DEVELOPMENT OF BIOLOGICALLY BASED THERAPIES FOR BASAL-LIKE BREAST TUMORS

Katherine A. Hoadley

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: Charles Perou
Reader: Robert Duronio
Reader: Rosann Farber
Reader: Jason Lieb
Reader: James Swenberg

ABSTRACT

Katherine A. Hoadley: Development of Biologically Based Therapies for Basal-like Breast Tumors
(Under the direction of Charles M. Perou)

There have been many experiments on breast cancer cell lines and tumors with respect to identifying genes/pathways that are involved in cancer initiation, progression and response to therapy; however, only a few actually make suggestions that might affect treatment. The knowledge that breast cancer actually represents several diseases that arise from at least two different epithelial cells has been a major stepping-stone for stratifying patients and identifying more selective and biology-based therapies. Drugs aimed at the estrogen receptor, estrogen production, and HER2 have been very successful in the many patients whose tumors are dependent upon these signaling pathways for growth. Unfortunately for tumors that lack these markers, such as basal-like subtype, there are few treatment options. Until recently, few studies had actually considered if there were subtype-specific differences in response to chemotherapy. This dissertation focuses on the basal-like subtype of cancer and examines responses to chemotherapeutics relative to the luminal subtypes and evaluates the EGFR pathway as a place for potential therapeutic intervention.

In response to two chemotherapeutics – doxorubicin and 5-fluorouracil – a general stress response was the dominant profile and this profile varied both *in vitro and in vivo* between the subtypes. The drug-specific response was more similar in the subtypes. A predictive gene list was identified that could predict both subtype and drug treatment with

fairly high accuracy suggesting some degree of subtype-specific mechanism of action. The different responses to doxorubicin and 5-fluorouracil led us to evaluate sensitivity to a larger panel of drugs and cell lines and we determined that the basal-like subtype was more sensitive to carboplatin. While identification of chemotherapy regimens that are beneficial to the basal-like subtype is needed, drugs targeted to specific deregulated pathways in this subtype will be more effective in the long run. My work evaluated the EGFR pathway and determined it is high in 90% of all basal-like tumors, but I also identified high expression of genes downstream of EGFR that can induce EGFR-independent activation of this pathway. My data suggest that inhibition of MEK or PI3K, along with chemotherapeutics, may be an effective regimen for basal-like patients.

ACKNOWLEDGEMENTS

I would like to thank:

My advisor, Charles Perou

The Perou Lab, past and present members:

Melissa Troester

Xiaping He, Olga Karginova

Aaron Thorner, Jason Herschowitz, Victor Weigman, George Chao

Joel Parker, Chris Fan, George Wu

Jerry Usary, David Darr, Young Hu, Mei Liu, Okey Ukairo

My Committee Members:

Robert Duronio
Rosann Farber
Jason Lieb
James Swenberg

My Family: Jim, Norleen, Charlie, and Meggie Hoadley

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

LIST OF ABBREVIATIONS AND SYMBOLS

5FU	5-fluorouracil
95% CI	95% confidence interval
CI	Combination Index
CV	cross validation
DMSO	dimethylsulfoxide
DOX	doxorubicin
DWD	Distance Weighted Discrimination
EGFR/HER1	epidermal growth factor receptor
ER	estrogen receptor
ETOP	etoposide
FDR	false discovery rate
GO	gene ontology
HER	human epidermal growth factor receptor
HER2	v-erb-b2 erythroblastic leukemia viral oncogene homolog 2
HMEC	human mammary epithelial cell
HR	hazard ratio
hTERT	human telomerase reverse transcriptase
IC50	inhibitory concentration 50%
ID1	inhibitor of DNA binding 1
ID3	inhibitor of DNA binding 3
K	cytokeratin/keratin

KNN	k-nearest neighbors
MTT	mitochondrial dye conversion assay [3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide]
NKI	Nederlands Kanker Instituut
OS	overall survival
PAM	Prediction Analysis of Microarrays
PR	progesterone receptor
RFS	relapse-free survival
RS	recurrence score
SAM	Significance Analysis of Microarrays
SMA	smooth muscle actin
TOP2A	topoisomerase 2A
UNC	University of North Carolina at Chapel Hill

CHAPTER I

INTRODUCTION

Breast cancer is the second most common cancer in women, affecting approximately 1 in 7. There has been a slight increase in the incidence of breast cancer over the past few years. This may be the result of better screening or an increase in longevity. Over this same time period, there has been a decline in the mortality rate (American Cancer Society), but there is still a great need to improve upon our ability to accurately predict patient outcomes and to treat patients more effectively. Even today, intensive chemotherapy regimens with side effects that are difficult to tolerate are the standard of care for some patients, and often a long-term response is not achievable. Therefore, my research has focused on two important treatment refinements, namely identifying the right chemotherapy for the main tumor subtypes, and on investigating the involvement of the epidermal growth factor receptor (EGFR)-pathway in breast tumors and identifying those patients that might benefit from EGFR inhibitor containing therapies.

Normal Breast Development

An understanding of the biology of the normal breast epithelial cells is needed as the basis for understanding the diversity of breast tumor subtypes. In the normal human breast, the architecture of breast tissue is very diverse with many types of cells present. The breast is made up of a series of ducts, as well as structural tissue, that develop during several stages

of a female's lifespan (Howard and Gusterson, 2000; Russo and Russo, 2004). Prenatally, rudimentary duct structures are formed with two layers of epithelial cells, the beginning of the basement membrane, as well as the stromal layer. This two cell layer thick duct is composed of the inner luminal epithelial cells (which are sometimes hormone receptor positive) and the outer myoepithelial cell layer (which shows characteristics of basal epithelia). The stroma is a connective tissue consisting of fibroblasts, endothelial cells, lymphocytes, and adipocytes that supports the ducts and lobules. Cross talk between the stromal fibroblasts and epithelial cells influences normal development (Wiseman and Werb, 2002). Around the onset of puberty, the rudimentary duct structures proliferate and differentiate into elongated ducts and lobules with the formation of terminal and alveolar buds. The stroma also proliferates and starts accumulating fat. During pregnancy, there is a massive increase in proliferation and formation of acini in the alveolar buds. The breast proliferation increases so much that it is difficult to even distinguish the main duct from its side branches. After giving birth, milk is secreted from the alveolar/luminal cells and travels through the ducts to the nipple. Once lactation is complete, the breast goes through a period of involution, whereby the massive secondary and tertiary breast structures are reduced to near pre-pregnancy levels through apoptosis (Schorr et al., 1999; Strange et al., 2001).

Microarray Analysis for Identification of Breast Cancer

In the past decade, the use of DNA microarrays has greatly expanded our understanding of breast tumor biology. This technology allows for the identification of gene expression changes by the binding of fluorescently labeled RNA or cDNA to complementary cDNA or synthesized oligonucleotides that have been spotted onto known positions on a

glass slide (Duggan et al., 1999; Lockhart and Winzeler, 2000). In dual-color arrays, the ratio of the intensity of the experimental sample versus a reference sample can be calculated to determine the expression levels of genes within any sample tested. Microarrays have now been used to examine gene expression in a number of breast tumor patient cohorts from multiple groups (Perou et al., 2000; Alizadeh et al., 2001; Sørlie et al., 2001; van de Vijver et al., 2002; Sørlie et al., 2003; Sotiriou et al., 2003; Weigelt et al., 2003; Ma et al., 2004; Chang et al., 2005; Paik et al., 2006). Since breast tumor samples contain the many cell types present in the breast, in addition to the tumor tissue, and each breast has differences in development and structure, the composition of each tumor section will vary from sample to sample. Gene expression analysis identified the various cell types within each tumor section such as endothelial cells, stromal cells, adipose-enriched cells, B and T lymphocytes, macrophages and epithelial cells (Perou et al., 2000); thus the gene expression patterns recapitulate the complex histology of tumors.

