes had microarray data present in the combined test set and were used in a two-way average-linkage hierarchical cluster analysis[41]. Cluster results were visualized using the program "Treeview".

Derivation of the Single Sample Predictor. The Single Sample Predictor (SSP) is a Nearest Centroid-based method based upon the work of Hastie and Tibshirani[3,42,43]. Our SSP classifies an individual sample according to its nearest centroid as determined by Spearman correlation. To derive our SSP, we utilized the 315-sample combined test set from Figure 2 to create centroids for each of the five intrinsic subtypes (LumA, LumB, HER2+/ER-, Basal-like and Normal Breast-like). Please note that we did not create a centroid for the IFN group because it failed significance in multivariate testing, but did create a centroid for the Normal Breast-like group because we feel it is important to be able to identify true normal samples; an H&E examination of most tumor samples falling into the Normal Breast-like category shows that this is occurring mainly because of too much normal tissue contamination.

To create each intrinsic subtype centroid, we averaged the gene expression profiles for samples clearly assigned to each subtype (limiting the analysis to 249 of the 315 samples) using the hierarchical clustering dendrogram as a guide (Figure 2). We then

applied the SSP to two independent test datasets: (1) the Ma et al. 60-sample ER+ tamoxifen-treated tumor dataset and (2) the Chang et al. 96-sample local only-treated tumor dataset. By matching UCIDs, microarray data for as many as possible of the 306 Intrinsic/UNC genes was obtained from these 2 datasets. To remove microarray platform/source systematic biases, we applied DWD to the 2 test datasets relative to the combined test set. The SSP was then used to classify tumors by intrinsic subtype in these 2 test datasets. Using similar methods, the SSP was also applied to the 105-sample training set used to derive the intrinsic/UNC gene set.

Survival analyses. Kaplan-Meier survival plots were compared using the Cox-Mantel log-rank test in WinSTAT for Excel (R. Fitch Software). Two-way contingency table analysis and unpaired Student's t-test were done using WinSTAT. For the "combined test set", multivariate Cox proportional hazards analysis was performed using SAS (Cary, NC).

RESULTS

Identification of the Intrinsic/UNC gene set. Our goals were to (1) create a new breast tumor intrinsic list, (2) validate this list on an independent dataset to show the clinical significance of the "intrinsic" classifications, and (3) to derive an objective "intrinsic subtype" classifier that could be used clinically (see Figure 1 for overview of analyses performed). An intrinsic analysis is a "within class" versus "across classes" analysis that identifies genes that show low variability within a group (i.e. a tumor-metastasis pair), but which show high variation in expression across different tumors; in essence, one is

selecting for genes that are consistently expressed when individual tumors are examined, but that vary in expression across different tumors. To develop a new breast tumor intrinsic gene set (Intrinsic/UNC), we assayed a training set of 105 breast tumor samples and 9 normal breast samples, which contained 26 sample pairs (146 microarray experiments in total), using Agilent oligo microarrays. Using the intrinsic analysis method as described in Sorlie et al. 2003[3], we identified an intrinsic gene set of 1410 microarray elements representing 1300 genes. We felt it important to create a new intrinsic list because first, we wanted to take advantage of newer microarrays (Agilent arrays with 17,000 genes vs. 8,000 gene cDNA microarrays previously used[3]), and second, we wanted to use paired tumor samples that were not before-and-after chemotherapy pairs, but were instead pre-treatment tumor pairs. The Intrinsic/UNC gene set showed overlap with a previous breast tumor intrinsic gene set (108 genes in common with the Intrinsic/Stanford gene set of Sorlie et al. 2003[3]), but also showed a significant increase in gene number likely due to the greater number of genes present on current microarrays.

Validation of the Intrinsic/UNC gene list. To evaluate the Intrinsic/UNC gene set on an independent test dataset, we applied it to a "combined test set" of 315 breast samples (311 tumors and 4 normal breast samples) using hierarchical clustering methods as have been done previously[1-3]. The "combined test set" of 315 breast samples was a single data set created by combining together the data from Sorlie *et al*. 2001 and 2003 (cDNA microarrays)[2,3], van't Veer *et al*. 2002 (custom Agilent oligo microarrays)[5] and Sotiriou *et al.* 2003 (cDNA microarrays)[19]. We created a single data table of these three sets by

first identifying the common genes present across all three microarray data sets (2800 genes). Next, we used Distance Weighted Discrimination (DWD) to combine these three data sets together[20]. DWD is a multivariate analysis tool that is able to identify systematic biases present in separate data sets and then make a global adjustment to compensate for these biases; in essence, each separate data set is a multi-dimensional cloud of data points, and DWD takes two points clouds and shifts one such that it more optimally overlaps the other. Finally, we determined that 306 of the 1300 unique Intrinsic/UNC genes were present in the combined test set and performed a hierarchical clustering analysis of these 306 genes and 315 samples (Figure 2). We analyzed the combined test set instead of analyzing each of the 3 datasets separately because we believed this would provide more statistical power to perform multivariate analysis, and would yield more meaningful results because any finding would need to be shared/present across all 3 datasets. Remarkably, despite the loss of genes in the Intrinsic/UNC list due to the requirement of having to be present on 4 different microarray platforms, the hierarchical clustering analysis in Figure 2 identified the five main subtypes/groups corresponding to the previously defined HER2+/ER-, Basal-like, LumA, LumB and Normal Breast-like tumor groups[2,3].

As shown in previous studies, a HER2+ expression cluster was observed in the cluster analysis of the "combined test set" and contained multiple genes from the 17q11 amplicon including *HER2/ERBB2* and *GRB7* (Figure 2D). The HER2+ intrinsic subtype (pink dendrogram branch in Figure 2B) was predominantly ER-negative (i.e. HER2+/ER-) as previously shown. A Basal-like expression cluster was also present and contained genes (i.e. *c-KIT, FOXC1* and *P-Cadherin*) previously identified to be characteristic of

basal epithelial cells (Figure 2F). Using the program EASE[21], the Gene Ontology (GO) categories "extracellular space" and "extracellular region" were over-represented relative to chance in the Basal epithelial gene cluster. As shown in previous studies, a Luminal/ER+ expression cluster was present and contained *ER*, *XBP1*, *FOXA1* and *GATA3* (Figure 2C). *GATA3* has recently been shown to be somatically mutated in some ER+ breast tumors, and some of the genes in Figure 2C are *GATA3*-regulated (*FOXA1* and *TFF3*)[22], thus showing the functional clustering of a transcription factor and some of its direct targets. The Gene Ontology (GO) categories "transcription regulator activity" and "DNA binding" were over-represented relative to chance in the Luminal/ER+ gene cluster.

The most significant difference between the previous Intrinsic/Stanford gene lists and the new Intrinsic/UNC gene list was that the latter contained a large proliferation signature (Figure 2G)[23-25]. As expected, EASE analysis showed that the GO categories "mitotic cell cycle" and "M phase" were over-represented relative to chance in the proliferation signature. The inclusion of proliferation genes in the Intrinsic/UNC gene set, but not in the Intrinsic/Stanford gene set, is likely due to the fact that the Intrinsic/Stanford lists were based upon before-and-after chemotherapy paired samples of the same tumor, while the Intrinsic/UNC list was based upon paired samples taken at the same time point with respect to chemotherapy (22/26 were pre-treatment pairs). This finding suggests that tumor cell proliferation rates do vary before and after chemotherapy, but that proliferation is a reproducible and intrinsic feature of a tumor's expression profile.

A possible new tumor group (IFN) characterized by the high expression of Interferon (IFN)-regulated genes was observed in the combined test set analysis (Figure 2E). According to EASE, the GO categories "immune response" and "defense response" were over-represented relative to chance in the interferon-regulated gene cluster. This cluster contained *STAT1*, which is thought to be the transcription factor responsible for mediating IFN-regulation of gene expression[26,27]. Genes in the IFN cluster have been linked to lymph node metastasis and poor prognosis[7,13]. In summary, the Intrinsic/UNC list contained more genes than previous lists, encompasses most features of the Intrinsic/Stanford list (i.e. Basal, Luminal/ER+, and HER2-amplicon gene clusters) and adds the biologically and clinically relevant proliferation signature.

Tumor subtypes identified by the Intrinsic/UNC gene set are predictive of outcome. To determine how many biologically relevant tumor subtypes/groups might be present within the cluster in Figure 2, we used 3 criteria, which resulted in the identification of 6 potential subtypes/groups. The first criteria was the simple and obvious dendrogram branching pattern (Figure 2B) suggesting six groups. Second was the observation that each of the six groups uniquely expressed distinct sets of known biologically relevant genes including the basal, luminal/ER+, HER2-amplicon, IFN-regulated, and proliferation-associated signatures. Third was our knowledge of the previous classifications made by the Sorlie et al. 2003 Intrinsic/Stanford list of the Stanford/Norway samples: there was a high concordance (78%) between the classification of these samples made using either the Sorlie et al. 2003 Intrinsic/Stanford list or the Intrinsic/UNC list (excluding the IFN samples). Therefore, the 311

tumors/patients were stratified into six groups, and we proceeded to look for differences in outcomes and associations with other clinical parameters between these six groups. The Intrinsic/UNC gene set identified tumor groups/subtypes that were predictive of Relapse-Free Survival (RFS, Figure 3A) and Overall Survival (OS, p=0.000001, data not shown) in Kaplan-Meier survival analysis on the combined test set. As previously seen in Sorlie et al.[2,3], the LumA group had the best outcome while the HER2+/ER-, Basal-like, and LumB groups had significantly worse outcomes. The new IFN class had a Kaplan-Meier survival curve similar to that of LumB, and both showed elevated proliferation rates when compared to LumA (Figure 2G).

In the combined test set, the standard clinical parameters of ER status, node status, grade, and tumor size (note: data for clinical HER2 status was not available) were significant predictors of RFS using Kaplan-Meier analysis (Figure 4), thus showing that the act of combining three different patient sets together did not destroy the prognostic abilities of these standard markers. In a multivariate Cox proportional hazards analysis of the combined test set using these standard clinical parameters, size, grade and ER status were significant predictors of RFS (Table 1A).

To further evaluate the prognostic/predictive value of the intrinsic subtype classification, we performed multivariate Cox proportional hazards analysis of the combined test set using the six intrinsic subtypes/groups defined above and the five standard clinical parameters with RFS, OS, or DSS as the endpoint (Table 1B shows analysis for RFS). The intrinsic subtypes, when added to the multivariate model containing the standard clinical variables, resulted in a model significantly more predictive of RFS, OS, and DSS (p=0.01, 0.009, and 0.04 respectively, by the likelihood-

ratio test). In multivariate analysis for RFS (Table 1B), the Basal-like, LumB and HER2+/ER- subtypes had hazard ratios significantly greater than 1 (LumA served as the reference group), while the IFN and Normal Breast-like groups were not significant. Thus, the intrinsic subtypes classifications of LumA, LumB, Basal-like and HER2+/ER- add new and important prognostic information beyond what the standard clinical predictors provide.

Associations of the Intrinsic subtypes with clinical and biological parameters. To further characterize and better understand the intrinsic subtypes, we determined whether an association existed between intrinsic subtype and grade, node status, ER status, age, and tumor size in the combined test set. Two-way contingency table analysis showed significant association between grade and subtype, with HER2+/ER- and Basal-like tumors more likely to be grade 3 (Table 2). The Cramer's V statistic[28], which measures the strength of association between two variables in a contingency table, indicated a substantial association (Cramer's V>0.36) between grade and subtype. Two-way contingency table analysis did not show significant association between node status and subtype ($p=0.44$), but did show significant association between ER status and subtype ($p<0.0001$; Cramer's $V=0.72$) and between tumor size and subtype ($p=0.01$; Cramer's $V=0.17$). As would be expected, ER+ tumors were more likely to be LumA or LumB. As indicated by the low Cramer's V (Cramer's V<0.19 indicates a low relationship), tumor size and subtype were not strongly correlated.

To determine association between age and subtype, we used an unpaired Student's t-test to compare the average ages of diagnosis of each tumor subtype.

Interestingly, the average age of diagnosis for HER2+/ER- tumors was significantly less than that for all other tumor types. The average age of diagnosis for LumA tumors was significantly greater than that for LumB tumors.

Derivation and application of a Single Sample Predictor. A caveat to the above analyses is that our classifications were based upon hierarchical clustering, which is a powerful tool for intrinsic class discovery, but which is not suited for individual sample classification because to classify a new sample would require a reanalysis of all samples. Therefore, we wanted to create an unchanging and objective method to classify tumors according to intrinsic subtype that could be clinically applicable. To this end, we developed a Single Sample Predictor (SSP) using the combined test set hierarchically clustered using the 306 Intrinsic/UNC genes (Figure 1). For the SSP, a mean expression profile (i.e. centroid) was created for each subtype that was significant in the multivariate analysis (LumA, LumB, Basal-like, HER2+/ER-) and for the Normal Breast-like group using the combined test set (Figure 2). Next, any new sample is then compared to each Centroid and assigned by the SSP to the nearest subtype/centroid as determined by Spearman correlation (note: this SSP is based on methods developed by Tibshirani and colleagues[3,29,30]); thus, the SSP contains five different idealized profiles, and any new sample is compared to each of the five profiles and assigned a profile label (i.e. subtype name) based upon the single idealized profile it most resembled.

To validate the SSP, we tested it on two additional datasets not used previously. The first was the 60-patient Ma et al. dataset, which represents a group of early stage ER+ tamoxifen-treated patients[6]. The SSP classified these samples as follows: 2 Basal-

like, 2 HER2+/ER-, 12 Normal Breast-like, 34 LumA, and 9 LumB. The 2 Basal-like and 2 HER2+/ER- assigned samples were excluded from a survival analysis because they were too few for a meaningful survival analysis and possibly were misclassified ER-negative tumors. Among the remaining samples the SSP classification was a significant predictor of RFS (p=0.04, Figure 3B), due to the poor outcome of the LumB group. Next, we applied the SSP to a 96-sample test set of local only (surgery)-treated patients from Chang et al.[31]. The tumor groups identified by the SSP showed significant differences in RFS (Figure 3C, p=0.0006) and OS (p=0.001, data not shown) in Kaplan-Meier analysis, with the poor outcome groups as expected: LumB, Basal-like, and HER2+/ER-. Thus, the SSP identified tumor groups that are truly prognostic and have significantly different outcomes as was seen before: namely, LumA always has the most favorable outcome, while LumB, Basal-like and HER2+/ER- do poorly[2,3,9,19].

We also applied the SSP onto the 105-sample dataset used to derive the Intrinsic/UNC gene list, which is technically not a test set for the SSP because it was used to derive the Intrinsic/UNC gene set. The tumor groups identified by the SSP showed significantly different RFS (Figure 3D, p=0.02) and OS (p=0.03, data not shown) in Kaplan-Meier analysis with the poor outcome groups again being LumB, Basal-like, and HER2+/ER-. A subset of the 105-sample dataset (48 in total) had been previously characterized using an immunohistochemical (IHC) analysis[32], which showed that (1) all 18 Basal-like tumors were ER-negative and HER2-negative (defined as not having a 3+ score on HER2 IHC analysis), (2) all 18 luminal subtype tumors were ER-positive and HER2-negative, and (3) all 12 HER2+/ER- subtype tumors were ER-negative and 11 out of these 12 showed HER2-overexpression (defined as having a 3+ score on HER2 IHC

analysis). Thus, the SSP correlated with many standard clinical parameters, and was also able to identify clinically relevant groups (i.e. LumA vs. LumB) not identifiable using the standard clinical assays, thus indicating potential value as an objective classification method that should be developed further as a clinically applicable test.

DISCUSSION

The development and validation of gene sets for cancer patients requires significant resources because large training and test sets are required to achieve robust results. In fact, microarray studies are often criticized for a lack of rigorous validation due to small sample sizes[17,18]. Therefore, we utilized a previously described microarray data set combining method (Distance Weighted Discrimination) to create a large validation test set of over 300 tumors, and used it to validate a newly derived gene list for breast cancer prognostication and prediction. This approach allowed us to perform a multivariate analysis in which we show for the first time that the intrinsic subtype classification adds valuable information in the presence of five standard clinical parameters. We believe this combined test set is a valid test set for use in our analysis because after the multiple data sets were combined, the prognostic abilities of the standard clinical variables such as ER and grade remained intact.

The remarkable power of our DWD-based approach is indicated by the fact that although samples came from different platforms, hierarchical clustering analysis of the combined data set managed to group samples and genes based upon biology, and not some artifact caused by combining the data sets together. Evidence that this grouping reflected biology and not some artifact comes from (1) the finding that various Gene

Ontology terms were significantly over-represented relative to chance in individual gene clusters seen in this analysis and (2) the groupings of the samples showed inter-dataset mixing and were significant predictors of outcome in univariate Kaplan-Meier and multivariate Cox analysis. It is also remarkable that this classification was successful in predicting outcome despite the fact that the Intrinsic/UNC gene set was reduced from 1300 genes to 306 genes in the combined test set; this indicates the robust nature of the intrinsic subtypes as defined by the new Intrinsic/UNC gene list.

One of the accomplishments of this manuscript was to develop an unchanging and objective intrinsic subtype predictor that could be used routinely in the clinical setting. This was accomplished by first identifying a new intrinsic gene set and then using this set to develop the Single Sample Predictor (SSP) that was shown here to be both prognostic on the local therapy-only patient subset from Chang et al.[31] and predictive of outcomes on the ER+ tamoxifen-treated data set of Ma et al.[6]. Many other gene expression based predictors for breast cancer patients have been developed, and in a complementary publication[33], we tested the intrinsic subtype SSP developed here, relative to those predictions made by four other previously published breast cancer prognostic/predictive gene sets using a single patient/tumor set of 295 cases; the four other expression-based predictors used were (1) the "70-gene" Good *vs.* Poor outcome predictor developed by van't Veer and colleagues[5,11], (2) the "Wound-Response" profile developed by Chang et al.[31,34], (3) the "Recurrence Score (RS)" profile developed by Paik et al.[10], and (4) the 2-gene (HOXB13:IL17BR) ratio predictor developed by Ma et al.[6]. The results showed that of samples classified as Basal-like, HER2+/ER-, or LumB by the SSP, 93-100% were classified by the 70-gene, RS and Wound-Response predictors as being in each

predictor's bad prognosis group. These data suggest that a high concordance exists across these multiple predictors, in particular the RS, 70-gene and Intrinsic Subtypes; thus, the new intrinsic gene list and classification method developed here, when compared to other predictors as accomplished in Fan et al.[33], showed that a high concordance across predictors exists, which provides additional validation for each predictor.