Since the unsupervised approach included the expression of non-tumor cells, additional classification gene lists have been developed that can predict patients' outcomes (Perou et al., 2000). Our approach has been to use an "intrinsic analysis" for gene/feature selection, which is an analysis based upon paired samples that identifies genes that vary little within repeated samples of the same tumor, but which vary significantly across different tumors (Sørlie et al., 2001; Sørlie et al., 2003; Hu et al., 2006). Hierarchical clustering of tumors based on the intrinsic gene list stratifies tumors into several subtypes (Figure 1.1).

The tumors separated into two main branches defined largely by the expression of the luminal epithelial gene set that contains the *estrogen receptor (ER)*, a key breast cancer

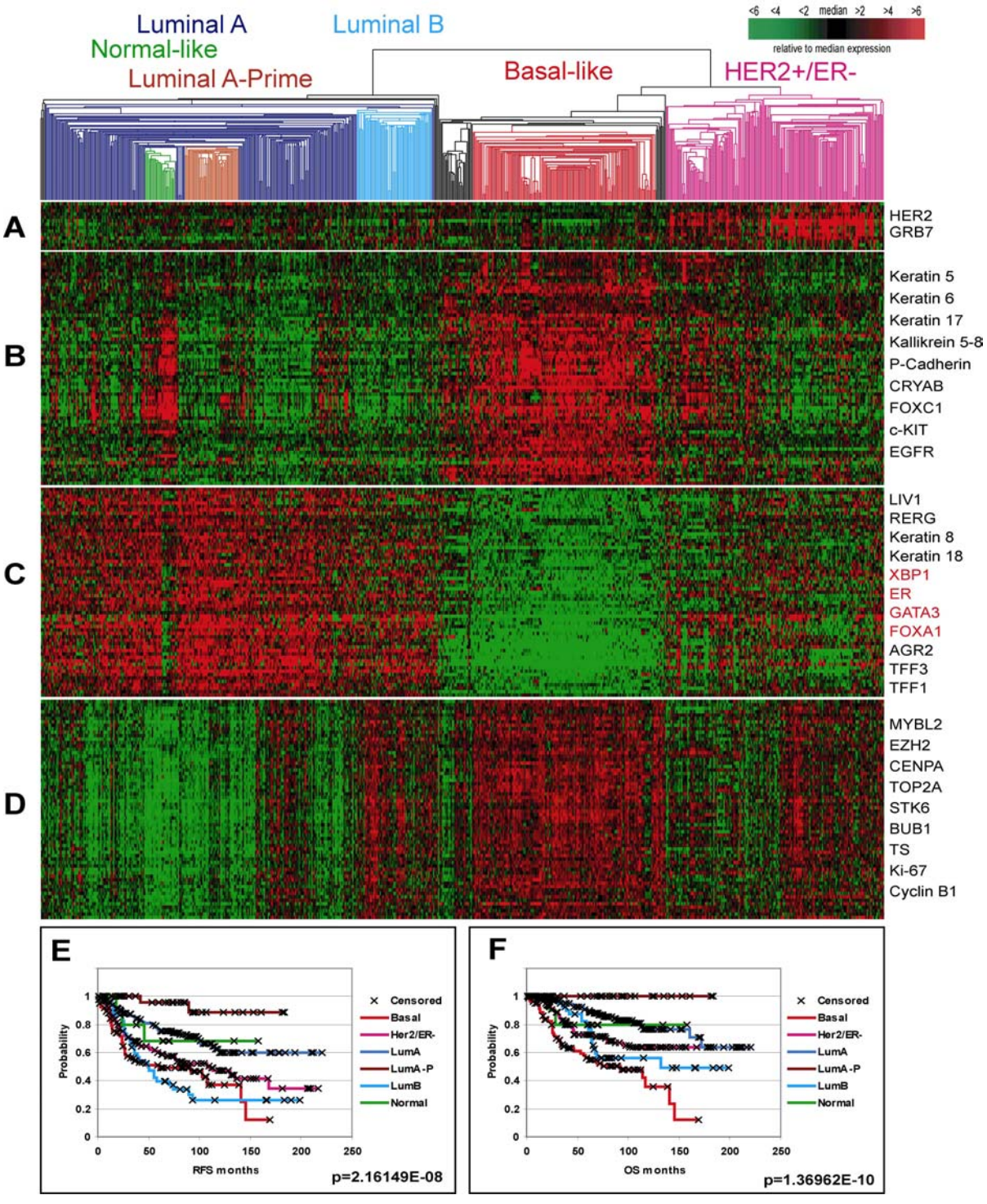
Luminal A
Normal-like
Luminal A-Prime
Luminal B
Basal-like
HER2+/ER-
<6 <4 <2 median >2 >4 >6
relative to median expression
A
HER2
GRB7
B
Keratin 5
Keratin 6
Keratin 17
Kallikrein 5-8
P-Cadherin
CRYAB
FOXC1
c-KIT
EGFR
C
LIV1
RERG
Keratin 8
Keratin 18
XBP1
ER
GATA3
FOXA1
AGR2
TFF3
TFF1
D
MYBL2
EZH2
CENPA
TOP2A
STK6
BUB1
TS
Ki-67
Cyclin B1
E
Probability
RFS months
p=2.16149E-08
F
OS months
p=1.36962E-10
Censored
Basal
Her2/ER-
LumA
LumA-P
LumB
Normal

Figure 1.1

Figure 1.1. Breast tumors analyzed using hierarchical clustering and the intrinsic gene list. A single data set of 340 samples from University of North Carolina at Chapel Hill and 337 from the Netherlands Cancer Institute were combined using Distance Weighted Discrimination (Benito et al., 2004), and then clustered together with the 1300 intrinsic gene list (Hu et al., 2006) to yield a large, homogenous data set containing over 470 different tumors with RFS and OS data. The clustering analysis identified the five major intrinsic subtype of Luminal A, Luminal B, Normal-like, Basal-like and HER2+/ER-. The Luminal A tumors can be further subdivided into Luminal A and Luminal A-Prime. *A*) *HER2*-amplicon gene set. *B*) Basal epithelial gene set. *C*) Luminal epithelial gene set. *D*) Proliferation gene set. *E)* Kaplan-Meier plots for Relapse-Free Survival for the six groups described in this Figure (Luminal A-Prime, Luminal A, Luminal B, Basal-like, HER2+/ER-, and Normal-like). *G*) Kaplan Meier plots for Overall Survival for the six groups.

marker that has been used for classifying breast cancer because it is a drug target (Figure 1.1C). Within the luminal/ER+ tumors were at least two sub-branches termed luminal A and luminal B. These tumors expressed *keratins 8* and *18*, as well as the *ER* and estrogen regulated-regulated genes including *RERG* and *LIV1*. The main difference between these two subtypes is that the luminal B subtype has reduced expression level of the luminal/ER+ cluster and has higher expression of a gene set that is indicative of cellular proliferation rates (Figure 1.1D) (Whitfield et al., 2002; Whitfield et al., 2006). Our lab has further stratified the luminal A tumors into two groups of which one we have termed luminal A-prime. These tumors are characterized by low proliferation and low recurrence score based on Paik et al.'s predictor (Paik et al., 2006). Also in this branch was the normal breast and normal-like breast tumor with expression of genes from adipose cells and *cytokeratin 5/6* expression (Figure 1.1B). The tumor samples that cluster with the normal tissue samples may have clustered in this branch because the sampled tumor section contained high amounts of surrounding normal tissue.