In summary, the results of this study advances our current knowledge of the intrinsic breast tumor subtypes and provides an objective method (SSP) for prospectively classifying tumors that could be used in the clinical setting. More broadly speaking, our findings show that while the individual brushstrokes (*i.e.* genes) may sometimes show discordance across data sets, the portraits created by the combined patterns of the individual brushstrokes is conserved and recognizable across datasets because of the similarities to the family portrait[24]. Moreover, these data show that the breast tumor intrinsic subtypes identified using the Intrinsic/UNC gene list can be generalized to many different patient sets, both treated and untreated.

Acknowledgements. C.M.P. was supported by funds from the NCI Breast SPORE program to UNC-CH (P50-CA58223-09A1), by the National Institute of Environmental Health Sciences (U19-ES11391-03) and by NCI (RO1-CA-101227-01). P.S.B. was supported by NCI R33-CA97769-01, A.N by NSF Grant DMS 0406361, J.S.M. by NSF Grant DMS-0308331, O.I.O. by the National Institute of Environmental Health Sciences (P50 ESO12382), L.A.C. by NIH M01RR00046. The Breast Cancer Research Foundation supported C.M.P., O.I.O and L.R.S.

REFERENCES

1. Perou CM, Sorlie T, Eisen MB, et al: Molecular portraits of human breast tumours. Nature 406:747-52, 2000

2. Sorlie T, Perou CM, Tibshirani R, et al: Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 98:10869-74., 2001

3. Sorlie T, Tibshirani R, Parker J, et al: Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A 100:8418-23, 2003

4. Sotiriou C, Powles TJ, Dowsett M, et al: Gene expression profiles derived from fine needle aspiration correlate with response to systemic chemotherapy in breast cancer. Breast Cancer Res 4:R3, 2002

5. van 't Veer LJ, Dai H, van de Vijver MJ, et al: Gene expression profiling predicts clinical outcome of breast cancer. Nature 415:530-6, 2002

6. Ma XJ, Wang Z, Ryan PD, et al: A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen. Cancer Cell 5:607-16, 2004

7. Huang E, Cheng SH, Dressman H, et al: Gene expression predictors of breast cancer outcomes. Lancet 361:1590-6, 2003

8. Zhao H, Langerod A, Ji Y, et al: Different gene expression patterns in invasive lobular and ductal carcinomas of the breast. Mol Biol Cell 15:2523-36, 2004

9. Bertucci F, Finetti P, Rougemont J, et al: Gene expression profiling identifies molecular subtypes of inflammatory breast cancer. Cancer Res 65:2170-8, 2005

10. Paik S, Shak S, Tang G, et al: A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 351:2817-26, 2004

11. van de Vijver MJ, He YD, van't Veer LJ, et al: A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 347:1999-2009, 2002

12. Bhattacharjee A, Richards WG, Staunton J, et al: Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. Proc Natl Acad Sci U S A 98:13790-5, 2001

13. Chung CH, Parker JS, Karaca G, et al: Molecular classification of head and neck squamous cell carcinomas using patterns of gene expression. Cancer Cell 5:489-500, 2004

14. Garber ME, Troyanskaya OG, Schluens K, et al: Diversity of gene expression in adenocarcinoma of the lung. Proc Natl Acad Sci U S A 98:13784-9, 2001

15. Michiels S, Koscielny S, Hill C: Prediction of cancer outcome with microarrays: a multiple random validation strategy. Lancet 365:488-92, 2005

16. Jenssen TK, Hovig E: Gene-expression profiling in breast cancer. Lancet 365:634-5, 2005

17. Simon R, Radmacher MD, Dobbin K, et al: Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. J Natl Cancer Inst 95:14-8, 2003

18. Ioannidis JP: Microarrays and molecular research: noise discovery? Lancet 365:454-5, 2005

19. Sotiriou C, Neo SY, McShane LM, et al: Breast cancer classification and prognosis based on gene expression profiles from a population-based study. Proc Natl Acad Sci U S A 100:10393-8, 2003

20. Benito M, Parker J, Du Q, et al: Adjustment of systematic microarray data biases. Bioinformatics 20:105-114, 2004

21. Hosack DA, Dennis G, Jr., Sherman BT, et al: Identifying biological themes within lists of genes with EASE. Genome Biol 4:R70, 2003

22. Usary J, Llaca V, Karaca G, et al: Mutation of GATA3 in human breast tumors. Oncogene 23:7669-78, 2004

23. Perou CM, Jeffrey SS, van de Rijn M, et al: Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. Proc Natl Acad Sci U S A 96:9212-9217, 1999

24. Chung CH, Bernard PS, Perou CM: Molecular portraits and the family tree of cancer. Nat Genet 32 Suppl:533-40, 2002

25. Whitfield ML, Sherlock G, Saldanha AJ, et al: Identification of genes periodically expressed in the human cell cycle and their expression in tumors. Mol Biol Cell 13:1977-2000, 2002

26. Bromberg JF, Horvath CM, Wen Z, et al: Transcriptionally active Stat1 is required for the antiproliferative effects of both interferon alpha and interferon gamma. Proc Natl Acad Sci U S A 93:7673-8, 1996

27. Matikainen S, Sareneva T, Ronni T, et al: Interferon-alpha activates multiple STAT proteins and upregulates proliferation-associated IL-2Ralpha, c-myc, and pim-1 genes in human T cells. Blood 93:1980-91, 1999

28. Van Belle G, Fisher L: Biostatistics: a methodology for the health sciences (ed 2nd). Hoboken, NJ, Wiley-Interscience John Wiley & Sons, 2004

29. Bullinger L, Dohner K, Bair E, et al: Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. N Engl J Med 350:1605-16, 2004

30. Bair E, Tibshirani R: Semi-supervised methods to predict patient survival from gene expression data. PLoS Biol 2:E108, 2004

31. Chang HY, Nuyten DS, Sneddon JB, et al: Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. Proc Natl Acad Sci U S A 102:3738-43, 2005

32. Livasy CA, Karaca G, Nanda R, et al: Phenotypic evaluation of the basal-like subtype of invasive breast carcinoma. Mod Pathol, 2005

33. Fan C, Oh DS, Wessels L, et al: Different gene expression-based predictors for breast cancer patients are concordant. N Engl J Med, In Press

34. Chang HY, Sneddon JB, Alizadeh AA, et al: Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. PLoS Biol 2:E7, 2004

35. Hu Z, Troester M, Perou CM: High reproducibility using sodium hydroxide-stripped long oligonucleotide DNA microarrays. Biotechniques 38:121-4, 2005

36. Novoradovskaya N, Whitfield ML, Basehore LS, et al: Universal Reference RNA as a standard for microarray experiments. BMC Genomics 5:20, 2004

37. Yang YH, Dudoit S, Luu P, et al: Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res 30:e15, 2002

38. Tusher VG, Tibshirani R, Chu G: Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A 98:5116-21, 2001

39. Troyanskaya O, Cantor M, Sherlock G, et al: Missing value estimation methods for DNA microarrays. Bioinformatics 17:520-5, 2001

40. Diehn M, Sherlock G, Binkley G, et al: SOURCE: a unified genomic resource of functional annotations, ontologies, and gene expression data. Nucleic Acids Res 31:219-23, 2003

41. Eisen MB, Spellman PT, Brown PO, et al: Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 95:14863-8, 1998

42. Hastie T, Tibshirani R, Friedman JH: The elements of statistical learning: data mining, inference, and prediction. New York, Springer, 2001

43. Tibshirani R, Hastie T, Narasimhan B, et al: Diagnosis of multiple cancer types by shrunken centroids of gene expression. Proc Natl Acad Sci U S A 99:6567-72, 2002

TABLES

Table 3.1. Multivariate Cox proportional hazards analysis for the 315-sample combined test set in relation to Relapse-Free Survival. Multivariate Cox proportional hazards analysis of (A) standard clinical factors alone, or with (B) the Intrinsic Subtypes. Size was a binary variable (0= diameter of 2cm or less, 1= greater than 2cm); node status was a binary variable (0= no positive nodes, 1= one or more positive nodes); age was a continuous variable formatted as decade-years. Hazard ratios for Intrinsic Subtypes were calculated relative to the Luminal A subtype. Variables found to be significant ($p<0.05$) in the Cox proportional hazards model are shown in bold.

A. Relapse-Free survival

Variable	Hazard Ratio (95% CI)	p-value
Age, per decade	1.04 (0.90-1.20)	0.64
ER status	**0.59** (0.41-0.83)	**0.003**
Node status	1.41 (0.98-2.04)	0.07
Tumor grade 2 *vs*. 1	**2.41** (1.08-5.36)	**0.032**
Tumor grade 3 *vs*. 1	**3.98** (1.80-8.82)	**0.0007**
Size	**1.60** (1.31-1.95)	**<0.0001**

B. Relapse-Free survival

Variable	Hazard Ratio (95% CI)	p-value
Age, per decade	1.08 (0.94-1.24)	0.29
ER status	0.69 (0.42-1.13)	0.14
Node status	1.35 (0.92-1.98)	0.13
Tumor grade 2 *vs*. 1	1.88 (0.82-4.32)	0.14
Tumor grade 3 *vs*. 1	**2.58** (1.08-6.12)	**0.03**
Size	**1.59** (1.30-1.95)	**<0.0001**
Basal-like vs. LumA	**2.02** (1.05-3.90)	**0.036**
HER2+/ER- vs. LumA	**3.47** (1.78-6.76)	**0.0003**
LumB vs. LumA	**1.92** (1.07-3.45)	**0.028**
IFN vs. LumA	1.40 (0.67-2.91)	0.37
Normal-like vs. LumA	1.56 (0.59-4.16)	0.37

Table 3.2. Association between tumor histologic grade and intrinsic subtype in the 315-sample combined test set.

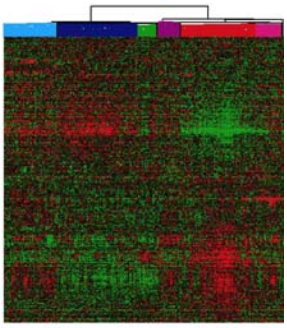
	Intrinsic Subtype				
Two-way contingency table	**LumA** (# of pts.)	**LumB** (# of pts.)	**IFN** (# of pts.)	**HER2+/ER-** (# of pts.)	**Basal-like** (# of pts.)
Grade					
1 (well)	29	2	1	0	1
2 (intermediate)	45	26	8	6	16
3 (poor)	15	32	16	21	67
Statistics for two-way contingency table analysis					
p-value†	<0.0001				
Cramer's V††	0.42				

† p-value calculated from Chi-square test on contingency table. †† Cramer's V statistic (value can range from 0 to 1) measures the strength of association between the two variables analyzed in the contingency table, with 1 indicating perfect association and 0 indicating no association.

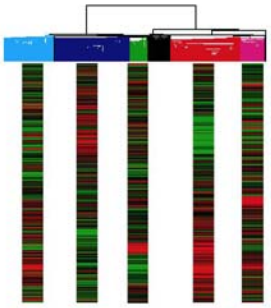
FIGURES

Training Set
A dataset of 105 breast tumor samples, 9 normal breast samples, and 26 sample pairs (each pair of samples is taken from the same patient), represented by 146 arrays, is used to derive the 1300-gene "Intrinsic/UNC" gene set.

Combined Test Set
A test set of 311 tumors and 4 normal breast samples represented by 315 arrays and 2800 genes in common, was created by combining the datasets of Sorlie et al. (2001; 2003), van't Veer et al. (2002) and Sotiriou et al. (2003). This "combined test set" was analyzed by hierarchical clustering using the subset of "Intrinsic/UNC" genes that were present within the combined test set (306 genes).



Single Sample Predictor (SSP)
The hierarchical clustering of the "combined test set" is used to create 5 Subtype Mean expression profiles (i.e. Centroids) based upon the expression of the 306 Intrinsic/UNC genes. New samples are then assigned to the nearest subtype/centroid as determined by Spearman correlation.



Validation of the SSP using 2 test datasets
The SSP is used to make subtype predictions on 2 test sets of homogenously treated patients. The resulting classifications were then analyzed using Kaplan-Meier Survival plots.

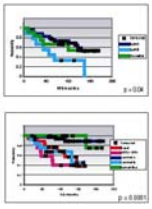


Figure 3.1. Overview of the analysis methods and datasets used.

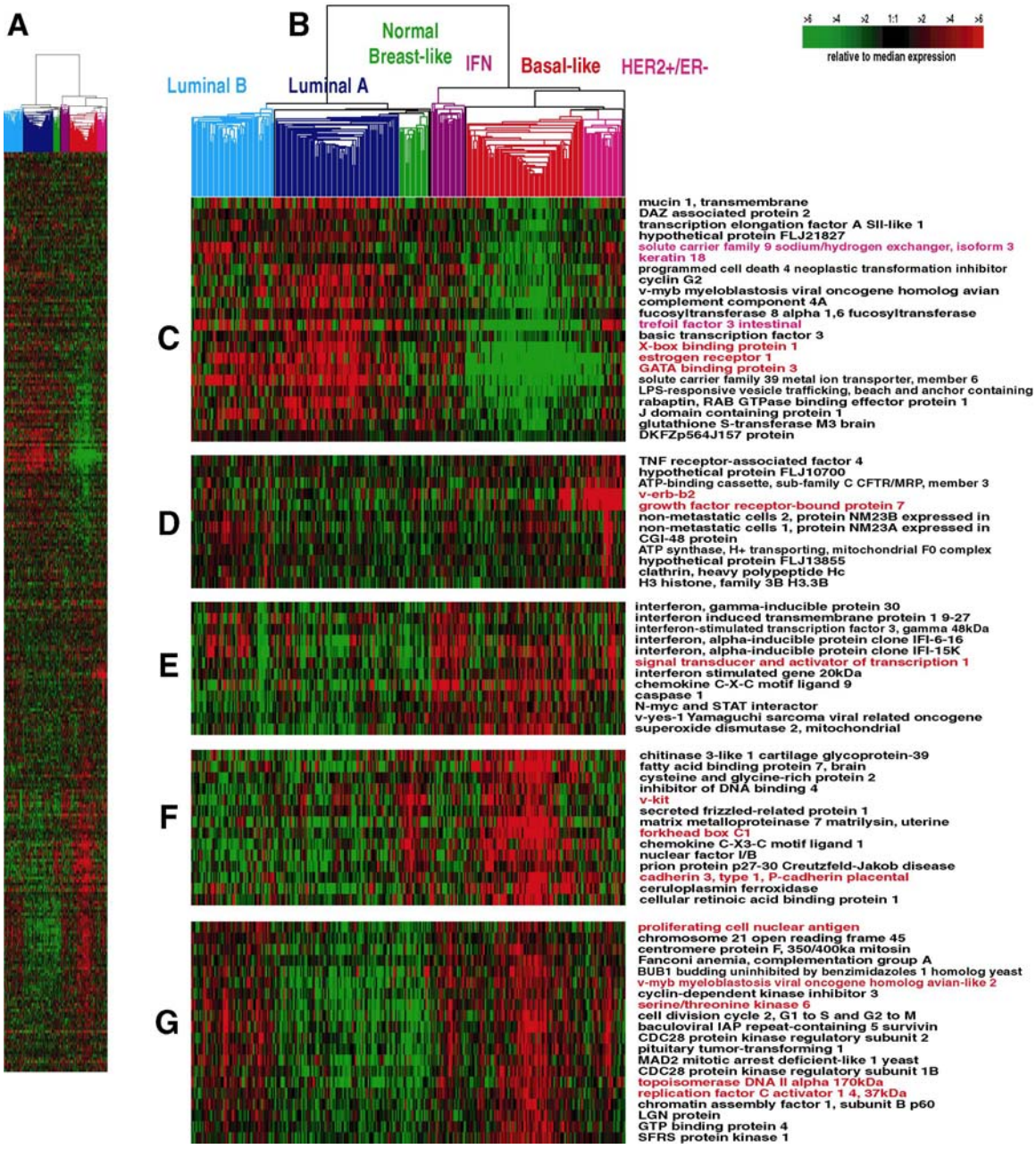


Figure 3.2. Hierarchical cluster analysis of the 315-sample combined test set using the Intrinsic/UNC gene set reduced to 306 genes. **(A)** Overview of complete cluster diagram. **(B)** Experimental sample-associated dendrogram. **(C)** Luminal/ER+ gene cluster with *GATA3*-regulated genes highlighted in pink. **(D)** *HER2* and *GRB7*-containing expression cluster. **(E)** Interferon-regulated cluster containing *STAT1*. **(F)** Basal epithelial cluster. **(G)** Proliferation cluster.

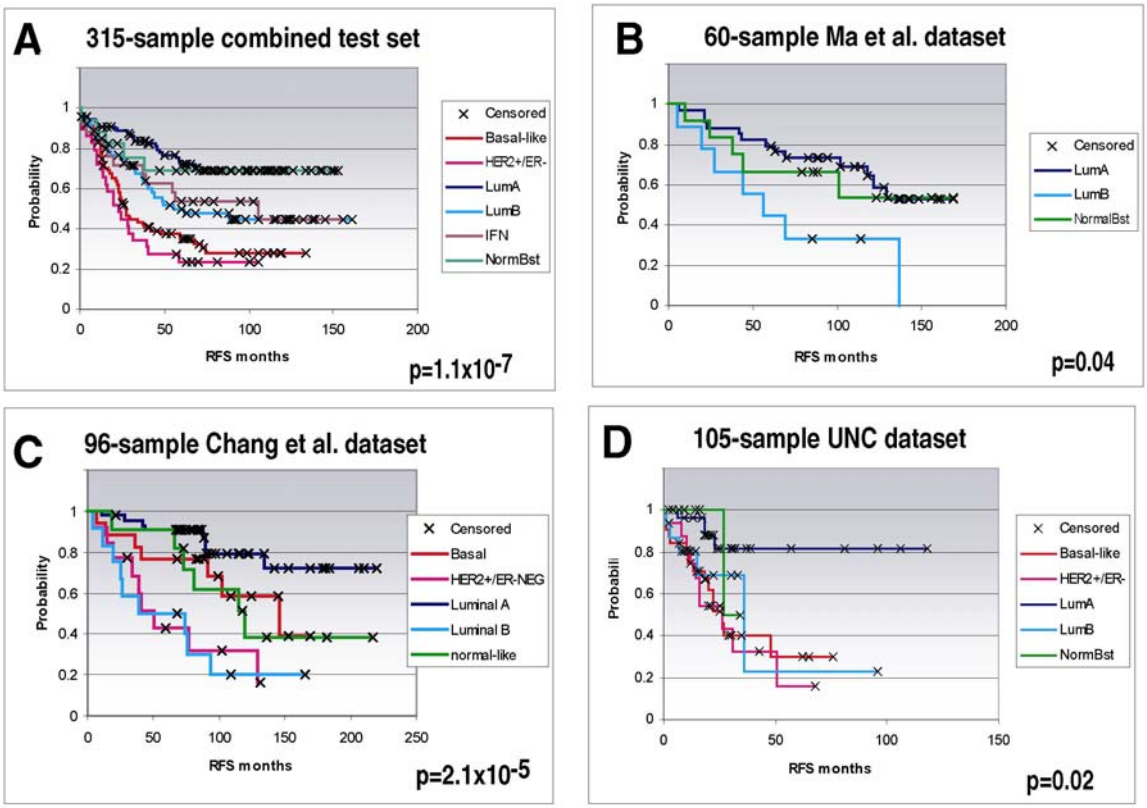


Figure 3.3. Kaplan-Meier survival curves of breast tumors classified by intrinsic subtype. Survival curves are shown for **(A)** the 315-sample combined test set classified by hierarchical clustering using the Intrinsic/UNC gene set and **(B)** the 60-sample Ma et al., **(C)** 96-sample Chang et al., and **(D)** 105-sample (used to derive the Intrinsic/UNC gene set) datasets classified by the Nearest-Centroid predictor (Single Sample Predictor).

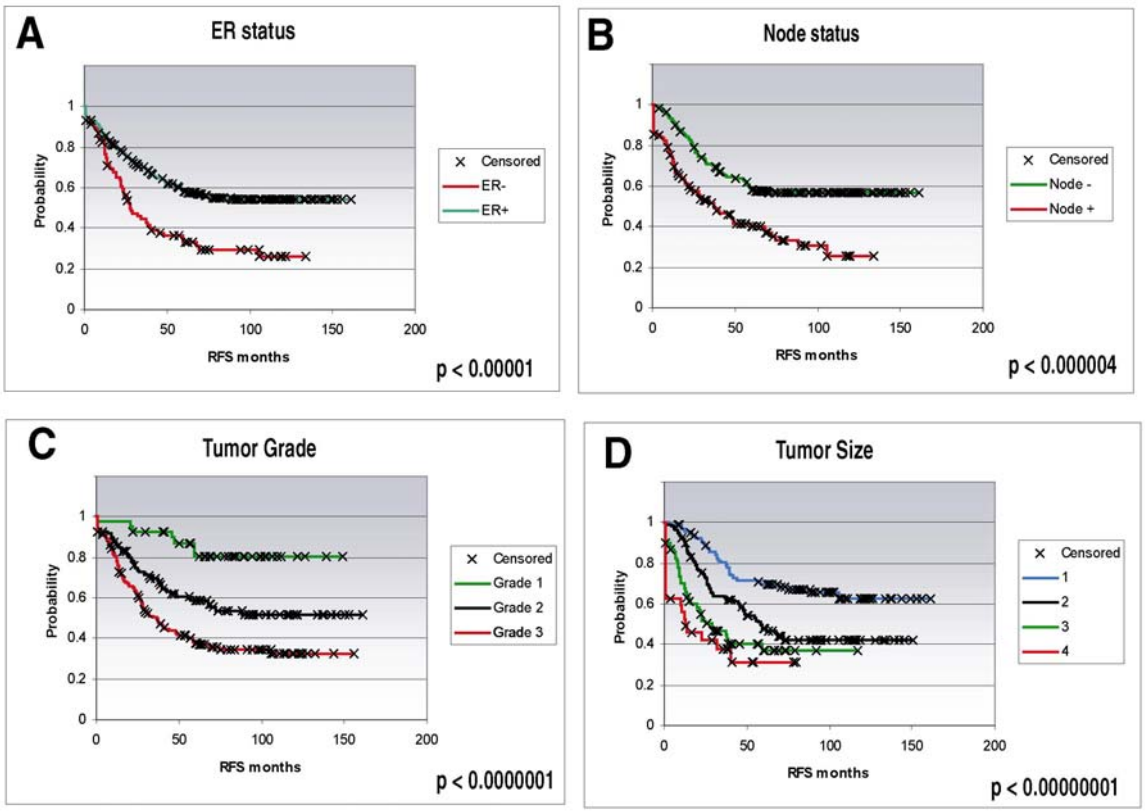


Figure 3.4. Kaplan-Meier survival curves using RFS as the endpoint, for the common clinical parameters present within the 315-sample combined test set. Survival curves are shown for **(A)** ER status, **(B)** node status, **(C)** histologic grade (1=well-differentiated, 2=intermediate, 3=poor), and **(D)** tumor size (1= diameter of 2cm or less; 2=diameter greater than 2cm and less than or equal to 5cm; 3=diameter greater than 5cm; 4=any size with direct extension to chest wall or skin).

CHAPTER 4: Different gene expression-based predictors for breast cancer patients are concordant

C. Fan*, D.S. Oh*, L. Wessels, B. Weigelt, D. Nuyten, A. Nobel, L.J. van't Veer, and C.M. Perou.

*These authors contributed equally to this work.

This work is in press in the New England Journal of Medicine. 2006.

ABSTRACT

Background. Gene expression profiling studies of primary breast tumors performed by different laboratories have resulted in the identification of many apparently different prognostic profiles/gene sets, which show little overlap in gene identity.

Methods. In order to compare the individual sample predictions made by these different gene sets, we applied to a single dataset of 295 samples, five different gene expression-based predictors: (1) Intrinsic Subtypes[1, 2], (2) 70-gene Good *vs.* Poor[3, 4], (3) Wound-Response Activated *vs.* Quiescent[5, 6], (4) Recurrence Score[7], and (5) the 2-gene ratio profile for tamoxifen-treated patients[8].

Results. There was high concordance in outcome predictions across most of these different predictors when the outcome predictions on individual samples were compared. In particular, patients of the Basal-like, HER2+/ER- and Luminal B Intrinsic Subtypes were almost all 70-gene Poor, Wound-Response Activated, and had a High Recurrence Score. The 70-gene and Recurrence Score predictors, which are beginning to be used in the clinical setting, showed 77-81% agreement.

Conclusions. These data show that even though different gene sets are being used for prognostication on breast cancer patients, four of the profiles tested here showed significant agreement in outcome predictions on individual patients and are likely tracking a common set of biological phenotypes.

INTRODUCTION

Many different gene expression studies have identified expression profiles/gene sets that are prognostic and/or predictive for breast cancer patients[2-12]. Comparison of the gene lists derived from some of these apparently similar studies show that they overlap with each other by a modest amount, if at all. The reasons for this lower-than-expected overlap are not completely known but must include differences in patient cohorts, microarray platforms and mathematical analysis methods. An important and unanswered question, however, is whether these predictors actually disagree or agree concerning outcome predictions for the individual patient. Here we describe an analysis of a single dataset on which five different prognostic/predictive gene expression-based predictors were simultaneously compared. This "across profile" analysis showed that different predictors had significant concordance when outcome predictions on individual patients were compared, despite the fact that these predictors had little gene overlap.

MATERIALS AND METHODS

Patient data set. For this study, we used a single dataset of 295 samples produced by researchers from the Netherlands Cancer Institute (NKI) using Agilent Oligo microarrays, and for which Relapse-Free Survival (RFS, scored as time to first event) and Overall Survival (OS) data were available[3-5]. The clinical information associated with these patients was obtained from the supporting website for the Chang *et al.* 2005 paper; this patient set contained predominantly Stage I and II patients who received either local therapy only (n=165), tamoxifen only (n=20), tamoxifen plus chemotherapy (n=20) or chemotherapy only (n=90).

Statistical methods. Five different prognostic/predictive gene sets (and methodologies) were tested on this single dataset and the results were recorded for each predictor on each patient (see Table 1 for a summary of the classifications made by the five predictors). The expression-based predictors used were (1) the "70-gene" Good *vs.* Poor outcome predictor developed by van't Veer, van de Vijver and colleagues[3, 4], (2) the "Wound-Response" profile developed by Chang *et al.*[5, 6], (3) the "Recurrence Score (RS)" profile developed by Paik, Shak *et al.*[7], (4) the "Intrinsic Subtype" classifications developed by Perou, Sorlie and colleagues[1, 2, 10, 13], and (5) the 2-gene (HOXB13:IL17BR) ratio predictor[8]. The RS and 2-gene ratio predictors were originally designed for outcome predictions on Estrogen Receptor (ER)-positive tamoxifen-treated patients[7, 8], and we therefore performed analysis for the ER+ patient subset on its own, in addition to the complete set of ER+ and ER– samples combined. Many other prognostic profiles exist for breast cancer patients. We excluded some of these for a lack of sufficient numbers of genes, the expression of which was captured in the NKI data set, or because the description of the predictor was too vague to be confidently applied to a new data set. Additionally, it is beyond the scope of this work to test every possible published breast cancer predictor.

For the 70-gene and Wound-Response predictions, we used the individual sample assignments provided by Chang *et al.*[5]. Briefly, the assignments made by the 70-gene and Wound-Response predictors were as follows: for the 70-gene predictor, a sample was classified according to the correlation of its expression levels of the 70 genes to a previously determined average centroid/profile of these genes in tumors from patients

with the "Good" prognosis profile. Patients with a correlation coefficient of >0.4 were classified as Good, and ≤0.4 as Poor[3, 4]. For the Wound-Response predictor, a sample was classified according to the Pearson correlation of its expression levels of the "core serum response (CSR)" genes to the serum-activated fibroblast centroid. Patients with >-0.15 correlation were classified as Wound-Response Activated, and ≤-0.15 as Wound-Response Quiescent[5].

We used a nearest centroid predictor[1] to classify tumors according to "Intrinsic Subtype". Briefly, a new "intrinsic" gene set was developed as described in Sorlie *et al.* 2003[2], using 24 new paired tumor samples assayed on Agilent Oligo microarrays, and 105 tumors in total. Next, this gene list was used in a hierarchical clustering analysis on a 311 tumor sample test set created by combining together the two-color DNA microarray data sets of Sorlie *et al.* 2001 and 2003[2, 10], van't Veer *et al.*[4] and Sotiriou *et al.*[11]. This cluster was then used as the starting point to create five Subtype Mean expression profiles/Centroids (Luminal A, Luminal B, HER2+/ER-, Basal-like and Normal-like) by averaging the gene expression profiles for the samples within each dendrogram branch/subtype. Finally, new samples like those in the Chang *et al.* dataset, are then individually compared to each centroid using the 306 intrinsic genes, and are assigned to the nearest centroid as determined by Spearman correlation. For more details on this nearest centroid predictor (also referred to as a Single Sample Predictor), the data and how to implement it, see https://genome.unc.edu/pubsup/breastTumor/.

To classify tumors using the Recurrence Score predictor, we used the microarray data for all 21 RS-genes and applied the algorithm and scaling methods described in Paik *et al.* (2004). Briefly, the expression of the 16 target genes was normalized relative to the

5 reference genes; next, the target genes were scaled as described for the qRT-PCR data, weighted averaging was performed and we then used these values and the RS algorithm to generate a Recurrence Score for each patient, which ranged from 0 to 100; scaling was done separately for the 295 patient group and for the 225 ER+ patient group. Using the cutoffs described in Paik *et al.* (0-18, 19-30, 31-100), we assigned each patient into the Low, Intermediate or High risk groups. Finally, we used the log-base-2 ratio of HOXB13:IL17BR as a means of patient stratification, using a cutoff of -0.15 as described in Ma *et al.*[8] to classify patients as having either a High or Low 2-gene ratio.

It should be noted that for the 70-gene, Wound-Response and Intrinsic Subtype profiles, a subset of the samples in this 295-sample dataset were used to train these predictors (75 of the 295 samples were previously published[4] and used to train the 70-gene profile, these same 75 samples were also part of the 311 tumor dataset used to derive the Intrinsic Subtype Centroid profiles[1], and 148 of the 295 samples were randomly selected to train the Wound-Response profile[5]). Therefore, their performance in the Kaplan-Meier and multivariate analyses described below is positively biased. However, as this paper is focused on comparing the actual predictions themselves and is not focused on identifying the "best" predictor, we believed it best to include as many samples as possible in the analysis as opposed to removing subsets of samples due to training and test set issues (if we removed training set samples, the resulting test dataset would be greatly reduced — to fewer than 147 samples and possibly as few as 72 samples). We acknowledge that the RS and 2-gene predictors are thus at a prognostic disadvantage relative to the other three because the 295 samples represent a true test set

for these two predictors; this point should be taken into consideration when interpreting the results of the survival analyses and hazard ratios from multivariate analyses.

Survival Analyses. To evaluate the prognostic value of each gene expression-based predictor, we performed univariate Kaplan-Meier analysis using the Cox-Mantel log-rank test in WinSTAT for Excel (R. Fitch Software, Staufen, Germany). We also performed a multivariate Cox proportional hazards analysis (SAS, Cary, NC) of each predictor individually in a model that included ER status (positive *vs*. negative), grade (1 *vs.* 2, and 1 *vs*. 3), node status (0 *vs.* 1-3 positive nodes, and 0 *vs*. >3 positive nodes), age (as a continuous variable), tumor size (≤2cm *vs*. >2cm), and treatment status (no adjuvant therapy *vs.* chemotherapy and/or hormonal therapy), with Relapse Free Survival (RFS, defined as time to first event) and Overall Survival (OS) as the endpoints (note: for multivariate analysis of the Intrinsic Subtypes and RS, ER status was not included as a variable because it was based upon the same microarray data that was used as part of these gene expression predictors). Two-way contingency table analyses and calculation of Cramer's V statistic were performed using WinSTAT for Excel. The Cramer's V statistic provides a quantitative measure of the strength of association between the two variables in a contingency table (which cannot be obtained from the p-value): Cramer's V values range from 0 to 1, with 0 indicating no relationship and 1 indicating perfect association. Traditionally, values between 0.36 and 0.49 indicate a substantial relationship and values >0.50 indicate a very strong relationship. The V statistic is a generalization of the more familiar phi statistic to non 2x2 contingency tables, and for 2x2 tables the V statistic is equal to the phi statistic[14].

RESULTS

Analysis of all 295 tumors. Each of the five predictors (except for the 2-gene ratio predictor), ER status, grade, tumor size (≤2cm *vs.* >2cm) and node status (0 *vs.* 1-3 *vs.* >3 nodes) were statistically significant predictors of Relapse-Free (RFS) and Overall Survival (OS) using univariate Kaplan-Meier survival analyses (Figure 1 and Table 1 shows a summary of how the five predictors classified the 295 samples). For the 2-gene ratio predictor, tumors with a High gene ratio were expected to be the poor outcome group (Figure 2 of Ma *et al.*, 2004), but this was not observed in the 295-tumor dataset (Figure 1I, J). For the other four predictors, the poor outcome groups observed in the 295-tumor dataset were as expected: (1) 70-gene Poor, (2) Wound-Response Activated, (3) High Recurrence Score, and (4) Basal-like, Luminal B, and HER2+/ER- Intrinsic Subtypes. To evaluate the prognostic value of each gene expression-based predictor, we next performed multivariate Cox proportional hazards analysis of each predictor individually in a model that included ER status, grade, node status, age, tumor size, and treatment status (Table 2). The Intrinsic Subtypes, 70-gene, Wound-Response and RS classification schemes were significant predictors in these models for both RFS and OS, showing that individually, each gene expression profile (except for the 2-gene predictor) adds new and important prognostic information beyond what the standard clinical predictors provide. In fact, the 70-gene, RS and Intrinsic Subtypes were the most predictive variables in each model as determined by the lowest nominal p-value.

As a point of reference, we next analyzed each predictor relative to the Intrinsic Subtype assignments, which is an assignment that is largely based upon an "unsupervised" analysis of breast tumor gene expression profiles (Table 3). For the 53

Basal-like tumors, all were classified as RS High, 70-gene Poor, and 50/53 were Wound-Response Activated. A nearly identical finding was also observed for the second ER-negative subtype (HER2+/ER-), and for the poor outcome, clinically ER+, Luminal B tumor group. Conversely, the Normal-like and Luminal A subtype tumors showed significant heterogeneity in terms of how samples were classified by the other predictors; however, 62/70 RS Low samples were in the Luminal A subtype. These data suggest that if a sample is classified as Basal-like, HER2+/ER-, or Luminal B, then one should be able to infer with high accuracy, that it would be classed in the bad prognosis groups of the 70-gene, Wound-Response and RS predictors.

We next compared the 70-gene, Wound-Response, RS and 2-gene predictor assignments to each other using two-way contingency table analyses (note: for these analyses, we combined the RS categories Low and Intermediate into a single group because they showed survival curves that were not significantly different [see Table 2E]). All comparisons yielded statistically significant correlations, with the least correlated profile being the 2-gene predictor. The RS, 70-gene and Wound-Response profiles were all highly correlated with each other (Table 4, Chi-square p-values <0.001). We then assessed the strength of correlation between the predictors using the Cramer's V statistic. Comparing the 70-gene *vs*. RS gave a Cramer's V=0.60 (indicates a strong relationship), RS *vs*. Wound-Response V=0.42 (substantial relationship), and the 70-gene *vs*. Wound-Response V=0.36 (substantial relationship). Thus, most tumors classified as having poor outcome by one of these three predictors were also classified as such by the other two. By comparing these Cramer's V values, the predictor showing the best agreement with the other two was the Recurrence Score (*i.e.* out of the three, the Recurrence Score was the

closest to functioning as a consensus predictor). To determine if combining the three predictors could result in a better predictor, we derived a predictor based on the majority vote of the three predictors. This predictor's performance in Kaplan-Meier analysis was comparable to that of the three predictors individually, but was not noticeably better.

Grade is an important clinical and biological feature of tumors, especially when one compares the clinical behavior of grade 1 *vs.* grade 3 breast tumors. Correlation with grade is an often-asked question regarding these new gene-expression based predictors. We therefore performed two-way contingency table analysis comparing each predictor to grade. Of the four predictors tested (70-gene, Wound-Response, 2-gene ratio, and RS as 2 classes [Low + Intermediate *vs.* High]), all showed significant correlation with grade ($p<0.001$). The profile with the strongest correlation with grade was the 70-gene, which gave a Cramer's V=0.52, next was RS (V=0.48), then Wound Response (V=0.35) and finally the 2-gene ratio (V=0.25). Thus, to varying degrees, all the predictors correlated with grade, however, it should be noted that because the 70-gene, RS, Intrinsic Subtypes and Wound-Response profiles were all significant predictors in the multivariate analyses that included grade, these predictors add prognostic information beyond that provided by grade. Moreover each of these predictors offers an assay that could be easily standardized across institutions and would be objective, quantitative, and automatable.

The 70-gene[3, 4] and RS[7, 15] predictors are the most advanced in terms of validation and are beginning to be used in the clinical setting to assist in making treatment decisions. We therefore specifically compared these two predictors to one another. A simple way to compare their predictions is to call a RS "Low" and "Intermediate" equivalent to a 70-gene "Good", a RS "High" equivalent to a 70-gene "Poor", and

determine how many samples agreed. Using this criteria, there was agreement in 239/298 (81%) of the cases. In particular, 81/103 RS "Low" and "Intermediate" patients were classified as 70-gene "Good". It should be noted that here we compared the predictors for their capacity to predict recurrence of disease in a group of patients that were both lymph node negative and positive. These two predictors, however, were developed to predict distant metastasis-free survival in lymph node-negative patients only and are either meant to be used to predict prognosis without adjuvant treatment (70-gene predictor) or for tamoxifen-treated patients (RS).