The *ER*-negative dendrogram branch contained at least two sub-branches. One of these branches included the other main breast cancer marker, *v-erb-b2 erythroblastic leukemia viral oncogene homolog 2 (HER2)* and defines the HER2+/ER- subtype, which showed high expression of *HER2* and many other genes on the 17q amplicon. The basal-like subtype lacked both expression of ER and the overexpression of *HER2*. This group had high expression levels of proliferation genes and was positive for *keratins 5, 6* and *17* (van de Rijn et al., 2002), which are typically expressed in basal epithelia of the human body. Using IHC (Perou et al., 2000) or immunofluorescence, these subtypes were hypothesized to arise from different cells within the breast based in part upon their unique keratin expression patterns

(Figure 1.2). Keratin 8 and 18 were present in the luminal cells that line the lumen and are hormonally responsive, and also present in all tumors called luminal by gene expression (Figure 1.2A and F). Conversely, keratin 5/6 were present in myoepithelial and basal-like cells surrounding the layer of luminal cells, as well as most tumors called basal-like by gene expression (Figure 1.2B and E). Overlay of the two images shows that the Keratin 8/18 and 5/6 expression were mostly mutually exclusive in normal breast tissues, thus providing evidence that breast cancer is a heterogeneous disease arising from at least two different epithelial cell lineages within the breast (Figure 1.2D).

Both the diversity of the breast tissue as well as the heterogeneity of epithelial breast cancers was observable at the gene expression level. In addition, individual variation was also observable. Gene expression signatures of before and after chemotherapy pairs, or tumor-lymph node metastasis pairs, were typically much more similar to each other and clustered on terminal branches of the dendrogram (Perou et al., 2000). This underscores the individuality of every tumor. Therefore, one-size-fits-all treatments will not work for all.

Breast Cancer Subtypes Predict Outcomes

Based on gene expression, breast tumors can be stratified into several subtypes, as described above. However, one may argue that the stratification of breast cancer will merely result in a host of subtypes without any clinical results, but these subtypes significantly predict survival outcomes. Kaplan-Meier survival analysis of overall survival (OS) and relapse free survival (RFS) data shows that these subtypes have significantly different

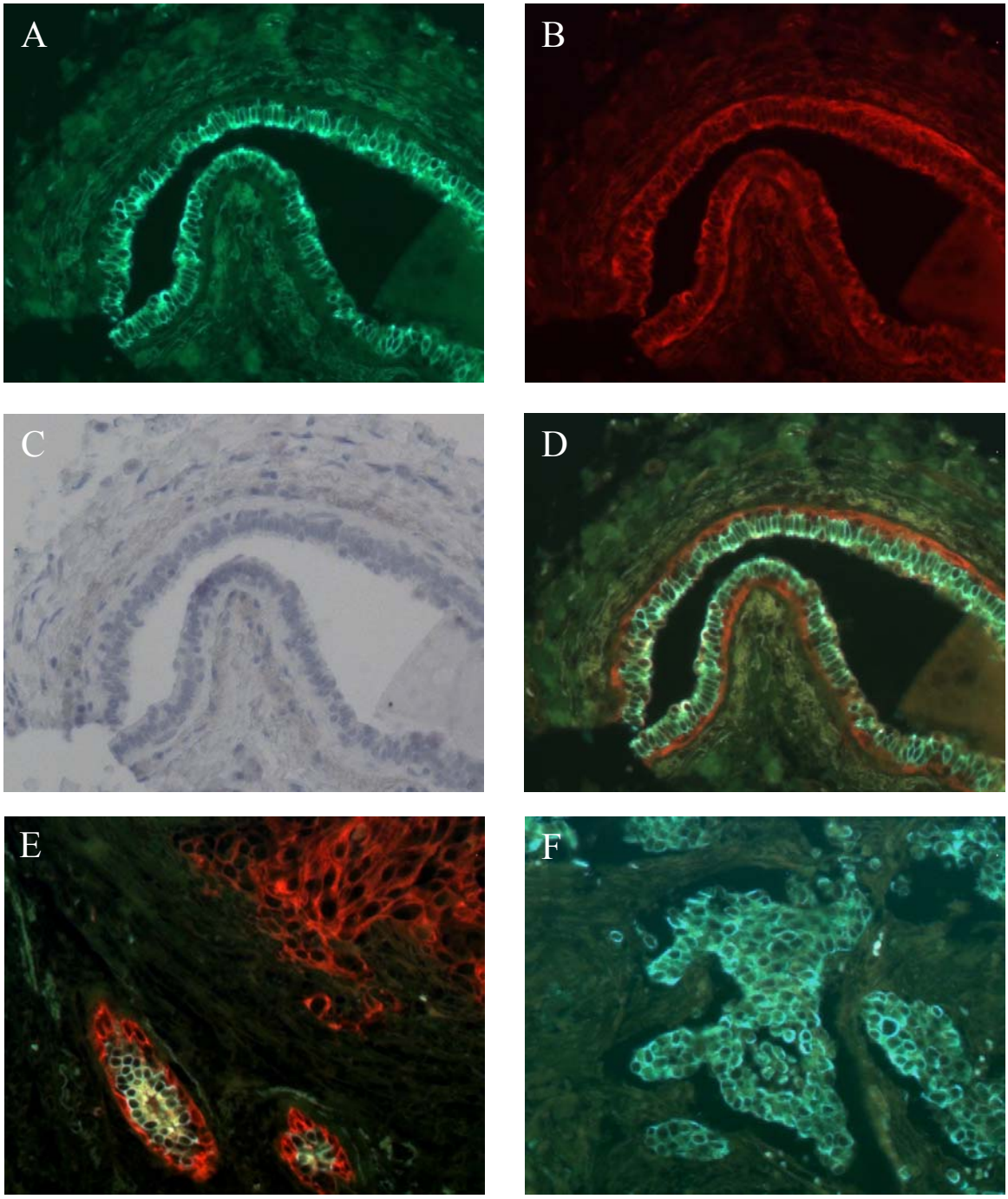

Figure 1.2. Immunofluorescence of normal breast ducts and breast tumors. *A*) A normal breast duct probed with FITC-conjugated antibodies to Cytokeratin (K) 8/18; K8/18 is expressed in luminal epithelial cells. *B*) Normal breast duct probed with Rhodamine-conjugated antibody to K5; K5 is expressed in myoepithelial cells. *C*) Light field image of normal breast duct. *D*) Merged Field, notice compartmentalization of K8/18 and K5. *E*) Infiltrating basal-like breast tumor (top right of image) with K5 expression and presence of normal breast duct (bottom left) with K5 expression in myoepithelial and K8/19 expression in luminal cells. *F*) Luminal tumors with only K8/18 expression. Immunofluorescence was performed by Xiaping He.

outcomes (Sørlie et al., 2001; Sørlie et al., 2003; Hu et al., 2006) (Figure 1.1). Luminal A-Prime tumors have the best survival outcomes while the luminal B, HER2+/ER-, and basal-like have poor outcomes. While each breast tumor has individual characteristics, they share subtype characteristics that can predict outcome.

Subtype identification through intrinsic analysis holds up among multiple independent, therapy-treated and non-treated patient sets (Hu et al., 2006). The most recent intrinsic list was derived from 105 UNC tumors on a large array platform to create a list with 1300 genes and included a proliferation signature (Hu et al., 2006). In all the data sets, the intrinsic subtypes were identified and they shared similar outcome predictions across data sets as well. The different data sets were also combined into a single data set using Distance Weighted Discrimination (DWD), which is a correction tool to allow direct comparison and integration of the different data sets (Benito et al., 2004). The intrinsic subtypes can be identified even in this large mixed dataset and the samples cluster based on subtype and not by platform or user. This once again showed difference in patient outcomes and supports the robustness of the intrinsic list to identify the prognostic subtypes.

Model of Breast Epithelial Differentiation

Several labs have hypothesized a model for the development of the epithelial layers of the breast, and we believe that the subtypes described above are highly related to potential stages of breast epithelial cell development. As mentioned above, the breast is a dynamic tissue that undergoes several stages of growth mostly at sexual maturity and during pregnancy where there is a large increase in proliferation followed by involution at the end of lactation (Russo and Russo, 2004). The ability of the breast to go through numerous rounds

of proliferation, differentiation, and involution has led to the hypothesis of a breast stem cell (Boecker and Buerger, 2003; Clarke, 2005; Woodward et al., 2005). Indeed, several labs have shown in mice the ability to regenerate an entire functional murine mammary gland in a cleared mammary fat pad from a single mammary cell (Shackleton et al., 2006; Stingl et al., 2006). Human mammary duct structures have also been generated from small numbers of organoids transplanted into in a cleared murine mammary fat pad of NOD/SCID mice (Kuperwasser et al., 2004).

Several labs have hypothesized models of breast differentiation. Upon signals for growth, a breast stem cell will differentiate into a committed progenitor that expresses keratins 5/6 (Figure 1.3). This cell has the potential for multi-potent differentiation. We postulate that one lineage option is to gain expression of smooth muscle actin and p63 to become a myoepithelial cell. Myoepithelial cells make up the layer of cells that surround the luminal cells and excrete collagens, intregrins, and laminins into the basement membrane which helps keep the epithelial cells correctly organized and oriented (Gudjonsson et al., 2002). Myoepithelial cells rarely develop into cancer, potentially due to the observation that myoepithelial cells have a number of tumor suppressors, which make them highly resistant to transformation (Sternlicht et al., 1997; Barsky, 2003). The other main lineage option is to gain expression of keratins 8/18, lose expression of keratins 5/6, and later gain expression of ER and other ER-regulated genes.

Based on this theory, we hypothesize that the luminal cells arise from the more differentiated cells with the keratins 8/18 and ER expression. The HER2 positive tumors may arise from the intermediates between the committed progenitor and the fully differentiated luminal cells since HER2+ tumors can be either ER+ or ER-. The basal-like

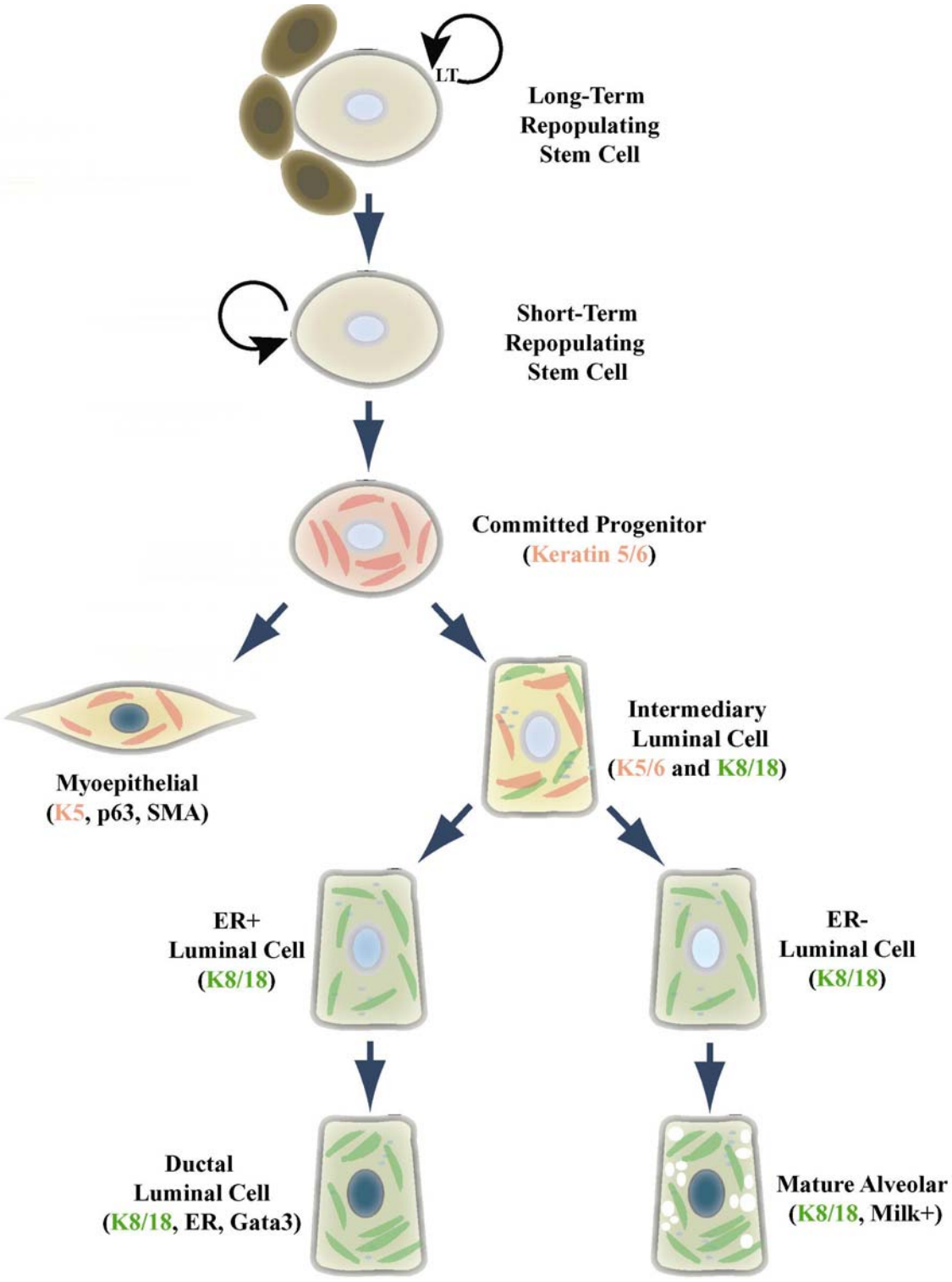
LT
Long-Term
Repopulating
Stem Cell
Short-Term
Repopulating
Stem Cell
Committed Progenitor
(Keratin 5/6)
Myoepithelial
(K5, p63, SMA)
Intermediary
Luminal Cell
(K5/6 and K8/18)
ER+
Luminal Cell
(K8/18)
ER-
Luminal Cell
(K8/18)
Ductal
Luminal Cell
(K8/18, ER, Gata3)
Mature Alveolar
(K8/18, Milk+)

Figure 1.3

Figure 1.3. Model of breast epithelial differentiation. A mammary stem cell has been hypothesized to give rise to the different epithelial lineages in the breast. In this model there is a long term repopulating stem cell within a stem cell niche. Progressive differentiation can lead to a committed progenitor that is positive for keratins 5/6. One possible avenue for further differentiation is to gain expression of keratins 8/18 and eventual loss of keratin 5/6 expression. The luminal intermediate cell may develop into an ER positive cell, an ER negative cell, or an alveolar secretory cell. The luminal cells line the lumen of the milk duct and the alveolar cells line the buds at the end of the duct and produce milk. Another avenue for differentiation is to gain smooth muscle actin (SMA) and p63 to become a myoepithelial cell that surrounds and provides support for the luminal and alveolar cells. Cell images were designed by Jason Herschkowitz.

tumors most probably arise from the committed progenitor as opposed to a mature myoepithelial cells because we know that the basal-like tumors lack the expression of p63 and smooth muscle actin, which define myoepithelial cells and their true tumors, seen as rare myoepitheliomas (Livasy et al., 2006).