Analysis of ER+ tumors. Two of the five predictors (RS and 2-gene ratio) were specifically designed for ER+ patients only. We therefore performed similar analyses as described above (Table 1) on only those 225 patients in this dataset who were classified as ER+ (which was based on a gene expression-based cutoff using the mRNA for ER, see Chang *et al.*[5]). Again, all the gene expression-based predictors, except for the 2-gene ratio predictor, were significant predictors of RFS and OS in univariate Kaplan-Meier analysis (Figure 2). In multivariate Cox proportional hazards analyses in which each predictor was evaluated individually in a model that included the standard clinical variables, the 70-gene, Wound-Response, Intrinsic Subtypes Luminal A *vs.* B, and RS added significant prognostic information regarding RFS and OS; again, each gene expression predictor typically gave the lowest p-value when compared with the traditional clinical variables (Table 5). Interestingly, when samples within the ER+ patient subset were classified according to Intrinsic Subtype (Table 6), 7 were Basal-like and 18 were HER2+/ER-, suggesting that approximately 10% of ER+ samples fell into

tumor subtype categories that we consider to be "ER-negative" as determined by hierarchical clustering analysis.

As was done for the 295-sample set, we did a pair-wise comparison of the 70-gene, Wound-Response, RS and 2-gene predictor assignments for the 225 ER+ patients using two-way contingency table analyses. All comparisons yielded statistically significant correlations except for the 2-gene predictor (Table 7). The RS, 70-gene and Wound-Response profiles were highly correlated with each other ($p<0.001$); the observed Cramer's V values were 0.54 for 70-gene *vs*. RS (very strong relationship), 0.38 for RS *vs*. Wound-Response (substantial relationship), and 0.34 for 70-gene *vs*. Wound-Response (moderate relationship). From the Cramer's V values, we again see that the predictor showing the best agreement with the other two predictors was the Recurrence Score. We again derived a predictor based on the majority vote of the three predictors and as was seen before, its performance in Kaplan-Meier analysis was comparable to the three individual predictors, but was not noticeably better.

When RS "Low" and "Intermediate" *vs*. "High" classification was compared to the 70-gene "Good" *vs*. "Poor", 173/225 samples (77%) showed agreement. In particular, of the 105 RS "Low" or "Intermediate" patients, 83 were classified as 70-gene "Good". Finally, we did not perform any multivariate Cox proportional hazards analyses using all predictors simultaneously to identify the "optimal model" because we believe that this would not be a fair test for either of those predictors (RS and 2-gene ratio) for which this was a true test set, or for those that were derived using a different platform (RS, 2-gene and Intrinsic Subtypes).

DISCUSSION

A plethora of gene expression-based prognosticators are being developed for outcome predictions in breast cancer patients. In this study, we took advantage of a single dataset that had enough genes assayed to allow the simultaneous analysis of five different gene expression-based predictors and determined that most of these assays were making similar predictions; that is, if one predictor assigned a sample to its "poor outcome" group, then another predictor also assigned that same sample to its "poor outcome" group. In the case of tumors of the Intrinsic Subtype classes Basal-like, HER2+/ER- and Luminal B, the assignments made by the 70-gene, RS and Wound-Response predictors were almost homogeneously into the relevant poor outcome groups (regardless of ER status). It is only within the Luminal A and Normal-like Intrinsic Subtypes where variability in outcome predictions was found. Of the five predictors analyzed in this report, only the 2-gene predictor failed to identify statistically significant outcome differences on this dataset; Reid *et al.* also reported that on their independent dataset of ER+ tamoxifen-treated patients, the 2-gene predictor failed to detect outcome differences[16].

When the 70-gene, Wound-Response, RS and 2-gene predictors were compared in a pair-wise fashion, it was determined that the 70-gene, Wound-Response and RS classifications were highly concordant. When the 70-gene and RS predictors were compared relative to each other, they agreed on sample predictions for 77% (ER+) and 81% (ER+ and ER-) of the patients. These analyses suggest that even though there is very little gene overlap (the 70-gene and RS overlap by only one gene: *SCUBE2*), and different algorithms are used, the majority of breast cancer patients would receive similar

outcome predictions when these different methods and models are used. It is also likely that the RS predictor, originally developed for only ER+ patients, "worked" on all patients because we can see that almost all (69/70) ER-negative patients were classified as being in the poor outcome RS High group (and they in fact do show poor outcomes). The outcome predictions by the different methods (with the exception of the 2-gene predictor) largely overlap when evaluated by the multivariate Cox proportional hazards analyses (95% Confidence Intervals of Hazard Ratios in Table 2). The proportion of patients identified as "poor outcome" or "good outcome" as outlined above is also highly similar. The discordance of up to 20% of the patients in different categories leads to slight differences in outcome prediction and underlines the need of further validations. In the coming years a NCI (PACCT), as well as an European Union (TRANSBIG-MINDACT) randomized clinical trial will prospectively address the power of the RS and 70-gene test respectively.

We believe that despite the lack of gene overlap, the different gene sets/profiles are making similar predictions largely because they are tracking common cellular phenotypes that are reflective of the dominant biology of breast tumors, which encompasses the consistent differences seen in ER-positive (i.e. Luminal) *vs*. ER-negative breast cancers (Basal-like and HER2+/ER-). While these distinctions are correlated with grade, which is another common biological phenotype that these gene profiles are tracking, it is also clear that these profiles are providing additional information beyond that provided by grade, as evidenced by their significance in Cox regression analyses. These findings also show that prognostic profiles can be readily detected by a great number of genes, and any sufficiently representative subset of these

genes could potentially be used as a predictor. This phenomenon has been observed in normal tissues: Son *et al.* reported that approximately 19,000 genes are differentially expressed between different organs, and any sufficiently large (approximately 100) randomly selected subset could reproduce the hierarchical clustering pattern produced when using the full gene set[17]. An important implication of these findings is that when comparing two "profiles", overlap in gene identity is not a good measure of reproducibility and that individual sample classifications is the relevant measure of concordance.

We find these results encouraging and interpret them to mean that although different gene sets are being used, they are each tracking a common set of biological characteristics that are present across different breast cancer patient sets and are making similar outcome predictions. The next question to ask is what outcome predictor(s) should be used and why. The answer cannot be determined based upon the analyses in this report, but the guiding principle should be to use the predictor(s) that will predict patient outcomes and assist in making therapeutic decisions (*i.e.* predictive assays) within a specified group. For example, if a patient is determined to be of the Intrinsic Subtypes of Basal-like, HER2+/ER- or Luminal B, or "poor outcome" by the RS or 70-gene test, then there is, as yet, little need to perform the other assays because they would all indicate a poor prognosis. Thus, future studies should focus on developing assays that can prognosticate within HER2+/ER-, Basal-like, and Luminal B patients. For patients of the Luminal A and Normal-like Intrinsic Subtypes, the RS, 70-gene and Wound-Response profiles provided additional information (Tables 3 and 6) that may be useful in guiding treatment decision. For example, Luminal A patients with a Low RS might be selected to

receive only hormone therapy because a Low RS is associated with good outcomes in tamoxifen-treated patients and little benefit from chemotherapy[7, 15]. Retrospective and prospective studies must now be done to determine which is the "best" predictor, but it is likely that the "best" predictor may be a combination of two or more different predictors.

REFERENCES

1. Hu Z, Fan C, Oh DS, et al. A New Breast Tumor Intrinsic Gene List Identifies Novel Characteristics that are Conserved Across Microarray Platforms. Submitted 2006.

2. Sorlie T, Tibshirani R, Parker J, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A 2003;100(14):8418-23.

3. van de Vijver MJ, He YD, van't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 2002;347(25):1999-2009.

4. van 't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002;415(6871):530-6.

5. Chang HY, Nuyten DS, Sneddon JB, et al. Robustness, scalability, and integration of a Wound-Response gene expression signature in predicting breast cancer survival. Proc Natl Acad Sci U S A 2005;102(10):3738-43.

6. Chang HY, Sneddon JB, Alizadeh AA, et al. Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. PLoS Biol 2004;2(2):E7.

7. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 2004;351(27):2817-26.

8. Ma XJ, Wang Z, Ryan PD, et al. A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen. Cancer Cell 2004;5(6):607-16.

9. Bertucci F, Finetti P, Rougemont J, et al. Gene expression profiling identifies molecular subtypes of inflammatory breast cancer. Cancer Res 2005;65(6):2170-8.

10. Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 2001;98(19):10869-74.

11. Sotiriou C, Neo SY, McShane LM, et al. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. Proc Natl Acad Sci U S A 2003;100(18):10393-8.

12. Wang Y, Klijn JG, Zhang Y, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet 2005;365(9460):671-9.

13. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. Nature 2000;406(6797):747-52.

14. Van Belle G, Fisher L. Biostatistics: a methodology for the health sciences. 2nd ed. Hoboken, NJ: Wiley-Interscience John Wiley & Sons; 2004.

15. Paik S, Shak S, Tang G, et al. Expression of the 21 genes in the Recurrance Score assay and prediction benefits from tamoxifen in NSABP study B-14 and chemotherapy in NSABP study B-20. In: San Antionio Breast Cancer Symposium; 2004 December 2005; San Antionio, TX: on line abstract; 2004.

16. Reid JF, Lusa L, De Cecco L, et al. Limits of predictive models using microarray data for breast cancer clinical treatment outcome. J Natl Cancer Inst 2005;97(12):927-30.

17. Son CG, Bilke S, Davis S, et al. Database of mRNA gene expression profiles of multiple human organs. Genome Res 2005;15(3):443-50.

TABLES

Table 4.1. Summary of the classifications of the NKI patient dataset using five different gene expression-based profiles.

Dataset **Classification Method**	**295-sample dataset** number of patients (%)	**ER+ 225-sample dataset** number of patients (%)
Intrinsic Subtype		
Luminal A	123 (41.7%)	121 (53.8%)
Luminal B	55 (18.6%)	55 (24.4%)
Normal-like	29 (9.8%)	24 (10.7%)
HER2+/ER-	35 (11.9%)	18 (8.0%)
Basal-like	53 (18.0%)	7 (3.1%)
Recurrence Score		
Low	70 (23.7%)	87 (38.7%)
Intermediate	33 (11.2%)	18 (8.0%)
High	192 (65.1%)	120 (53.3%)
70-gene		
Good	115 (39.0%)	113 (50.2%)
Poor	180 (61.0%)	112 (49.8%)
Wound Response		
Quiescent	67 (29.8%)	60 (26.7%)
Activated	228 (77.3%)	165 (73.3%)
2-gene ratio		
Low	137 (46.4%)	122 (54.2%)
High	158 (53.6%)	103 (45.8%)

Table 4.2. Multivariate Cox proportional hazards analysis for the 295-sample Chang *et al.* 2005 dataset in relation to Relapse-Free Survival and Overall Survival. Multivariate Cox proportional hazards analysis of (A) standard clinical prognostic factors alone, or with (B) the 70-gene predictor, (C) Wound-response predictor, (D) Ma *et al.* 2-gene ratio predictor, (E) Recurrence Score, or (F) Intrinsic Subtypes in relation to Relapse-Free Survival and Overall Survival. Size was a binary variable (0= diameter of 2cm or less, 1= greater than 2cm); age was a continuous variable formatted as decade-years; hazard ratios for Intrinsic Subtypes were calculated relative to the Luminal A subtype. Variables found to be significant ($p<0.05$) in the Cox proportional hazards model are shown in bold.

A.	**Relapse-Free survival**		**Overall survival**	
Variable	**Hazard Ratio (95% CI)**	**p-value**	**Hazard Ratio (95% CI)**	**p-value**
Age, per decade	**0.59** (0.43-0.82)	**0.001**	**0.67** (0.45-0.98)	**0.042**
ER status	**0.65** (0.42-0.99)	**0.045**	**0.44** (0.27-0.71)	**0.001**
Tumor grade 2 *vs*. 1	**2.45** (1.33-4.50)	**0.004**	**4.31** (1.49-12.47)	**0.007**
Tumor grade 3 *vs*. 1	**2.53** (1.35-4.74)	**0.004**	**5.96** (2.06-17.21)	**0.001**
Size	1.40 (0.96-2.05)	0.083	1.52 (0.94-2.43)	0.086
1-3 *vs*. 0 positive nodes	1.32 (0.72-2.41)	0.37	1.06 (0.48-2.36)	0.88
>3 *vs*. 0 positive nodes	**2.24** (1.12-4.49)	**0.023**	1.85 (0.78-4.38)	0.16
Hormonal or chemotherapy *vs*. no adjuvant therapy	0.56 (0.31-1.01)	0.055	0.81 (0.38-1.74)	0.59

B.	Relapse-Free survival		Overall survival	
Variable	**Hazard Ratio (95% CI)**	**p-value**	**Hazard Ratio (95% CI)**	**p-value**
Age, per decade	**0.64** (0.46-0.88)	**0.006**	0.71 (0.48-1.05)	0.085
ER status	0.86 (0.56-1.31)	0.47	**0.59** (0.36-0.95)	**0.031**
Tumor grade 2 *vs*. 1	1.57 (0.82-2.97)	0.17	2.55 (0.86-7.63)	0.093
Tumor grade 3 *vs*. 1	1.32 (0.68-2.59)	0.41	2.84 (0.94-8.54)	0.064
Size	1.44 (0.99-2.11)	0.059	1.58 (0.98-2.53)	0.058
1-3 *vs*. 0 positive nodes	1.20 (0.66-2.18)	0.55	1.01 (0.46-2.21)	0.97
>3 *vs*. 0 positive nodes	**2.19** (1.07-4.47)	**0.032**	1.97 (0.81-4.79)	0.14
Hormonal or chemotherapy *vs*. no adjuvant therapy	**0.54** (0.30-0.99)	**0.048**	0.75 (0.34-1.62)	0.46
70-gene predictor (poor *vs*. good)	**3.44** (1.98-5.99)	**<0.001**	**4.71** (2.02-11.00)	**<0.001**

C.	Relapse-Free survival		Overall survival	
Variable	**Hazard Ratio (95% CI)**	**p-value**	**Hazard Ratio (95% CI)**	**p-value**
Age, per decade	**0.56** (0.40-0.77)	**<0.001**	**0.62** (0.41-0.92)	**0.019**
ER status	0.69 (0.45-1.06)	0.089	**0.46** (0.28-0.76)	**0.002**
Tumor grade 2 *vs*. 1	**1.89** (1.02-3.52)	**0.045**	**3.32** (1.13-9.71)	**0.028**
Tumor grade 3 *vs*. 1	**1.92** (1.02-3.62)	**0.045**	**4.46** (1.53-13.00)	**0.006**
Size	1.39 (0.95-2.03)	0.090	1.52 (0.95-2.44)	0.083
1-3 *vs*. 0 positive nodes	1.19 (0.65-2.17)	0.58	0.96 (0.43-2.12)	0.92
>3 *vs*. 0 positive nodes	1.74 (0.85-3.57)	0.13	1.48 (0.61-3.59)	0.38
Hormonal or chemotherapy *vs*. no adjuvant therapy	0.58 (0.32-1.07)	0.080	0.84 (0.39-1.81)	0.65
Wound-response signature (activated *vs*. quiescent)	**2.88** (1.50-5.52)	**0.002**	**3.25** (1.27-8.27)	**0.014**

D.	Relapse-Free survival		Overall survival	
Variable	**Hazard Ratio (95% CI)**	**p-value**	**Hazard Ratio (95% CI)**	**p-value**
Age, per decade	**0.60** (0.43-0.82)	**0.001**	**0.67** (0.45-0.98)	**0.042**
ER status	**0.63** (0.41-0.98)	**0.040**	**0.44** (0.26-0.72)	**0.001**
Tumor grade 2 *vs*. 1	**2.43** (1.32-4.47)	**0.004**	**4.31** (1.49-12.48)	**0.007**
Tumor grade 3 *vs*. 1	**2.57** (1.37-4.82)	**0.003**	**5.96** (2.06-17.23)	**0.001**
Size	1.40 (0.96-2.05)	0.082	1.52 (0.94-2.43)	0.086
1-3 *vs*. 0 positive nodes	1.32 (0.72-2.42)	0.36	1.06 (0.48-2.36)	0.88
>3 *vs*. 0 positive nodes	**2.26** (1.13-4.54)	**0.022**	1.85 (0.78-4.38)	0.16
Hormonal or chemotherapy *vs*. no adjuvant therapy	0.55 (0.30-1.00)	0.051	0.81 (0.38-1.74)	0.59
Ma *et al*. 2-gene ratio (high *vs*. low)	0.91 (0.61-1.34)	0.62	1.00 (0.61-1.63)	0.99

E.	Relapse-Free survival		Overall survival	
Variable	**Hazard Ratio (95% CI)**	**p-value**	**Hazard Ratio (95% CI)**	**p-value**
Age, per decade	**0.57** (0.42-0.79)	**<0.001**	**0.63** (0.42-0.94)	**0.023**
Tumor grade 2 *vs*. 1	1.61 (0.85-3.04)	0.14	2.95 (0.99-8.73)	0.051
Tumor grade 3 *vs*. 1	1.50 (0.79-2.86)	0.21	**3.81** (1.30-11.1)	**0.014**
Size	**1.51** (1.03-2.20)	**0.035**	**1.66** (1.03-2.67)	**0.036**
1-3 *vs*. 0 positive nodes	1.24 (0.68-2.26)	0.48	0.95 (0.43-2.09)	0.90
>3 *vs*. 0 positive nodes	**2.10** (1.04-4.25)	**0.039**	1.59 (0.66-3.82)	0.30
Hormonal or chemotherapy *vs*. no adjuvant therapy	**0.54** (0.30-0.98)	**0.044**	0.80 (0.37-1.73)	0.57
Intermediate *vs*. Low recurrence score	1.81 (0.70-4.68)	0.22	1.81 (0.39-8.27)	0.45
High *vs*. Low recurrence score	**4.27** (2.05-8.92)	**<0.001**	**6.14** (1.84-20.4)	**0.003**

F.	Relapse-Free survival		Overall survival	
Variable	**Hazard Ratio (95% CI)**	**p-value**	**Hazard Ratio (95% CI)**	**p-value**
Age, per decade	**0.59** (0.42-0.82)	**0.002**	0.67 (0.45-1.00)	0.051
Tumor grade 2 *vs*. 1	1.80 (0.96-3.39)	0.068	**3.51** (1.19-10.36)	**0.023**
Tumor grade 3 *vs*. 1	1.80 (0.92-3.50)	0.087	**4.47** (1.48-13.49)	**0.008**
Size	**1.55** (1.05-2.29)	**0.027**	1.55 (0.96-2.51)	0.076
1-3 *vs*. 0 positive nodes	1.20 (0.65-2.21)	0.55	1.01 (0.45-2.28)	0.98
>3 *vs*. 0 positive nodes	2.01 (0.96-4.21)	0.064	1.81 (0.73-4.50)	0.20
Hormonal or chemotherapy *vs*. no adjuvant therapy	**0.49** (0.26-0.92)	**0.025**	0.69 (0.31-1.54)	0.37
Luminal B	**3.79** (2.17-6.61)	**<0.001**	**2.55** (1.25-5.22)	**0.010**
Normal-like	**2.86** (1.49-5.50)	**0.002**	2.00 (0.76-5.31)	0.16
Her2+/ER-	**3.16** (1.61-6.18)	**<0.001**	**3.54** (1.59-7.85)	**0.002**
Basal-like	**2.45** (1.33-4.51)	**0.004**	**3.05** (1.49-6.27)	**0.002**

Table 4.3. Comparison of predictors for all 295 samples.