Treatment Options

Much of the inability to accurately predict patient outcomes is due to the fact that breast cancer is heterogeneous. For years, clinicians have used the presence of two markers, ER and HER2, as well as size, grade, node status, age, and morphology, to help categorize breast tumors in order to guide treatment. ER is high in a subset of patients while HER2 is amplified in another subset of tumors. The identification of these markers led to the development of several effective inhibitors that are specific for these molecules. Tamoxifen, an estrogen-like compound that binds ER and prevents growth, was developed in the 1960s (Jordan, 2003). In addition, aromatase inhibitors have been developed that prevent the production of estrogen from androgens in postmenopausal women (Brueggemeier, 2006). Hormonal therapy has become a standard treatment regimen for ER positive tumors. Trastuzumab, a monoclonal antibody to the extracellular ligand-binding domain of HER2, prevents subsequent downstream signaling that leads to increased growth and decreased apoptosis. Trastuzumab clinical trials showed significant improvement with treatment and was recently approved for wide spread clinical use (Romond et al., 2005). These drugs have made huge advancements in breast cancer patient survival because they target the selected cancer cells with fewer side effects than chemotherapy, and because they are only given to the patients who express the drugs' target. Thus, the development of diagnostics to identify

the patients that might benefit from these drugs was almost as important as the development of the drugs themselves. However, some patients eventually do succumb while on these therapies suggesting that additional therapies are needed. Obviously, there is a lot of variation among tumors and a better understanding of the underlying genetic variation across the tumors might give more insight into how to treat breast tumors.

Chemotherapy is the standard of care for almost all intermediate and high-risk breast cancer patients; however, treatment regimens vary from location to location and from physician to physician. Guidelines have been developed for adjuvant chemotherapy from consensus panels at several conferences, such as St. Gallen and National Institute of Health, to help guide treatment options based on clinical and pathological information (Eifel et al., 2001; Goldhirsch et al., 2005). Two online websites, Nottingham Prognostic Index (Galea et al., 1992) and Adjuvant! Online (Ravdin et al., 2001), have also been developed to help guide treatment regimens. However, these methods could be improved upon if they could incorporate some of the genomic information that was described above.

To best determine treatment, we need to have a better understanding of the biology of these subtypes and how the subtypes interact with the current biomarkers that guide therapy. The gene expression data identified in the intrinsic subtypes have provided additional knowledge about the different subtypes. The luminal/ER+ tumors typically have better outcomes than the other subtypes. However, even within the luminal tumors gene expression studies have been able to further elucidate treatment options. Paik et al classified ER positive, node-negative tumors according to a Recurrence Score (RS) based on the gene expression levels of 21 genes that predicts the tumors propensity to recur (Paik et al., 2006). They showed that high RS scores correlated with a high likelihood of recurrence and a large

benefit from chemotherapy, whereas, low or moderate RS scores gave lower propensities to recur, no benefit from adjuvant chemotherapy and these patients gained a benefit from tamoxifen. High RS trended to show high proliferation and we have shown that the high RS ER-positive tumors are typically luminal B (Fan et al., 2006). This sort of knowledge can help improve the treatment option for these patients by directing chemotherapy regimens to those patients that will actually benefit.

For ER-negative tumors, the benefit of chemotherapy is better established. The basal-like group, which lacks ER and HER2 expression, is currently limited to the cytotoxic chemotherapy. A major concern with a chemotherapy-only option for the basal-like subtype is that the majority of the preclinical and clinical testing was performed on unselected subtype populations that contained high percentages of the more prevalent and better outcome luminal subtype. Therefore, it is necessary to have a better understanding of the biology behind the basal-like subtype and how it responds to chemotherapy. There is also a need for the development of targeted therapies for the basal-like subtype. Unlike chemotherapy which affects all rapidly dividing cells, thus leading to side effects, therapies that target specific molecules unique to the tumor will allow for a more efficient treatment with fewer side effects and potentially a better outcome.

Research Introduction

Noting these issues with treatment decisions, especially concerning basal-like tumors, Chapter 2 and Chapter 3 are devoted to evaluating the chemotherapeutic-induced gene expression responses in two subtypes of breast cancer, the basal-like and luminal. As mentioned previously, most *in vitro* and *in vivo* studies have included the luminal subtype

when observing effects of chemotherapeutics; however, a focus on the basal-like subtype has been lacking. Therefore, one main goal was to determine if similar or distinct responses to common chemotherapeutics between these two subtypes occurred. 5-fluorouracil (5FU) and doxorubicin (DOX) were chosen based on availability of *in vivo* tumor gene expression data from patients who had been treated with these drugs. Therefore, we could observe gene expression in both cell lines and tumors. We detail the dramatically different responses that we observed. In Chapter 2, a general stress response to both 5FU and DOX was noted to be the dominant profile, and in Chapter 3 we identify smaller, drug-specific signatures. The basal-like tumors are ER-, progesterone receptor (PR)- and HER2-, and thus, they are in desperate need of a biologically targeted agent(s). In Chapter 4, the focus turns to the EGFR pathway as a possible biologically based target pathway present within basal-like tumors.

REFERENCES

Alizadeh, A. A., Ross, D. T., Perou, C. M. and van De Rijn, M. (2001). Towards a novel classification of human malignancies based on gene expression patterns. J Pathol 195(1): 41-52.

Barsky, S. H. (2003). Myoepithelial mRNA expression profiling reveals a common tumor-suppressor phenotype. Exp Mol Pathol 74(2): 113-122.

Benito, M., Parker, J., Du, Q., Wu, J., Xiang, D., Perou, C. M. and Marron, J. S. (2004). Adjustment of systematic microarray data biases. Bioinformatics 20(1): 105-114.

Boecker, W. and Buerger, H. (2003). Evidence of progenitor cells of glandular and myoepithelial cell lineages in the human adult female breast epithelium: a new progenitor (adult stem) cell concept. Cell Prolif 36 Suppl 1: 73-84.

Brueggemeier, R. W. (2006). Update on the use of aromatase inhibitors in breast cancer. Expert Opin Pharmacother 7(14): 1919-1930.

Chang, H. Y., Nuyten, D. S., Sneddon, J. B., Hastie, T., Tibshirani, R., Sorlie, T., Dai, H., He, Y. D., van't Veer, L. J., Bartelink, H., van de Rijn, M., Brown, P. O. and van de Vijver, M. J. (2005). Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. Proc Natl Acad Sci U S A 102(10): 3738-3743.

Clarke, R. B. (2005). Isolation and characterization of human mammary stem cells. Cell Prolif 38(6): 375-386.

Duggan, D. J., Bittner, M., Chen, Y., Meltzer, P. and Trent, J. M. (1999). Expression profiling using cDNA microarrays. Nat Genet 21(1 Suppl): 10-14.

Eifel, P., Axelson, J. A., Costa, J., Crowley, J., Curran, W. J., Jr., Deshler, A., Fulton, S., Hendricks, C. B., Kemeny, M., Kornblith, A. B., Louis, T. A., Markman, M., Mayer, R. and Roter, D. (2001). National Institutes of Health Consensus Development Conference Statement: adjuvant therapy for breast cancer, November 1-3, 2000. J Natl Cancer Inst 93(13): 979-989.

Fan, C., Oh, D. S., Wessels, L., Weigelt, B., Nuyten, D. S., Nobel, A. B., van't Veer, L. J. and Perou, C. M. (2006). Concordance among gene-expression-based predictors for breast cancer. N Engl J Med 355(6): 560-569.

Galea, M. H., Blamey, R. W., Elston, C. E. and Ellis, I. O. (1992). The Nottingham Prognostic Index in primary breast cancer. Breast Cancer Res Treat 22(3): 207-219.

Goldhirsch, A., Glick, J. H., Gelber, R. D., Coates, A. S., Thurlimann, B. and Senn, H. J. (2005). Meeting highlights: international expert consensus on the primary therapy of early breast cancer 2005. Ann Oncol 16(10): 1569-1583.

Gudjonsson, T., Ronnov-Jessen, L., Villadsen, R., Rank, F., Bissell, M. J. and Petersen, O. W. (2002). Normal and tumor-derived myoepithelial cells differ in their ability to interact with luminal breast epithelial cells for polarity and basement membrane deposition. J Cell Sci 115(Pt 1): 39-50.