Intrinsic Subtype	#	Recurrence Score		70-gene		Wound-Response		2-gene	
		Low	0	good	0	Quiescent	3	Low	11
		Intermediate	0						
Basal-like	53	high	53	poor	53	Activated	50	high	42
		Low	62	good	87	Quiescent	45	Low	78
		Intermediate	25						
Luminal A	123	high	36	poor	36	Activated	78	high	45
		Low	1	good	9	Quiescent	4	Low	30
		Intermediate	4						
Luminal B	55	high	50	poor	46	Activated	51	high	25
		Low	0	good	3	Quiescent	0	Low	7
		Intermediate	0						
HER2+/ER-	35	high	35	poor	32	Activated	35	high	28
		Low	7	good	16	Quiescent	15	Low	11
		Intermediate	4						
Normal-like	29	high	18	poor	13	Activated	14	high	18

Table 4.4. Two-way contingency table analysis measuring the association between the 70-gene, Wound-response, and Recurrence Score predictors in the 295-sample dataset.

A.

Two-way contingency table	**Wound-Response**	
	Quiescent	Activated
	(# of patients)	
70-gene predictor		
Good	48	67
Poor	19	161
Statistics for two-way contingency table analysis		
p-value†	<0.001	
Cramer's V††	0.36	

B.

Two-way contingency table	**Recurrence Score**	
	Low or Int.	High
	(# of patients)	
70-gene predictor		
Good	81	34
Poor	22	158
Statistics for two-way contingency table analysis		
p-value†	<0.001	
Cramer's V††	0.60	

C.

Two-way contingency table	**Recurrence Score**	
	Low or Int. (# of patients)	High
Wound Response		
Quiescent	48	19
Activated	55	173
Statistics for two-way contingency table analysis		
p-value†	<0.001	
Cramer's V††	0.42	

† p-value calculated from Chi-square test on contingency table. †† Cramer's V statistic (value can range from 0 to 1) measures the strength of association between the two variables analyzed in the contingency table, with 1 indicating perfect association and 0 indicating no association.

Table 4.5. Multivariate Cox proportional hazards analysis for the 225 ER+ samples in the Chang *et al.* 2005 dataset in relation to Relapse-Free Survival and Overall Survival. Multivariate Cox proportional hazards analysis of (A) standard clinical prognostic factors alone or with (B) the 70-gene predictor, (C) Wound-response predictor, (D) Ma et al.'s 2-gene predictor, (E) Recurrence Score predictor, or (F) intrinsic subtypes in relation to Relapse-Free Survival and Overall Survival. Size was a binary variable (0= diameter of 2cm or less, 1= greater than 2cm); age was a continuous variable formatted as decade-years; hazard ratios for intrinsic subtypes were calculated relative to the Luminal A subtype. Variables found to be significant ($p<0.05$) in the Cox proportional hazards model are shown in bold.

A.	**Relapse-Free survival**		**Overall survival**	
Variable	**Hazard Ratio (95% CI)**	**p-value**	**Hazard Ratio (95% CI)**	**p-value**
Age, per decade	**0.54** (0.35-0.82)	**0.005**	**0.56** (0.31, 0.99)	**0.047**
Tumor grade 2 *vs*. 1	**2.11** (1.12-3.98)	**0.021**	**3.28** (1.08, 9.94)	**0.035**
Tumor grade 3 *vs*. 1	**2.83** (1.49-5.38)	**0.002**	**7.36** (2.51, 21.5)	**<0.001**
Size	1.41 (0.88-2.26)	0.15	1.33 (0.71, 2.49)	0.36
1-3 *vs*. 0 positive nodes	**2.11** (1.06-4.21)	**0.034**	2.11 (0.81, 5.53)	0.13
>3 *vs*. 0 positive nodes	**2.92** (1.26-6.73)	**0.012**	2.16 (0.69, 6.72)	0.18
Hormonal or chemotherapy *vs*. no adjuvant therapy	**0.37** (0.18-0.73)	**0.004**	0.52 (0.20, 1.32)	0.17

B.

	Relapse-Free survival		Overall survival	
Variable	**Hazard Ratio (95% CI)**	**p-value**	**Hazard Ratio (95% CI)**	**p-value**
Age, per decade	**0.65** (0.42-0.99)	**0.047**	0.68 (0.38, 1.22)	0.20
Tumor grade 2 *vs*. 1	1.24 (0.63-2.44)	0.53	1.75 (0.55, 5.55)	0.34
Tumor grade 3 *vs*. 1	1.20 (0.59-2.48)	0.61	2.72 (0.85, 8.66)	0.091
Size	1.45 (0.91-2.31)	0.12	1.41 (0.76, 2.61)	0.28
1-3 *vs*. 0 positive nodes	1.73 (0.88-3.40)	0.11	1.81 (0.71, 4.60)	0.21
>3 *vs*. 0 positive nodes	**2.70** (1.12-6.49)	**0.027**	2.32 (0.69, 7.80)	0.17
Hormonal or chemotherapy *vs*. no adjuvant therapy	**0.37** (0.18-0.74)	**0.005**	0.47 (0.18, 1.22)	0.12
70-gene predictor (poor *vs*. good)	**3.88** (2.15-7.02)	**<0.001**	**5.47** (2.13, 14.1)	**<0.001**

C.

	Relapse-Free survival		Overall survival	
Variable	**Hazard Ratio (95% CI)**	**p-value**	**Hazard Ratio (95% CI)**	**p-value**
Age, per decade	**0.50** (0.32-0.78)	**0.002**	**0.49** (0.27, 0.90)	**0.022**
Tumor grade 2 *vs*. 1	1.60 (0.83-3.05)	0.16	2.41 (0.78, 7.39)	0.12
Tumor grade 3 *vs*. 1	**2.10** (1.09-4.05)	**0.026**	**5.24** (1.76, 15.6)	**0.003**
Size	1.39 (0.87-2.21)	0.17	1.32 (0.71, 2.46)	0.38
1-3 *vs*. 0 positive nodes	1.87 (0.94-3.73)	0.076	1.85 (0.71, 4.82)	0.21
>3 *vs*. 0 positive nodes	2.19 (0.92-5.20)	0.075	1.61 (0.50, 5.17)	0.42
Hormonal or chemotherapy *vs*. no adjuvant therapy	**0.39** (0.20-0.78)	**0.008**	0.56 (0.22, 1.45)	0.23
Wound-response (activated *vs*. quiescent)	**2.95** (1.42-6.14)	**0.004**	**4.03** (1.20, 13.5)	**0.024**

D.

	Relapse-Free survival		Overall survival	
Variable	**Hazard Ratio (95% CI)**	**p-value**	**Hazard Ratio (95% CI)**	**p-value**
Age, per decade	**0.54** (0.35-0.84)	**0.006**	**0.56** (0.31, 0.99)	**0.048**
Tumor grade 2 *vs*. 1	**2.08** (1.10-3.93)	**0.023**	**3.28** (1.08, 9.92)	**0.036**
Tumor grade 3 *vs*. 1	**2.99** (1.55-5.74)	**0.001**	**7.42** (2.51, 21.9)	**<0.001**
Size	1.42 (0.89-2.26)	0.14	1.34 (0.72, 2.49)	0.36
1-3 *vs*. 0 positive nodes	**2.15** (1.08-4.29)	**0.030**	2.12 (0.80, 5.55)	0.13
>3 *vs*. 0 positive nodes	**2.99** (1.29-6.92)	**0.010**	2.16 (0.69, 6.70)	0.18
Hormonal or chemotherapy *vs*. no adjuvant therapy	**0.36** (0.18-0.72)	**0.004**	0.52 (0.20, 1.32)	0.17
Ma *et al*. 2-gene ratio (high *vs*. low)	0.81 (0.51-1.29)	0.38	0.97 (0.52, 1.79)	0.91

E.

	Relapse-Free survival		Overall survival	
Variable	**Hazard Ratio (95% CI)**	**p-value**	**Hazard Ratio (95% CI)**	**p-value**
Age, per decade	**0.49** (0.32-0.75)	**0.001**	**0.50** (0.28, 0.90)	**0.021**
Tumor grade 2 *vs*. 1	1.42 (0.72, 2.79)	0.32	1.83 (0.58, 5.77)	0.30
Tumor grade 3 *vs*. 1	1.69 (0.83, 3.41)	0.15	**3.26** (1.04, 10.2)	**0.042**
Size	1.52 (0.96, 2.42)	0.073	1.54 (0.83, 2.86)	0.17
1-3 *vs*. 0 positive nodes	**1.97** (1.00, 3.87)	**0.049**	1.75 (0.68, 4.47)	0.24
>3 *vs*. 0 positive nodes	**3.10** (1.32, 7.26)	**0.009**	2.06 (0.66, 6.39)	0.21
Hormonal or chemotherapy *vs*. no adjuvant therapy	**0.40** (0.20, 0.80)	**0.009**	0.62 (0.24, 1.58)	0.32
Intermediate vs. Low recurrence score	0.82 (0.27, 2.46)	0.72	1.42 (0.27, 7.50)	0.68
High *vs*. Low recurrence score	**2.59** (1.44, 4.65)	**0.001**	**4.95** (1.82, 13.4)	**0.002**

F.	Relapse-Free survival		Overall survival	
Variable	**Hazard Ratio (95% CI)**	**p-value**	**Hazard Ratio (95% CI)**	**p-value**
Age, per decade	**0.52** (0.33, 0.81)	**0.004**	0.57 (0.31, 1.04)	0.065
Tumor grade 2 *vs*. 1	1.40 (0.72, 2.71)	0.32	2.48 (0.80, 7.70)	0.12
Tumor grade 3 *vs*. 1	1.62 (0.82, 3.18)	0.16	**4.92** (1.62, 14.9)	**0.005**
Size	**1.66** (1.03, 2.65)	**0.036**	1.51 (0.80, 2.82)	0.20
1-3 *vs*. 0 positive nodes	1.82 (0.92, 3.59)	0.085	1.89 (0.73, 4.89)	0.19
>3 *vs*. 0 positive nodes	2.34 (0.96, 5.68)	0.061	1.98 (0.60, 6.54)	0.26
Hormonal or chemotherapy *vs*. no adjuvant therapy	**0.34** (0.16, 0.68)	**0.003**	0.49 (0.18, 1.29)	0.15
Luminal B	**4.40** (2.47, 7.84)	**<0.001**	**2.81** (1.33, 5.91)	**0.006**
Normal-like, HER2+/ER-, or Basal-like	**2.51** (1.38, 4.58)	**0.003**	1.92 (0.84, 4.37)	0.12

Table 4.6. Comparison of predictors for 225 ER+ patients.

Intrinsic Subtype	#	Recurrence Score		70-gene		Wound-Response		2-gene	
		Low	1	good	0	Quiescent	0	Low	1
		Intermediate	1						
Basal-like	7	high	5	poor	7	Activated	7	high	6
		Low	68	good	87	Quiescent	45	Low	77
		Intermediate	13						
Luminal A	121	high	40	poor	34	Activated	76	high	44
		Low	2	good	9	Quiescent	4	Low	30
		Intermediate	2						
Luminal B	55	high	51	poor	46	Activated	51	high	25
		Low	1	good	2	Quiescent	0	Low	5
		Intermediate	0						
HER2+/ER-	18	high	17	poor	16	Activated	18	high	13
		Low	15	good	15	Quiescent	11	Low	9
		Intermediate	2						
Normal-like	24	high	7	poor	9	Activated	13	high	15

Table 4.7. Two-way contingency table analysis measuring the association between the 70-gene, Wound-response, and Recurrence Score predictors in the ER+ 225-sample dataset.

A.

Two-way contingency table	**Wound-Response** (# of patients)	
	Quiescent	Activated
70-gene predictor		
Good	47	66
Poor	13	99
Statistics for two-way contingency table analysis		
p-value†	<0.001	
Cramer's V††	0.34	

B.

Two-way contingency table	**Recurrence Score** (# of patients)	
	Low or Int.	High
70-gene predictor		
Good	83	30
Poor	22	90
Statistics for two-way contingency table analysis		
p-value†	<0.001	
Cramer's V††	0.54	

C.

Two-way contingency table	**Recurrence Score**	
	Low or Int. (# of patients)	High
Wound Response		
Quiescent	47	13
Activated	58	107
Statistics for two-way contingency table analysis		
p-value†	<0.001	
Cramer's V††	0.38	

† p-value calculated from Chi-square test on contingency table. †† Cramer's V statistic (value can range from 0 to 1) measures the strength of association between the two variables analyzed in the contingency table, with 1 indicating perfect association and 0 indicating no association.

FIGURES

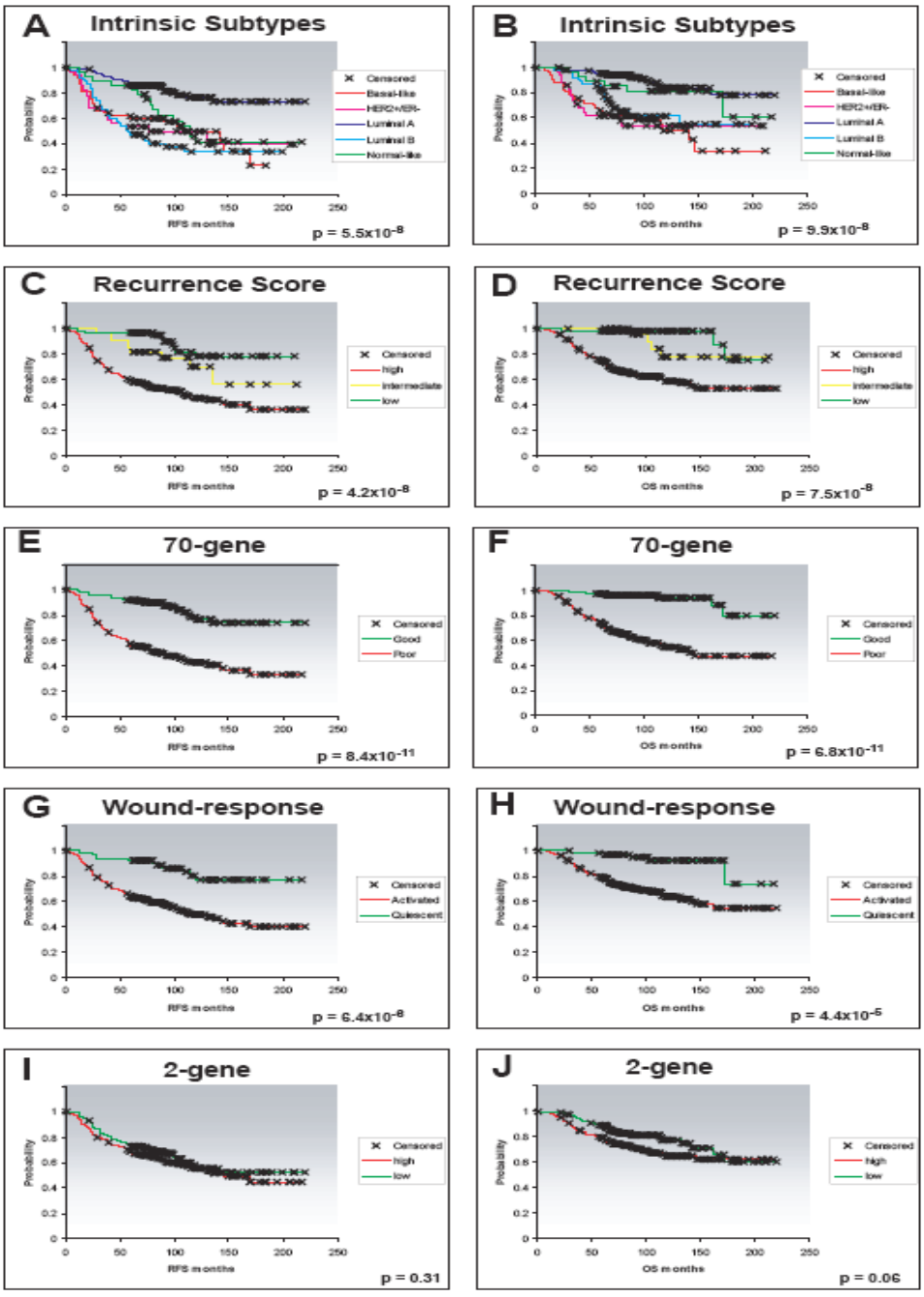


Figure 4.1. Survival analysis of the 295 patients stratified according to 5 different gene expression based predictors. Kaplan-Meier survival plots for Relapse Free Survival (left panels) and Overall Survival (right panels) are shown for the 295 patients stratified according to Intrinsic Subtypes (A, B), Recurrence Score (C, D), 70-gene profile (E, F), Wound-Response (G, H) and 2-gene ratio predictor (I, J). All reported p-values are based upon a log-rank test.

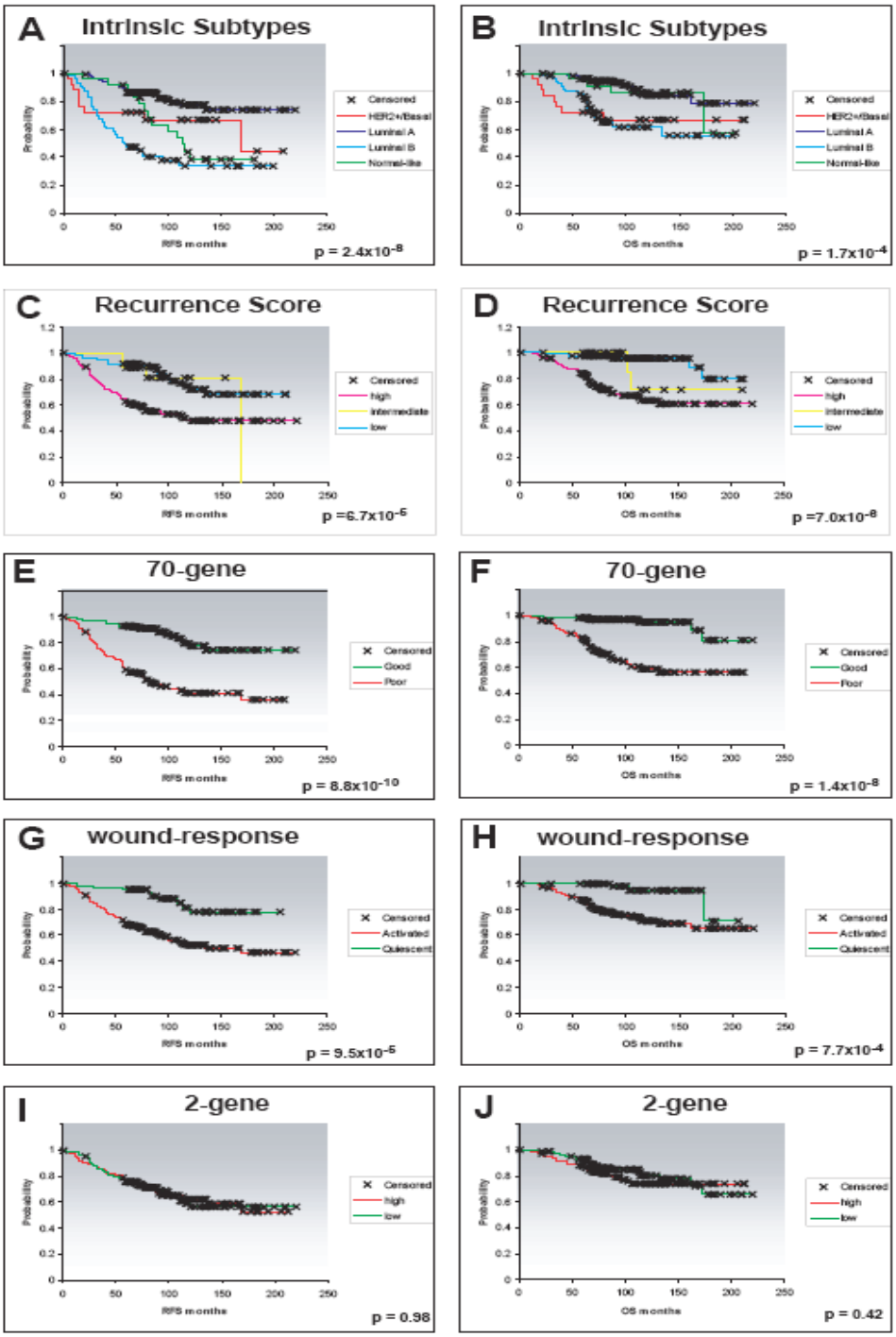


Figure 4.2. Survival analysis of the 225 ER+ patients stratified according to 5 different gene expression based predictors. Kaplan-Meier survival plots for Relapse Free Survival (left panels) and Overall Survival (right panels) are shown for the 225 ER+ patients stratified according to Intrinsic Subtypes (A, B), Recurrence Score (C, D), 70-gene profile (E, F), Wound-Response (G, H) and 2-gene ratio predictor (I, J). All reported p-values are based upon a log-rank test.

CHAPTER 5: Expression profiles can predict response to neoadjuvant chemotherapy in breast cancer patients

ABSTRACT

Background

The identification of markers predictive of neoadjuvant chemotherapy response would be an important advance. Therefore, we sought to identify gene expression profiles predictive of neoadjuvant response.

Methods

DNA microarray analysis was performed on pre-treatment core biopsies from locally advanced breast cancer patients receiving four cycles of neoadjuvant doxorubicin plus cyclophosphamide (AC) followed by four cycles of paclitaxel (T) or paclitaxel and trastuzumab (TH). In total, 44 patients receiving chemotherapy yielded successful pre-treatment core biopsy and microarray.

Results

Ten-fold cross-validated supervised analyses using the pre-treatment microarray data identified gene expression patterns that accurately predicted (1) clinical response after four cycles of treatment, (2) clinical response after successful completion of all eight

cycles of treatment, and (3) overall clinical response for these 44 patients. In contrast, no significant association was detected between any of the response outcomes measured and the standard clinical parameters of ER status, node status, or grade.

Conclusions

These results suggest that gene expression profiling may lead to clinically useful predictors of neoadjuvant chemotherapy response. The gene expression patterns reported here may provide the means of selecting patients for AC-T(H) neoadjuvant therapy.

INTRODUCTION

Neoadjuvant chemotherapy (treatment before primary surgery) has been widely used as a component of the standard of care for locally advanced breast cancer patients. Although neoadjuvant and adjuvant chemotherapy have similar efficacy in terms of disease-free and overall survival rates, neoadjuvant chemotherapy has been shown to improve breast-conserving operability in locally advanced breast cancers[1]. Another advantage of neoadjuvant chemotherapy is that it allows for the direct and timely observation of tumor treatment response. Response (pathologic complete response [pCR] or clinical complete response) to neoadjuvant chemotherapy has been correlated with improved long-term disease-free and overall survival[1-7]. Currently, there is no clinically useful predictor of neoadjuvant chemotherapy response. Such a predictor would be of significant value; by identifying patients unlikely to benefit from therapy, it would spare them from treatment-associated toxicities and allow them to be more efficiently selected to receive alternative approaches.

Recently, studies have focused on using gene expression profiling to identify expression patterns predictive of chemotherapy response[8-13]. Using gene expression profiling shows promise in identifying patterns predictive of chemotherapy response as it has already allowed the classification of breast tumors into five molecular subtypes (Luminal A, Luminal B, Basal-like, HER2+/ER- and Normal Breast-like) that show significant differences in patient outcome[14,15]. Therefore, we determined if pre-treatment gene expression patterns could predict response to the following neoadjuvant chemotherapy regimen: doxorubicin plus cyclophosphamide followed by paclitaxel with or without trastuzumab.

MATERIALS AND METHODS

Patients and study design. A phase II study run at UNC-CH (Trial L9819; Trial PI: Lisa Carey) of the neoadjuvant administration of four cycles of doxorubicin plus cyclophosphamide (AC), followed by four cycles of paclitaxel (T) or paclitaxel plus trastuzumab (TH) was performed on a set of locally advanced operable breast cancer patients (for complete details see Carey et al.[16]). Patients received AC-TH if their tumors showed HER2-positivity as defined by Carey et al.[16] The main purpose of this phase II study was to determine the cardiotoxicity of neoadjuvant doxorubicin plus cyclophosphamide followed by paclitaxel plus trastuzumab. As part of this study, pretreatment core biopsies were obtained from patients; RNA from these biopsies was used for our prediction of response analysis described here. Clinical response to neoadjuvant therapy was evaluated after the first four cycles of treatment and after all eight cycles of treatment, following Response Evaluation Criteria in Solid Tumors (RECIST), with complete response (CR) defined as no clinical evidence of tumor, partial response (PR) as ≥30% decrease in the longest diameter, progression (PD) as ≥20% increase in the longest diameter, and stable disease (SD) as all other tumor responses[17]. Pathologic response in the post-neoadjuvant chemotherapy surgical specimen was defined by residual disease in the breast or axillary lymph nodes according to the revised 2003 AJCC TNM staging system.

RNA isolation and microarray hybridization. Total RNA samples from the pretreatment biopsies were prepared using Qiagen RNAeasy kits. An Agilent Bioanalyzer was used to determine sample quality. Only those samples giving ≥1μg Total RNA and

discernable 18S and 28S peaks were used for microarray analysis. Total RNA amplification and labeling were done as previously described[18]. Microarray hybridizations were performed on Agilent Human microarrays using 2μg of Cy3-labeled common reference sample[19] and 2μg of Cy5-labeled experimental sample. Microarrays were hybridized overnight, washed, dried, and scanned as described[18]. Microarray image files were analyzed with GenePix Pro 4.1 and loaded into the UNC-CH Microarray Database (https://genome.unc.edu/).

Microarray analysis and prediction of response. Data from microarray experiments were calculated as described[18]. Genes were excluded from data analysis if they did not have signal intensity ≥30 in both channels for ≥70% of the experiments. To predict response, the gene expression data for the 44 pre-treatment samples was used and the "supervising parameters" were clinical response after cycles 4 and 8, overall clinical response, and pathologic complete response (pCR). The difference between overall clinical response and clinical response after cycle 8 is that for a patient to be evaluated for the latter, she would had to have completed all 8 therapy cycles. In contrast, overall clinical response is evaluated after the last successfully completed therapy cycle (not necessarily cycle 8).

Four statistical classification methods were used to predict chemotherapy response using the pre-treatment gene expression data: a k-Nearest Neighbor Classifier (k-NN with k=1, 3, 5, or 7) with either Euclidean distance or one-minus-Spearman-correlation as the distance function and a Class Nearest Centroid (CNC) classifier with either Euclidean distance or one-minus-Spearman-correlation as the distance function[20].

To evaluate prediction accuracy, each of the four prediction methods underwent 10-fold cross validation (CV); in a given round of CV, each predictor using *n* genes (how the *n* genes were selected is described below) was trained on 90% of the samples and used to make predictions on the remaining 10%, with this procedure repeated 9 more times such that every sample was "left out" exactly once. The prediction accuracies for each of the 10 iterations were averaged together and this average prediction accuracy was recorded for each prediction method with *n* genes. *n* was increased for subsequent rounds of CV. For each response variable, the set of *n* genes that gave the highest average prediction accuracy during CV was determined and reported for each prediction method (Table 3).

Each prediction method required a gene/feature selection step to identify genes associated with each "class" (i.e. CR vs. PR+SD). For all 4 prediction methods, we used a gene selection method first described by Dudoit *et al.*[21]; the genes were identified in the training set according to the ratio of between-class to within-class sums of squares. The top *n*-ranked genes were used during each round of CV. The number of cases in our study was relatively small (44), therefore, we did not break our data into training and test sets but instead, performed 10-fold CV using the four statistical prediction methods to avoid over-fitting caused by using a single prediction method or fortuitous training and test set randomizations.

RESULTS

Patient characteristics and response rates. 44 patients enrolled in the L9819 study gave a successful pre-treatment core biopsy and microarray. Patient characteristics are summarized in Table 1. Of these 44 patients, 24 and 11 successfully completed AC-T and

AC-TH therapy, respectively. The patient subgroups receiving AC-T or AC-TH did not significantly differ from each other in age, grade, stage, ER or PR status. Clinical response data for the 44 patients is summarized in Table 2. Patient subgroups receiving AC-T or AC-TH did not significantly differ from each other in response rates (overall clinical response, clinical response after 4 and 8 treatment cycles, and pCR).

Analysis of tumor samples using the breast intrinsic gene set. Chemotherapy response is likely a multi-factorial process, therefore, we examined whether gene expression patterns were able to capture response-associated biological features. To investigate the gene expression data, we first hierarchically clustered[22] the 44 pre-treatment samples using the 1300-gene "breast intrinsic" gene set (developed by Hu *et al.*[23]) that identifies the intrinsic breast tumor subtypes (Luminal, HER2+/ER-neg, Basal-like). The results show that the main intrinsic subtypes (Figure 1) were identifiable in this patient dataset. As seen in previous studies[23], the proliferation gene cluster was found to have the highest expression in Basal-like tumors.

Association of response with clinical parameters and breast intrinsic subtype. We explored how conventional clinical parameters performed in predicting response. Using either (1) all 44 patients, (2) just those receiving AC-T, or (3) just those receiving AC-TH, neither ER status, PR status, node status, grade, or tumor size were significantly correlated with pCR or any other response variable according to Chi-squared analysis. Other studies have also found that the standard clinical parameters show weak to no association with neoadjuvant chemotherapy response[24].

We next examined response rates within the intrinsic molecular subtypes. Rouzier *et al.*[25] reported a strong association (p=0.002) between pCR rate and the Basal-like and HER2+/ER- subtypes. In the L9819 study, using either (1) all 44 patients, (2) just those receiving AC-T, or (3) just those receiving AC-TH, we did not see a statistically significant association between subtype and pCR or any other response variable (overall clinical response, clinical response after 4 and 8 treatment cycles). Among all 44 patients, 2/11 (18.2%) basal-like, 2/15 (13.3%) HER2+/ER-, and 2/16 (12.5%) luminal tumors showed pCR. Among all 44 patients, the association between subtype and clinical response after 4 treatment cycles was not significant (p=0.17), but a trend was evident: 10/12 (83.3%) basal-like, 8/16 (50%) HER2+/ER-, and 9/16 (56.3%) luminal tumors showed response (complete or partial) after 4 treatment cycles, which mimics the finding of Rouzier *et al.*[25] in that the highest response rates were seen in the Basal-like subtype.

Prediction of neoadjuvant chemotherapy response. We performed "supervised analyses" on the pretreatment gene expression data and determined the 10-fold Cross Validation (CV) error rates for predicting (1) pCR, (2) overall response, (3) response after 4 treatment cycles, and (4) response after 8 treatment cycles. Table 3 shows that 10-fold CV analyses using the Class Nearest Centroid and k-Nearest Neighbor classification methods yielded gene expression profiles/predictors that accurately classified tumors according to (1) overall response: clinical complete response (CR) vs. non-CR (75-77% accuracy), (2) clinical response (partial or complete) vs. non-response after cycle 4 (73-77% accuracy), and (3) CR vs. non-CR after cycle 8 (79-83% accuracy). We could not accurately classify tumors according to pCR (<60% accuracy), which we speculate is due

to the fact that the low number of samples showing pCR (n=6) was not enough to generate a reliable predictor. Each of the four prediction methods achieved similar accuracies when used to predict a given response variable (Table 3). Prediction of ER status is included as a positive control for our gene expression-based predictors; it represents the upper threshold of how good a predictor can be (86-89% accuracy) on this dataset, providing a benchmark against which the chemotherapy response predictors can be judged.

Using the gene lists identified in 10-fold CV as being predictive of response (Table 3, highlighted in blue), we hierarchically clustered[22] the pretreatment biopsy samples to better understand the predictive genes and their relationships to each other. Figure 2 shows the hierarchical clustering of tumors using the 54-gene set predictive of response after treatment cycle 4 (75% accuracy, 78% sensitivity, 71% specificity in 10-fold CV analysis using the Euclidean nearest centroid method) (Note: the classification of samples into clusters and the associated accuracies observed in Figures 2-4 are different from those observed in the 10-fold CV analysis shown in Table 3. The clusters are for illustrative purposes/better understanding of the predictive gene sets only). Using the program EASE[26], the Gene Ontology (GO) categories "DNA binding" and "nucleotide/nucleic acid metabolism" were over-represented relative to chance in the gene set highly expressed in tumors showing response at cycle 4 (Figure 2, top gene dendrogram branch).

Figure 3 shows the hierarchical clustering of tumors using the 70-gene set predictive of overall response (75% accuracy, 72% sensitivity, 77% specificity in 10-fold CV analysis using the Euclidean nearest centroid method). Interestingly, the complete

responders in Figure 3 showed high expression of caspase-9, which promotes apoptosis[27-29]. According to EASE, the GO category "ATP-dependent helicase activity" was over-represented relative to chance in the gene set highly expressed in tumors showing overall complete response.

Figure 4 shows the hierarchical clustering of tumors using the 72-gene set predictive of response after 8 treatment cycles (79% accuracy, 80% sensitivity, 79% specificity in 10-fold CV analysis using the Euclidean nearest centroid method). EASE showed that the GO categories "apoptosis/programmed cell death" and "positive regulation of apoptosis" were over-represented relative to chance in the gene set highly expressed in tumors showing complete response at cycle 8. Some of the pro-apoptotic genes highly expressed in complete responders relative to non-complete responders included caspases 4 and 5 and PACAP. These results suggest that in the pre-treatment samples, the high expression of proapoptotic genes are associated with and may partly explain chemotherapy response.

DISCUSSION

In this work, we examined whether pre-treatment gene expression patterns could predict response to AC-T(H) neoadjuvant therapy. 10-fold CV analysis identified gene expression patterns with prediction accuracy rates of 75-77% for overall response, 73-77% for response after cycle 4, and 79-83% for response after cycle 8 (Table 3). The accuracy rates achieved are encouraging and warrant further validation.

We speculate that the 72-gene set predictive of response after 8 treatment cycles may represent a general mechanism of chemotherapy response. This gene set was

significantly enriched for pro-apoptotic genes. Findings from other studies indicate that apoptosis genes may be important in a general chemotherapy response; several reports studying breast tumor response to different chemotherapy regimens all indicate that in pretreatment samples, the high expression of apoptosis-related genes is associated with response[9,10,30-32]. Currently, we are testing the hypothesis that our 72-gene set may represent a general predictor of chemotherapy response using additional data sets of neoadjuvantly treated patients.

To date, three other published studies have attempted to predict neoadjuvant chemotherapy response using gene expression[8,10,11]. The prediction accuracies achieved by our gene expression-based predictors are similar to those achieved by the predictors developed by Chang *et al.*[10] and Ayers *et al.*[8]: their predictors showed accuracies of 88% for predicting docetaxel clinical response and 78% for predicting pCR to paclitaxel + fluorouracil + doxorubicin + cyclophosphamide, respectively. In contrast, Hannemann *et al.*[11] could not find a gene expression pattern from pretreatment FNAC samples that was capable of predicting pCR for patients receiving doxorubicin and cyclophosphamide or doxorubicin and docetaxel. Clearly, additional studies are needed with larger sample sizes, however, some common themes in the predictive gene expression patterns are evident, including apoptosis as an important feature. It is encouraging that genes identified in the predictors make biological sense and suggest our predictors are appropriately tracking response.

Our results indicate that gene expression profiling may lead to clinically useful predictors of neoadjuvant chemotherapy response. Information regarding the intrinsic subtypes may also contribute to prediction assay development. We believe that the gene

expression-based predictors identified here have the potential to be clinically useful and warrant further validation using additional datasets as they emerge.