Howard, B. A. and Gusterson, B. A. (2000). Human breast development. J Mammary Gland Biol Neoplasia 5(2): 119-137.

Hu, Z., Fan, C., Oh, D. S., Marron, J. S., He, X., Qaqish, B. F., Livasy, C., Carey, L. A., Reynolds, E., Dressler, L., Nobel, A., Parker, J., Ewend, M. G., Sawyer, L. R., Wu, J., Liu, Y., Nanda, R., Tretiakova, M., Ruiz Orrico, A., Dreher, D., Palazzo, J. P., Perreard, L., Nelson, E., Mone, M., Hansen, H., Mullins, M., Quackenbush, J. F., Ellis, M. J., Olopade, O. I., Bernard, P. S. and Perou, C. M. (2006). The molecular portraits of breast tumors are conserved across microarray platforms. BMC Genomics 7: 96.

Jordan, V. C. (2003). Tamoxifen: a most unlikely pioneering medicine. Nat Rev Drug Discov 2(3): 205-213.

Kuperwasser, C., Chavarria, T., Wu, M., Magrane, G., Gray, J. W., Carey, L., Richardson, A. and Weinberg, R. A. (2004). Reconstruction of functionally normal and malignant human breast tissues in mice. Proc Natl Acad Sci U S A 101(14): 4966-4971.

Livasy, C. A., Karaca, G., Nanda, R., Tretiakova, M. S., Olopade, O. I., Moore, D. T. and Perou, C. M. (2006). Phenotypic evaluation of the basal-like subtype of invasive breast carcinoma. Mod Pathol 19(2): 264-271.

Lockhart, D. J. and Winzeler, E. A. (2000). Genomics, gene expression and DNA arrays. Nature 405(6788): 827-836.

Ma, X. J., Wang, Z., Ryan, P. D., Isakoff, S. J., Barmettler, A., Fuller, A., Muir, B., Mohapatra, G., Salunga, R., Tuggle, J. T., Tran, Y., Tran, D., Tassin, A., Amon, P., Wang, W., Enright, E., Stecker, K., Estepa-Sabal, E., Smith, B., Younger, J., Balis, U., Michaelson, J., Bhan, A., Habin, K., Baer, T. M., Brugge, J., Haber, D. A., Erlander, M. G. and Sgroi, D. C. (2004). A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen. Cancer Cell 5(6): 607-616.

Paik, S., Tang, G., Shak, S., Kim, C., Baker, J., Kim, W., Cronin, M., Baehner, F. L., Watson, D., Bryant, J., Costantino, J. P., Geyer, C. E., Jr., Wickerham, D. L. and Wolmark, N. (2006). Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. J Clin Oncol 24(23): 3726-3734.

Perou, C. M., Sørlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., Fluge, O., Pergamenschikov, A., Williams, C., Zhu, S. X., Lonning, P. E., Borresen-Dale, A. L., Brown, P. O. and Botstein, D. (2000). Molecular portraits of human breast tumours. Nature 406(6797): 747-752.

Ravdin, P. M., Siminoff, L. A., Davis, G. J., Mercer, M. B., Hewlett, J., Gerson, N. and Parker, H. L. (2001). Computer program to assist in making decisions about adjuvant therapy for women with early breast cancer. J Clin Oncol 19(4): 980-991.

Romond, E. H., Perez, E. A., Bryant, J., Suman, V. J., Geyer, C. E., Jr., Davidson, N. E., Tan-Chiu, E., Martino, S., Paik, S., Kaufman, P. A., Swain, S. M., Pisansky, T. M., Fehrenbacher, L., Kutteh, L. A., Vogel, V. G., Visscher, D. W., Yothers, G., Jenkins, R. B., Brown, A. M., Dakhil, S. R., Mamounas, E. P., Lingle, W. L., Klein, P. M., Ingle, J. N. and Wolmark, N. (2005). Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. N Engl J Med 353(16): 1673-1684.

Russo, J. and Russo, I. H. (2004). Development of the human breast. Maturitas 49(1): 2-15.

Schorr, K., Li, M., Krajewski, S., Reed, J. C. and Furth, P. A. (1999). Bcl-2 gene family and related proteins in mammary gland involution and breast cancer. J Mammary Gland Biol Neoplasia 4(2): 153-164.

Shackleton, M., Vaillant, F., Simpson, K. J., Stingl, J., Smyth, G. K., Asselin-Labat, M. L., Wu, L., Lindeman, G. J. and Visvader, J. E. (2006). Generation of a functional mammary gland from a single stem cell. Nature 439(7072): 84-88.

Sørlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Thorsen, T., Quist, H., Matese, J. C., Brown, P. O., Botstein, D., Eystein Lonning, P. and Borresen-Dale, A. L. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 98(19): 10869-10874.

Sørlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J. S., Nobel, A., Deng, S., Johnsen, H., Pesich, R., Geisler, S., Demeter, J., Perou, C. M., Lonning, P. E., Brown, P. O., Borresen-Dale, A. L. and Botstein, D. (2003). Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A 100(14): 8418-8423.

Sotiriou, C., Neo, S. Y., McShane, L. M., Korn, E. L., Long, P. M., Jazaeri, A., Martiat, P., Fox, S. B., Harris, A. L. and Liu, E. T. (2003). Breast cancer classification and prognosis based on gene expression profiles from a population-based study. Proc Natl Acad Sci U S A 100(18): 10393-10398.

Sternlicht, M. D., Kedeshian, P., Shao, Z. M., Safarians, S. and Barsky, S. H. (1997). The human myoepithelial cell is a natural tumor suppressor. Clin Cancer Res 3(11): 1949-1958.

Stingl, J., Eirew, P., Ricketson, I., Shackleton, M., Vaillant, F., Choi, D., Li, H. I. and Eaves, C. J. (2006). Purification and unique properties of mammary epithelial stem cells. Nature 439(7079): 993-997.

Strange, R., Metcalfe, T., Thackray, L. and Dang, M. (2001). Apoptosis in normal and neoplastic mammary gland development. Microsc Res Tech 52(2): 171-181.

van de Rijn, M., Perou, C. M., Tibshirani, R., Haas, P., Kallioniemi, O., Kononen, J., Torhorst, J., Sauter, G., Zuber, M., Kochli, O. R., Mross, F., Dieterich, H., Seitz, R., Ross, D., Botstein, D. and Brown, P. (2002). Expression of cytokeratins 17 and 5 identifies a group of breast carcinomas with poor clinical outcome. Am J Pathol 161(6): 1991-1996.

van de Vijver, M. J., He, Y. D., van't Veer, L. J., Dai, H., Hart, A. A., Voskuil, D. W., Schreiber, G. J., Peterse, J. L., Roberts, C., Marton, M. J., Parrish, M., Atsma, D., Witteveen, A., Glas, A., Delahaye, L., van der Velde, T., Bartelink, H., Rodenhuis, S., Rutgers, E. T., Friend, S. H. and Bernards, R. (2002). A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 347(25): 1999-2009.

Weigelt, B., Glas, A. M., Wessels, L. F., Witteveen, A. T., Peterse, J. L. and van't Veer, L. J. (2003). Gene expression profiles of primary breast tumors maintained in distant metastases. Proc Natl Acad Sci U S A 100(26): 15901-15905.

Whitfield, M. L., George, L. K., Grant, G. D. and Perou, C. M. (2006). Common markers of proliferation. Nat Rev Cancer 6(2): 99-106.

Whitfield, M. L., Sherlock, G., Saldanha, A. J., Murray, J. I., Ball, C. A., Alexander, K. E., Matese, J. C., Perou, C. M., Hurt, M. M., Brown, P. O. and Botstein, D. (2002). Identification of genes periodically expressed in the human cell cycle and their expression in tumors. Mol Biol Cell 13(6): 1977-2000.