REFERENCES

1. Kaufmann M, von Minckwitz G, Smith R, et al: International expert panel on the use of primary (preoperative) systemic treatment of operable breast cancer: review and recommendations. J Clin Oncol 21:2600-8, 2003

2. Kuerer HM, Newman LA, Smith TL, et al: Clinical course of breast cancer patients with complete pathologic primary tumor and axillary lymph node response to doxorubicin-based neoadjuvant chemotherapy. J Clin Oncol 17:460-9, 1999

3. Fisher B, Bryant J, Wolmark N, et al: Effect of preoperative chemotherapy on the outcome of women with operable breast cancer. J Clin Oncol 16:2672-85, 1998

4. Ferriere JP, Assier I, Cure H, et al: Primary chemotherapy in breast cancer: correlation between tumor response and patient outcome. Am J Clin Oncol 21:117-20, 1998

5. Chang J, Powles TJ, Allred DC, et al: Biologic markers as predictors of clinical outcome from systemic therapy for primary operable breast cancer. J Clin Oncol 17:3058-63, 1999

6. Pierga JY, Mouret E, Laurence V, et al: Prognostic factors for survival after neoadjuvant chemotherapy in operable breast cancer. the role of clinical response. Eur J Cancer 39:1089-96, 2003

7. Cleator SJ, Makris A, Ashley SE, et al: Good clinical response of breast cancers to neoadjuvant chemoendocrine therapy is associated with improved overall survival. Ann Oncol 16:267-72, 2005

8. Ayers M, Symmans WF, Stec J, et al: Gene expression profiles predict complete pathologic response to neoadjuvant paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide chemotherapy in breast cancer. J Clin Oncol 22:2284-93, 2004

9. Bertucci F, Finetti P, Rougemont J, et al: Gene expression profiling for molecular characterization of inflammatory breast cancer and prediction of response to chemotherapy. Cancer Res 64:8558-65, 2004

10. Chang JC, Wooten EC, Tsimelzon A, et al: Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer. Lancet 362:362-9, 2003

11. Hannemann J, Oosterkamp HM, Bosch CA, et al: Changes in gene expression associated with response to neoadjuvant chemotherapy in breast cancer. J Clin Oncol 23:3331-42, 2005

12. Iwao-Koizumi K, Matoba R, Ueno N, et al: Prediction of docetaxel response in human breast cancer by gene expression profiling. J Clin Oncol 23:422-31, 2005

13. Sorlie T, Perou CM, Fan C, et al: Differential sensitivity to doxorubicin and mitomycin among breast cancers belonging to the intrinsic subtypes identified by gene expression profiling. Submitted, 2005

14. Sorlie T, Perou CM, Tibshirani R, et al: Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 98:10869-74, 2001

15. Sorlie T, Tibshirani R, Parker J, et al: Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A 100:8418-23, 2003

16. Carey LA, Hanna K, Kimmick G, et al: Clinical and pathologic response rate and acute and chronic cardiotoxicity of neoadjuvant trastuzumab given in conjunction with weekly paclitaxel after AC. Submitted, 2006

17. Therasse P, Arbuck SG, Eisenhauer EA, et al: New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. J Natl Cancer Inst 92:205-16, 2000

18. Hu Z, Troester M, Perou CM: High reproducibility using sodium hydroxide-stripped long oligonucleotide DNA microarrays. Biotechniques 38:121-4, 2005

19. Novoradovskaya N, Whitfield ML, Basehore LS, et al: Universal Reference RNA as a standard for microarray experiments. BMC Genomics 5:20, 2004

20. Chung CH, Parker JS, Karaca G, et al: Molecular classification of head and neck squamous cell carcinomas using patterns of gene expression. Cancer Cell 5:489-500, 2004

21. Dudoit S, Fridlyand J, Speed TP: Comparison of discrimination methods for the classification of tumors using gene expression data. Journal of the American Statistical Association 97:77-87, 2002

22. Eisen MB, Spellman PT, Brown PO, et al: Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 95:14863-8, 1998

23. Hu Z, Fan C, Oh DS, et al: The Molecular Portraits of Breast Tumors Are Conserved Across Microarray Platforms. Submitted, 2006

24. Hortobagyi GN, Hayes DF, Pusztai L: Integrating newer science into breast cancer prognosis and treatment: a review of current molecular predictors and profiles. American Society of Clinical Oncology, Annual Meeting:191-202, 2002

25. Rouzier R, Perou CM, Symmans WF, et al: Breast cancer molecular subtypes respond differently to preoperative chemotherapy. Clin Cancer Res 11:5678-85, 2005

26. Hosack DA, Dennis G, Jr., Sherman BT, et al: Identifying biological themes within lists of genes with EASE. Genome Biol 4:R70, 2003

27. Shinoura N, Sakurai S, Asai A, et al: Over-expression of APAF-1 and caspase-9 augments radiation-induced apoptosis in U-373MG glioma cells. Int J Cancer 93:252-61, 2001

28. Cardone MH, Roy N, Stennicke HR, et al: Regulation of cell death protease caspase-9 by phosphorylation. Science 282:1318-21, 1998

29. Hakem R, Hakem A, Duncan GS, et al: Differential requirement for caspase 9 in apoptotic pathways in vivo. Cell 94:339-52, 1998

30. Folgueira MA, Carraro DM, Brentani H, et al: Gene expression profile associated with response to doxorubicin-based therapy in breast cancer. Clin Cancer Res 11:7434-43, 2005

31. Dressman HK, Hans C, Bild A, et al: Gene expression profiles of multiple breast cancer phenotypes and response to neoadjuvant chemotherapy. Clin Cancer Res 12:819-26, 2006

32. Thuerigen O, Schneeweiss A, Toedt G, et al: Gene expression signature predicting pathologic complete response with gemcitabine, epirubicin, and docetaxel in primary breast cancer. J Clin Oncol 24:1839-45, 2006

TABLES

Table 5.1. Patient characteristics of the L9819 dataset.

	L9819 dataset (n=44) No. of pts (%)
Age	
Median (range)	47 (30 to 79)
Stage	
I	0 (0%)
IIA	5 (11.4%)
IIB	10 (22.7%)
IIIA	18 (40.9%)
IIIB	7 (15.9%)
IIIC	1 (2.3%)
IV	3 (6.8%)
ER status	
+	21 (48%)
-	23 (52%)
n.a.	-
Grade	
1	5 (11.4%)
2	7 (15.9%)
3	25 (56.8%)
n.a.	7 (15.9%)

Table 5.2. Neoadjuvant Chemotherapy Response data for the L9819 dataset.

	Response after 4 cycles	Response after 8 cycles	Overall response	pCR
Response data for all 44 pts.	8 CR (18.2%) 19 PR (43.2%) 15 SD (34.1%) 2 PD (4.5%)	10 CR (34.5%) 12 PR (41.4%) 6 SD (20.7%) 1 PD (3.4%)	18 CR (40.9%) 16 PR (36.4%) 9 SD (20.5%) 1 PD (8.3%)	6 pCR (13.6%) 36 non-pCR (81.8%) 2 n.a. (4.5%)

Abbreviations: pCR=pathologic complete response; CR=clinical complete response; PR=partial response; SD=stable disease; PD=progressive disease.

Table 5.3. Accuracy (as determined by 10-fold cross validation) of various classification methods for the 44 tumor samples that yielded pre-treatment microarray data from the L9819 study. Classification methods were used to predict (A) response at cycle 4, (B) response after completion of all 8 treatment cycles, (C) overall response, and (D) ER status.

A. Prediction of Response at Cycle 4 (27 responders [8 CRs and 19 PRs] vs. 17 non-responders [15 SDs and 2 PDs])

Classification method	**gene #**	**acc.**	**True +**	**True -**	**False +**	**False -**	**sens.**	**spec.**	**PPV**	**NPV**
Spearman Nearest Centroid	53	0.727	20	12	5	7	0.741	0.706	0.800	0.632
Euclidean Nearest Centroid	54	0.75	21	12	5	6	0.778	0.706	0.808	0.667
Spearman k-NN (k=7)	37	0.727	21	11	6	6	0.778	0.647	0.778	0.647
Euclidean k-NN (k=7)	53	0.773	21	13	4	6	0.778	0.765	0.840	0.684
Average		0.744					0.769	0.706	0.806	0.657

B. Prediction of Response at Cycle 8 (10 CRs vs. 19 non-CRs [12 PRs, 6 SDs, and 1 PD])

Classification method	**gene #**	**acc.**	**True +**	**True -**	**False +**	**False -**	**sens.**	**spec.**	**PPV**	**NPV**
Spearman Nearest Centroid	55	0.793	8	15	4	2	0.800	0.789	0.667	0.882
Euclidean Nearest Centroid	72	0.793	8	15	4	2	0.800	0.789	0.667	0.882
Spearman k-NN (k=5)	76	0.793	8	15	4	2	0.800	0.789	0.667	0.882
Euclidean k-NN (k=5)	63	0.828	8	16	3	2	0.800	0.842	0.727	0.889
Average		0.802					0.800	0.803	0.682	0.884

C. Prediction of Overall Response (18 CRs vs. 26 non-CRs [16 PRs, 9 SDs, and 1 PD])

Classification method	**gene #**	**acc.**	**True +**	**True -**	**False +**	**False -**	**sens.**	**spec.**	**PPV**	**NPV**
Spearman Nearest Centroid	54	0.75	14	19	7	4	0.778	0.731	0.667	0.826
Euclidean Nearest Centroid	70	0.75	13	20	6	5	0.722	0.769	0.684	0.800
Spearman k-NN (k=5)	55	0.773	14	20	6	4	0.778	0.769	0.700	0.833
Euclidean k-NN (k=5)	37	0.75	12	21	5	6	0.667	0.808	0.706	0.778
Average		0.756					0.736	0.769	0.689	0.809

D. Prediction of ER status (21 ER+ vs. 23 ER-)

Classification method	**gene #**	**acc.**	**True +**	**True -**	**False +**	**False -**	**sens.**	**spec.**	**PPV**	**NPV**
Spearman Nearest Centroid	50	0.864	19	19	4	2	0.905	0.826	0.826	0.905
Euclidean Nearest Centroid	50	0.864	19	19	4	2	0.905	0.826	0.826	0.905
Spearman k-NN (k=3)	50	0.886	19	20	3	2	0.905	0.870	0.864	0.909
Euclidean k-NN (k=3)	50	0.864	19	19	4	2	0.905	0.826	0.826	0.905
Average		0.870					0.905	0.837	0.835	0.906

Note: Accuracies, etc. highlighted in blue are for the gene lists used to cluster tumors in Figures 2-4. Average accuracies, etc. across the prediction methods are highlighted in red. "True +" refers to number of responders correctly identified. "True –" refers to number of non-responders correctly identified. "False +" refers to number of non-responders incorrectly identified. "False –" refers to number of responders incorrectly identified.

Abbreviations: acc.=accuracy; sens.=sensitivity; spec.=specificity; PPV=positive predictive value; NPV=negative predictive value; k-NN=k-nearest neighbors; CR=clinical complete response; PR=partial response; SD=stable disease; PD=progressive disease.

FIGURES

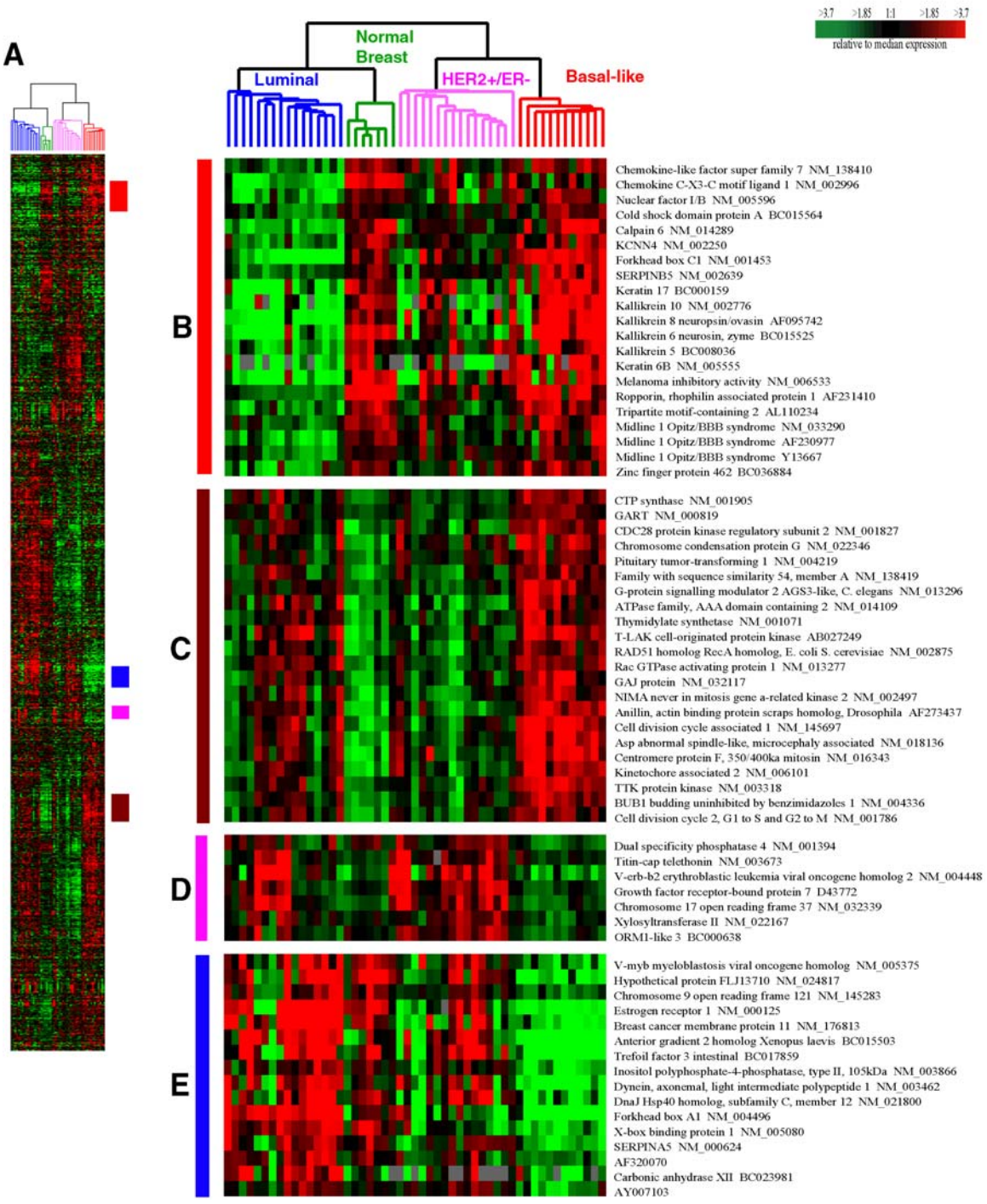


Figure 5.1. Hierarchical cluster analysis of L9819 pre-treatment tumor samples using the 1300-gene "intrinsic breast" gene set developed by Hu et al. groups tumors into the intrinsic subtypes. **A.** Scaled-down representation of the complete cluster diagram. **B.** Basal epithelial, **C.** Proliferation, **D.** HER2+, and **E.** Luminal gene clusters.

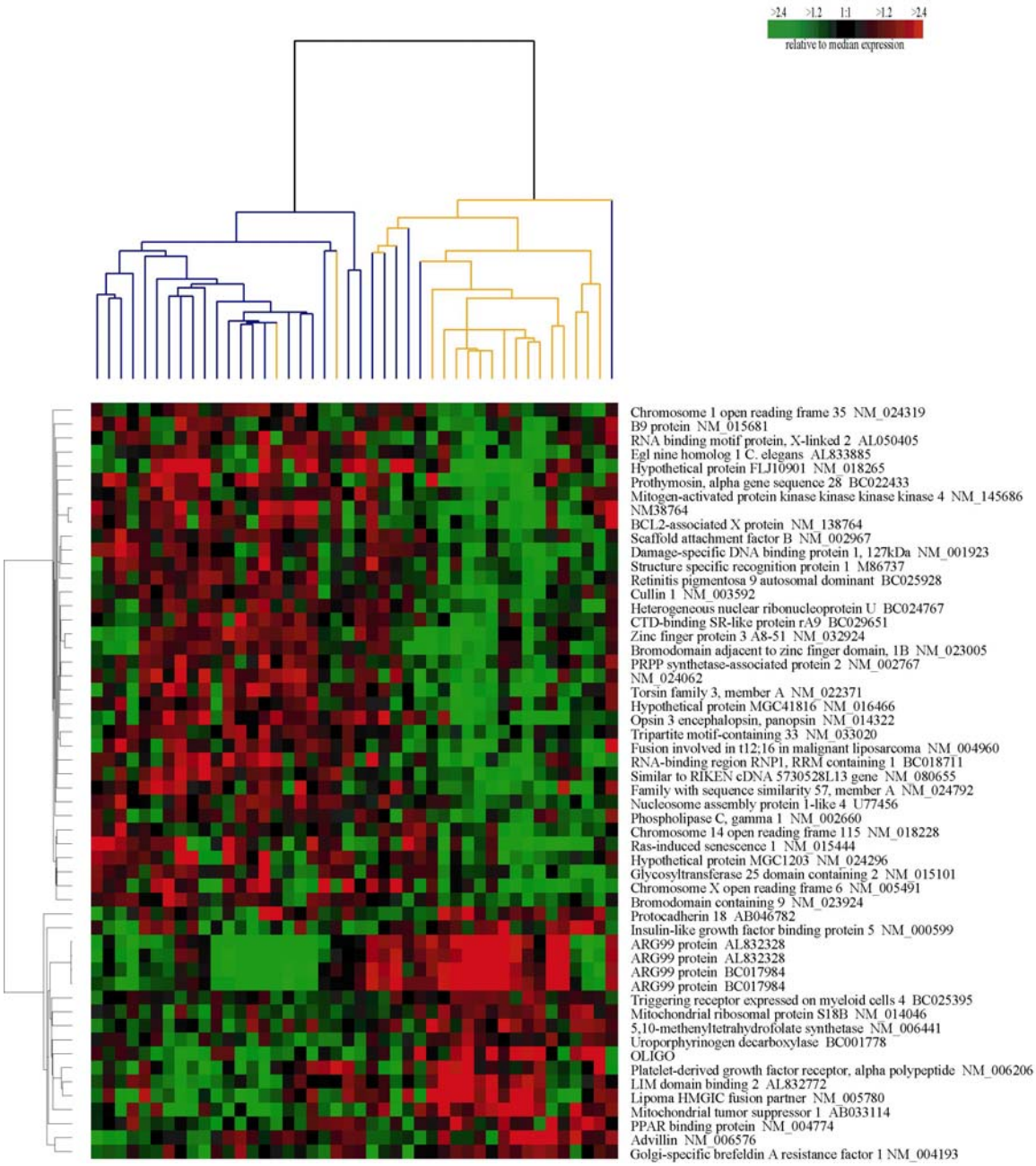


Figure 5.2. Hierarchical cluster analysis of L9819 pre-treatment tumor samples using the 54-gene set predictive of clinical response after 4 treatment cycles. Blue and yellow dendrogram branches indicate responders (complete or partial) and non-responders (stable or progressive disease), respectively.