Wiseman, B. S. and Werb, Z. (2002). Stromal effects on mammary gland development and breast cancer. Science 296(5570): 1046-1049.

Woodward, W. A., Chen, M. S., Behbod, F. and Rosen, J. M. (2005). On mammary stem cells. J Cell Sci 118(Pt 16): 3585-3594.

CHAPTER II

CELL-TYPE-SPECIFIC RESPONSES TO CHEMOTHERAPEUTICS IN BREAST CANCER

PREFACE

This work was previously published and represents a co-first author effort between Melissa Troester and myself. My role in this project included the 5-fluorouracil treatments and microarray experiments and I performed the tumor analysis. I participated in the data analysis and manuscript preparation and revisions. Melissa Troester performed the doxorubicin treatments, data analysis, and wrote the manuscript. Therese Sørlie, Anne-Lise Børresen-Dale, Per Eystein Lønning, and Charles Perou collected the tumor samples and performed the microarray experiments. Brittney-Shea Herbert and Jerry Shay provided the ME16C line. William Kaufmann tested the karyotyping of the HMEC lines. Charles Perou conceived the project.

Melissa A. Troester, Katherine A. Hoadley, Therese Sørlie, Brittney-Shea Herbert, Anne-Lise Børresen-Dale, Per Eystein Lønning, Jerry W. Shay, William K. Kaufmann, and Charles M. Perou. (2004). Cell-type-specific responses to chemotherapeutics in breast cancer. Cancer Research 64(12): 4218-4226.

ABSTRACT

Recent microarray studies have identified distinct subtypes of breast tumors that arise from different cell types and that show statistically significant differences in patient outcome. To gain insight into these differences, we identified *in vitro* and *in vivo* changes in gene expression induced by chemotherapeutics. We treated two cell lines derived from basal-like epithelium (immortalized human mammary epithelial cells) and two lines derived from luminal epithelium (MCF-7 and ZR-75-1) with chemotherapeutics used in the treatment of breast cancer and assayed for changes in gene expression using DNA microarrays. Treatment doses for doxorubicin and 5-fluorouracil were selected to cause comparable cytotoxicity across all four cell lines. The dominant expression response in each of the cell lines was a general stress response; however, distinct expression patterns were observed. Both cell types induced DNA damage response genes such as $p21^{waf1}$, but the response in the luminal cells showed higher fold changes and included more p53-regulated genes. Luminal cell lines repressed a large number of cell cycle regulated genes and other genes involved in cellular proliferation, whereas the basal-like cell lines did not. Instead, the basal-like cell lines repressed genes that were involved in differentiation. These *in vitro* responses were compared with expression responses in breast tumors sampled before and after treatment with doxorubicin or 5-fluorouracil /mitomycin C. The *in vivo* data corroborated the cell-type specific responses to chemotherapeutics observed *in vitro*, including the induction of $p21^{waf1}$. Similarities between *in vivo* and *in vitro* responses help to identify important response mechanisms to chemotherapeutics.

INTRODUCTION

The response of breast tumors to cytotoxic chemotherapeutic agents such as doxorubicin (DOX) and 5-fluoruracil (5FU) varies significantly across individuals. Sensitivity to these compounds is associated with *HER2* overexpression (Thor et al., 1998), *p53* status (Aas et al., 1996), and *topoisomerase IIα* amplification or deletion (Jarvinen et al., 2000), but the mechanisms of chemoresistance are still poorly understood. To better understand variations in clinical responses to treatment, recent studies have used gene expression patterns to identify major biological subtypes of breast cancer. These studies identified a previously unrecognized tumor subtype with characteristics of breast basal-like epithelium (Perou et al., 2000; Sørlie et al., 2001; van 't Veer et al., 2002; Sørlie et al., 2003; Sotiriou et al., 2003). Basal-like tumors are estrogen receptor α (ERα)-negative, do not overexpress *HER2*, and they have a poor prognosis compared to tumors derived from luminal epithelium (Sørlie et al., 2001; Sørlie et al., 2003).

Basal-like and luminal breast tumors are often treated with the same chemotherapeutic agents, but little is known about how each cell type responds to these drugs. To improve our understanding of how basal-like and luminal epithelium differ in their responses to chemotherapy, we selected two representative cell lines from each of these breast epithelial cell types to study; two human mammary epithelial (HME) cell lines immortalized by the overexpression of the catalytic subunit of telomerase (hTERT) represent the basal-like subtype, and two breast tumor-derived cell lines (MCF-7 and ZR-75-1) represent the luminal subtype (Ross et al., 2000). All four cell lines express wildtype p53 protein. True to their corresponding tumor subtypes, the HME lines are ERα-negative and the

luminal cancer cell lines are ERα-positive. We treated all four cell lines with DOX and 5FU and performed expression profiling to identify patterns of response.

Transcriptional profiling is a powerful approach for investigating cellular responses to drugs. This approach has led to greater understanding of pathway inhibition and off-target drug effects (Hughes et al., 2000), the response of yeast to genotoxic agents and environmental stresses (Gasch et al., 2000), and the effects of different kinds of DNA-damaging agents in human cells (Heinloth et al., 2003). Our analyses showed that the transcriptional responses of the basal-like and luminal cell lines to chemotherapeutics are quite distinct. We also correlated our *in vitro* data with *in vivo* data on breast tumors sampled before and after treatment with DOX or 5FU/mitomycin C (Perou et al., 2000; Sørlie et al., 2001; Sørlie et al., 2003) and we identified commonalities. Taken together, these *in vitro* and *in vivo* data sets illustrate that cell type is an important determinant of response to commonly used chemotherapeutics.

MATERIALS AND METHODS

Cells and Culture Conditions

HME31-hTERT no. 16C (ME16C) cells were derived from a clone of a finite lifespan HME cell culture isolated from the uninvolved tissue of a 53-year-old woman with unilateral breast cancer and no family history of breast cancer; HME31 postselection cells were infected with the retrovirus pBABE-puro-hTERT and an immortal population was established (ME16C). A second immortal HME clone, HME-CC, was a gift from Christopher Counter (Duke University, Durham NC); to derive the HME-CC cells, a HME cell isolate (Clonetics) was infected with the retrovirus pBabe-hygro-hTERT and an

immortal population was established. ME16C and HME-CC cells were maintained in mammary epithelium growth media (Cambrex Bio Science, Walkersville, MD). Karyotyping on the ME16C and HME-CC lines was conducted as described in Wang and Fedoroff (Wang and Fedoroff, 1972) at the University of North Carolina at Chapel Hill Chromosome Imaging Core Facility. A single isolate of the HME-CC was found to be trisomic for chromosome 20 in 65% of metaphase spreads and contained 9q+ and 18q+ in 25% of metaphase spreads. Two different isolates of ME16C were studied, and both were shown to be trisomic for chromosome 20 in 50% of metaphase spreads and marker chromosomes 3p- and iso10q were recognized in 25% of metaphase spreads. MCF-7 cells (a gift from F. Tamanoi, UCLA) and ZR-75-1 cells (American Type Culture Collection) were maintained in RPMI 1640 supplemented with L-glutamine (GIBCO), 10% Fetal Bovine Serum (Sigma), and 50 unit/mL penicillin/50 unit/mL streptomycin. Before conduction these experiments and at regular intervals, thereafter, all cell lines were tested by the University of North Carolina at Chapel Hill Tissue Culture Facility and were found to be negative for *Mycoplasma* contamination. Cells were maintained at 37 °C and 5% carbon dioxide.

Cytotoxicity Assay

A mitochondrial dye conversion assay (Cell Titer 96, Promega) was used to quantitate cell line responses to chemotherapeutics. Five thousand cells were seeded per well of a 96-well plate. The cells were allowed to adhere overnight and then the media was replaced with fresh media containing a range of drug doses (DOX: 0 – 10 μM, 5FU: 0 – 10 mM). After 36h of drug treatment, 15 μL of tetrazolium dye solution were added, and incubated at 37°C for 1h before adding Cell Titer 96 Stop Solution. Dye conversion products

were allowed to solubilize in a humidified chamber overnight, and absorbance was measured at 570 nm (minus background absorbance at 650 nm).