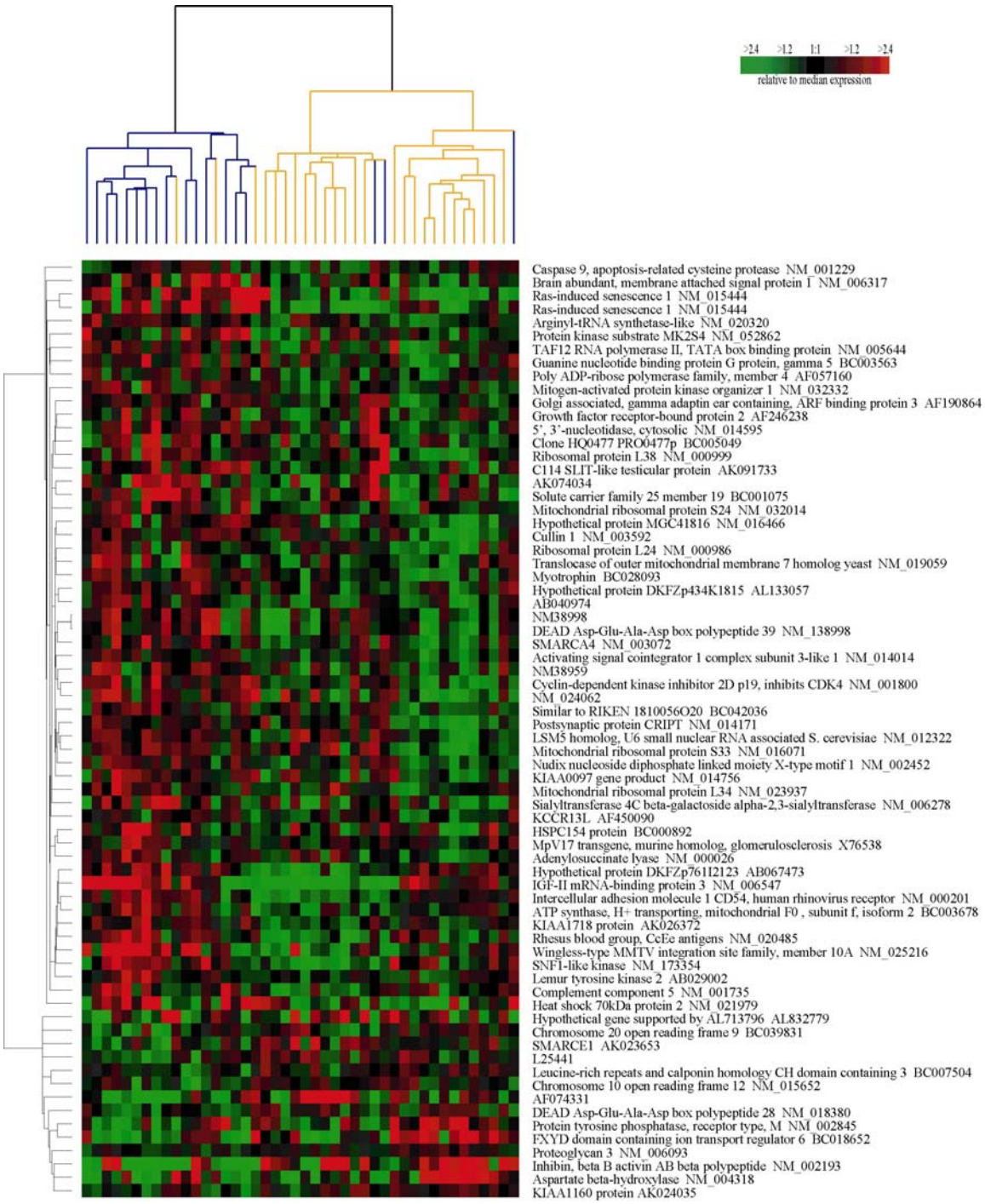


Figure 5.3. Hierarchical cluster analysis of L9819 pre-treatment tumor samples using the 70-gene set predictive of overall response. Blue and yellow dendrogram branches indicate complete and non-complete overall responders, respectively.

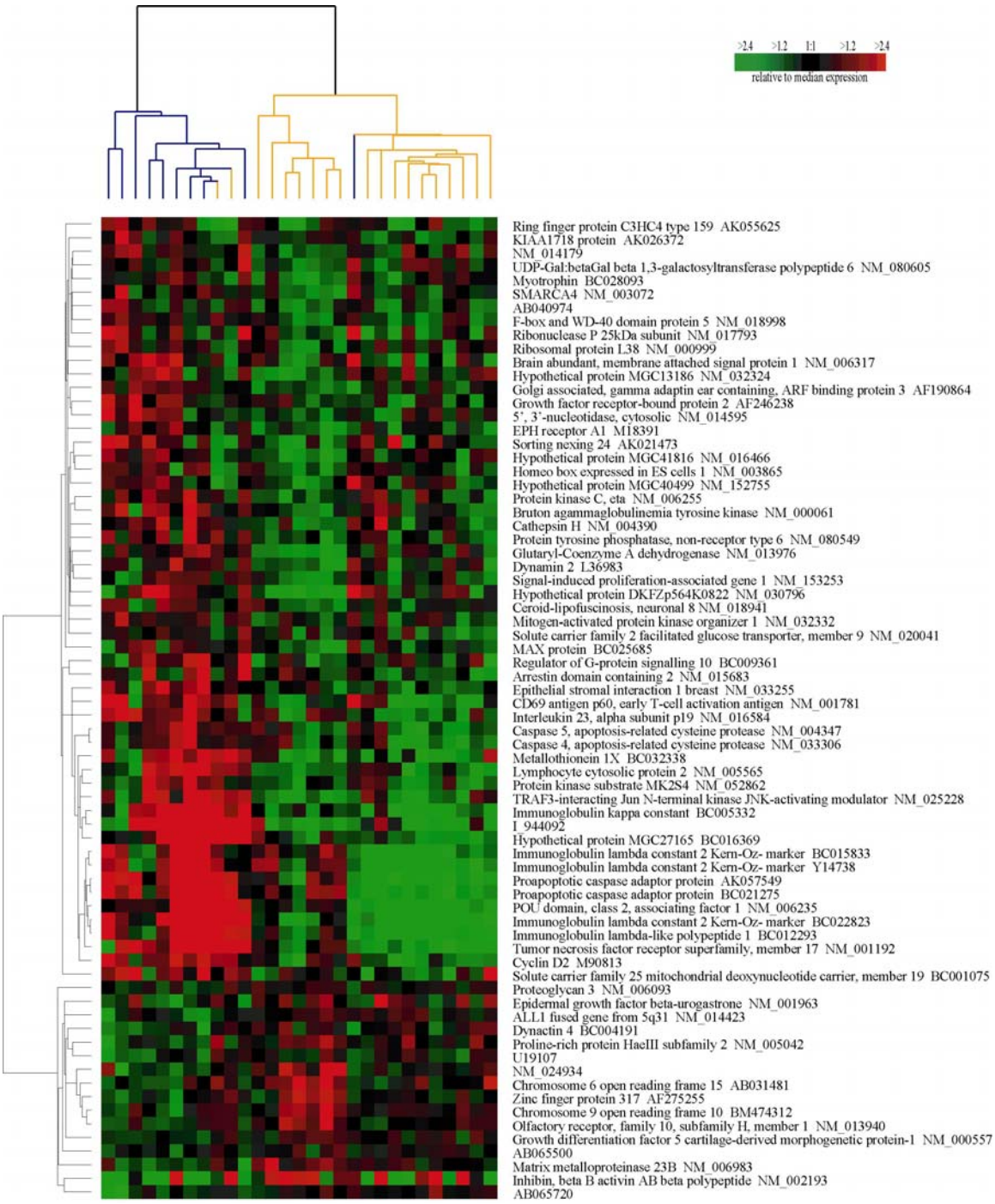


Figure 5.4. Hierarchical cluster analysis of L9819 pre-treatment tumor samples using the 72-gene set predictive of clinical response after completion of all 8 treatment cycles. Blue and yellow dendrogram branches indicate complete and non-complete responders, respectively.

CHAPTER 6: DISCUSSION

Currently, many breast cancer patients are being given lengthy and expensive treatments associated with significant toxicity and morbidity. Some of these patients will receive no benefit in survival while others would have achieved good outcomes without additional treatments. This work has aimed to remedy this situation using gene expression profiling. First, we developed an expression-based predictor of outcome for Estrogen Receptor (ER) and/or Progesterone Receptor (PR)-positive breast cancer patients using biological differences among these tumors. Second, we used a recently developed multivariate analysis tool (DWD) to validate and objectively define the "intrinsic" subtypes as a predictor/prognosticator of breast cancer patient outcomes by using independent datasets generated on differing microarray platforms. Third, using a single patient dataset, we determined that there was significant concordance in outcome predictions made by several different gene expression profiles (developed on different platforms by different laboratories), which showed little overlap in gene identity. Lastly, we developed gene expression-based predictors for response to neoadjuvant chemotherapy using pre-treatment microarray data.

From the results of this work, we propose the following decision tree for how to treat breast cancer patients (Figure 1), which will of course, need much more validation and testing before routine clinical use. First, using the intrinsic subtype single sample predictor (SSP) introduced in Chapter 3, the patient would be objectively classified as

having either luminal, HER2+/ER-, or basal-like breast cancer. If the patient's tumor were determined to be of the luminal subtype, we would then apply the Group IE-IIE predictor introduced in Chapter 2. If the tumor is classified as Group IE, we hypothesize that all that is needed for treatment would be hormone therapy (i.e. tamoxifen), and that this patient might be spared the chemotherapy regimen that they would normally be prescribed. If the tumor is classified as Group IIE, we would then first examine the tumor's HER2 status to determine whether trastuzumab (monoclonal antibody that binds to HER2) should be part of the treatment regimen (i.e. HER2+ tumors will receive trastuzumab). Regardless of HER2 status, if the tumor is classified as Group IIE, we would also apply the L9819 72-gene chemotherapy response predictor (introduced in Chapter 5), which we believe can be used to predict general chemotherapy response. If the tumor is predicted to respond to chemotherapy by the L9819 predictor, we would include (neo)adjuvant chemotherapy and tamoxifen in the treatment regimen for these patients (however, the data suggests that Group IIE tumors are resistant to tamoxifen, which must be formally tested in randomized trials). If the tumor is predicted to not respond to chemotherapy by the L9819 predictor, the patient would be given an alternative therapy and perhaps be considered for entry into trials containing new biological agents. The hope is that these patients unlikely to benefit from conventional chemotherapy regimens will in addition to avoiding unnecessary chemotherapy-associated morbidity, have the opportunity to benefit from exposure to potentially effective novel agents.

If the patient's tumor were determined to be of the HER2+/ER- subtype, the patient will receive trastuzumab (currently some HER2+/ER- tumors are believed to be

resistant to trastuzumab, but no predictor to determine resistance yet exists). In addition, we would use the L9819 general chemotherapy response predictor to determine if conventional (neo)adjuvant chemotherapy should be added to the treatment regimen in the case of those predicted to respond, and for those predicted to not respond, alternative regimens with novel agents might be considered. Finally, if the patient's tumor were determined to be of the Basal-like subtype, at present we believe the best course of action would be to use the L9819 general chemotherapy response predictor to determine if (neo)adjuvant chemotherapy should be given to the patient, and again, those predicted not to respond might be offered therapies containing novel biological agents like HER1 inhibitors. We hope that with improved understanding of the basal-like subtype, novel therapies tailored to this particular subtype of breast cancer will be developed to complement or perhaps supplant chemotherapy as the treatment of choice. For example, recent studies suggest that this subtype may benefit from therapy targeting the epidermal growth factor receptor (HER1) and/or the pathway it regulates[1,2]. Clinical trials are currently underway to determine the efficacy of HER1 inhibitors (i.e. cetuximab, gefitinib, and erlotinib) in basal-like tumors[3].

As stated earlier, the decision tree in Figure 1 will require rigorous validation and testing before routine clinical use. These validation studies should be done using cohorts with large enough sample size so that (1) the full diversity of the target population will be represented and (2) the study will have the necessary statistical power to determine if the predictor in question does indeed improve current treatment decision-making strategies. When possible, randomized controlled trials should be used in the validation process[4]. For example, to test our hypothesis discussed earlier that Group IE tumors may benefit

from tamoxifen but not benefit significantly from chemotherapy, we could use a clinical trial in which patients classified as Group IE are randomly assigned to receive either tamoxifen or tamoxifen and chemotherapy. The survival outcomes of those receiving tamoxifen alone or tamoxifen with chemotherapy would be compared. To test our hypothesis that Group IIE tumors do not benefit from tamoxifen but may benefit significantly from chemotherapy, we could use a clinical trial in which patients classified as Group IIE are randomly assigned to receive either tamoxifen alone, chemotherapy alone, or both tamoxifen and chemotherapy. The survival outcomes of these three treatment groups would then be compared, and our expected result would be that Group IIE patients receiving tamoxifen plus chemotherapy would have the same outcomes as those receiving chemotherapy alone. Other clinical trial designs as described by Sargent et al.[4] could also be used to test these same hypotheses. When randomized clinical trials are not possible, retrospective case-control studies may be considered.

We acknowledge that before gene expression-based predictors or prognosticators can be introduced to the clinic, the current expenses and required training involved need to be made less prohibitory for routine clinical use. The requirement for fresh frozen material for microarray analysis is also prohibitory. However, we are confident that these difficulties can be overcome in time with technological advances. For example, recent progress has been made to use formalin fixation and paraffin-embedded (FFPE) tissue samples to amplify RNA for microarray analysis[5]. In addition, RT-PCR (real-time reverse transcriptase polymerase chain reaction) assays can be used as an alternative to microarrays for the simultaneous analysis of hundreds of genes and can be employed to validate “expression signatures” initially identified in microarray analysis. Importantly,

RT-PCR assays can examine gene expression using limited amounts of RNA extracted from FFPE sections.

In summary, this work has advanced the current knowledge of the heterogeneity of breast cancer and provides a means for improved prediction and prognostication for breast cancer patients. This work has shown that gene expression profiling can and will be clinically useful and will improve the treatment decision-making process for breast cancer patients. Provided that properly designed and rigorous validation studies are performed, we are confident gene expression profiling will be accepted for routine clinical use.

REFERENCES

1. Sartor C, Zhou H, Perou CM, et al: Basal-like breast tumor-derived cell lines are growth inhibited and radiosensitized by epidermal growth factor receptor (EGFR) tyrosine kinase inhibitors., San Antonio Breast Cancer Symposium. San Antonio, TX, 2004

2. Nielsen TO, Hsu FD, Jensen K, et al: Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. Clin Cancer Res 10:5367-74, 2004

3. Brenton JD, Carey LA, Ahmed AA, et al: Molecular classification and molecular forecasting of breast cancer: ready for clinical application? J Clin Oncol 23:7350-60, 2005

4. Sargent DJ, Conley BA, Allegra C, et al: Clinical trial designs for predictive marker validation in cancer treatment trials. J Clin Oncol 23:2020-7, 2005

5. Murphy N, Millar E, Lee CS: Gene expression profiling in breast cancer: towards individualising patient management. Pathology 37:271-7, 2005

FIGURES

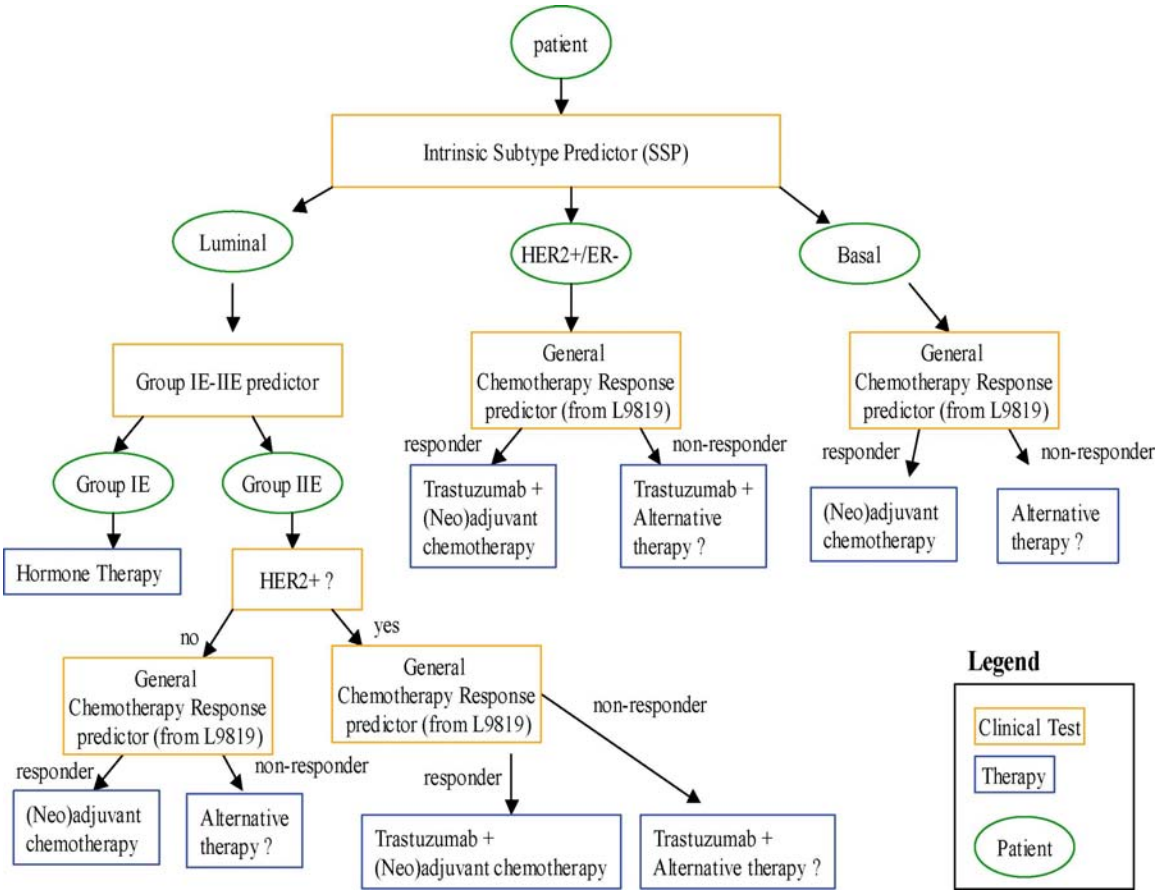


Figure 6.1. Proposed decision tree to determine therapy for the newly diagnosed breast cancer patient. This decision tree makes use of three different gene expression-based predictors discussed in this work: (1) the Intrinsic Subtype Predictor (SSP) introduced in Chapter 3, (2) the Group IE-IIE predictor introduced in Chapter 2, and (3) the general chemotherapy response predictor (from the L9819 clinical study) introduced in Chapter 5.