Estimating the Inhibitory Concentration 50% (IC50)

The IC50 for 36h of treatment for each drug in each cell line was estimated using nonlinear regression (SAS Statistical Software, Cary, NC) and the following relationship:

$$y = \frac{k}{1 + (x / x_0)^b}$$

where y is the absorbance value corrected for media-only wells, and x is the dose (in μM for DOX and in mM for 5FU; Ref. (Van Ewijk and Hoekstra, 1993)). The parameter k represents the value of y (in absorbance units) when x is zero. The IC50 value is represented by x_0, and $-b$ is a unitless scalar representing the slope of the line on logit-log scale. In our experiments, if b is greater than zero, the response is monotonically decreasing.

Collection of mRNA for Microarray Experiments

Cell lines were grown in 150-mm dishes to 70-80% confluence and then were treated for 3h, 12h, 24h, or 36h with DOX (doxorubicin hydrochloride) or 5FU (Sigma) at the IC50 concentration. Cells were harvested by scraping and mRNA was isolated using a Micro-FastTrack kit (Invitrogen). To generate feeding control (sham) mRNA samples for each cell line, cells were treated with media only, in parallel with drug-treated samples. Individual harvests of treated or sham mRNA were not pooled prior to microarray analysis. However, a reference mRNA sample was generated for each of the four cell lines by harvesting untreated mRNA from each cell line at 80% confluence and then pooling four harvests together (i.e. four MCF-7 harvests were pooled and served as the reference mRNA for all MCF-7

experiments), using each cell line as its own reference controlled for baseline differences between the cell lines.

Microarray Experiments

Syntheses of labeled cDNA were performed as described previously (Perou et al., 2000), with reference cDNAs labeled with Cy3-dUTP and treated and sham cDNAs labeled with Cy5-dUTP. Each cDNA sample mix was hybridized overnight to an oligonucleotide microarray created in the University of North Carolina at Chapel Hill Genomics Core Facility [http://genomicscore.unc.edu/]. These microarrays were created by spotting the Compugen Human oligomers library representing 18,861 human genes [http://www.labonweb.com/chips/libraries.html] onto coated microarray slides (Corning no. 40016). All of the microarray raw data tables are available at the UNC Microarray Database [https://genome.unc.edu/], at the supporting website for this chapter [https://genome.unc.edu/pubsup/TOX/], and have been deposited in the Gene Expression Omnibus under the accession number of GSE763 (submitter C. Perou). The direction of gene expression change was verified by real-time reverse transcription PCR for a subset of samples using commercially available primers (Applied Biosystems) for *$p21^{waf1}$, ferredoxin reductase, prostate differentiation factor, inhibitor of DNA binding 3,* and *chitinase 3-like 1*. To normalize the target sample variation, we used the average of three control genes: *splicing factor 3A subunit 1 (SF3A1), pumilio homolog 1 (PUM1),* and *β-actin. SF3A1* and *PUM1* were selected as control genes because they had the lowest variation across the tumor data set presented in Perou *et al.* (Perou et al., 2000). Sham-adjusted real-time PCR values were

regressed on the average of sham-adjusted $\log_2$ (red/green ratio) array values for each gene; the regression yielded a positive slope of 4.3, Pearson $r = 0.75$ (data not shown).

SAM Using Cell Line Data

Genes that were significantly induced or repressed were identified using the Significance Analysis of Microarrays (SAM) package Add-In for Microsoft Excel (Tusher et al., 2001). Before conducting SAM, genes were excluded that did not have a mean signal intensity greater than twice the median background value for both the red and green channel in at least 70% of the experiments. For genes that passed these filtering criteria, the log-base-2 of median red intensity over median green intensity was calculated.

The gene expression changes in the 3h time points were very modest (data not shown); therefore, this time point was excluded from all analyses. To identify genes whose steady-state expression was altered, we combined the 12h, 24h, and 36h time points for each cell line and treatment group into a single class. This eliminated artifacts caused by random temporal variation in steady state RNA levels. Two or three replicate arrays were used for each treatment condition for each cell line.

To identify a general stress response pattern, DOX-treated and 5FU-treated experiments were combined into a single class and compared against sham experiments for each cell line (i.e. MCF-7 DOX-treated and 5FU-treated vs. MCF-7 sham). Missing data were imputed using SAM with 100 permutations and 10 *k*-nearest neighbors. A two-class unpaired SAM analysis was conducted on the imputed data set. The SAM delta values were adjusted to obtain the largest gene list that gave a false discovery rate of less than 5%.

SAM Using Breast Tumor Data

All tumor data were previously published (Sørlie et al., 2001; Sørlie et al., 2003) except for data from five new tumors samples collected after chemotherapy, which are now publicly available at the Stanford Microarray Database [http://genome-www5.stanford.edu/]. This breast tumor dataset encompassed two different cohorts of breast cancer patients, one of which received neoadjuvant doxorubicin and a second of which received neoadjuvant 5FU and Mitomycin C (Geisler et al., 2001; Geisler et al., 2003); in both cohorts, we obtained samples of the tumors before therapy and at the time of surgical resection (after therapy sample). All before and after samples were labeled with Cy5-dUTP, mixed with Cy3-dUTP-labeled Stanford common reference sample, and hybridized to cDNA microarrays produced at Stanford University (Perou et al., 2000). The gene expression patterns of all before samples were compared to the gene expression patterns of all after samples using a two-class, unpaired SAM analysis. A total of 81 before and 50 after samples were assessed, representing all of the tumor subtypes identified in Sørlie *et al*.(Sørlie et al., 2001). Consistent with the *in vitro* data analyses, SAM delta values were adjusted to obtain the largest gene list that gave a false discovery rate of less than 5%.

To study the genes differentially regulated in basal-like or luminal tumor subtypes separately, we also classified each tumor into one of two groups using the intrinsic list of Sørlie *et al*.(Sørlie et al., 2003): one group contained those tumors that represented the luminal epithelium-derived tumors (both Luminal A and B for a total of 51 before and 30 after samples) and a second group representing the basal-like subtype (for a total of 11 before and 10 after). Each of these two groups was then analyzed using a two-class unpaired SAM

analysis; gene expression patterns of before samples were compared to gene expression patterns of after samples and false discovery rates were estimated.

Hierarchical Clustering of Gene Expression Responses

Average linkage hierarchical cluster analysis using Pearson correlation was conducted using the program Cluster, and the data were visualized in Treeview (Eisen et al., 1998; Eisen and Brown, 1999). To visualize the gene expression patterns for the luminal cell lines, the data from the union of the genes identified by SAM for MCF-7 and ZR-75-1 were identified, combined into a non-redundant list, and clustered. These clusters illustrate the fold change relative to control levels for each gene. Following the same procedure, data from the union of the genes sets identified for ME16C and HME-CC were extracted, combined into a non-redundant list, and clustered. Cluster analysis was also performed using the top 100 genes identified by SAM for distinguishing between luminal versus basal-like cell lines responses to DOX-treatment and 5FU-treatment. For all of the clusters, genes were excluded that did not have a mean intensity greater than twice the median background for both the red and green channel in at least 80% of the experiments. For the breast tumor data, similar gene filtering, SAM, and clustering analyses were performed.

Western Blot Analysis

Cells were treated for 36h with DOX or 5FU at the 36h IC50 concentration. Cells were rinsed with PBS and then harvested with M-PER Mammalian Protein Extraction Reagent (Pierce) containing Halt Protease Inhibitor and 5 mM EDTA (Pierce). Protein concentrations were determined using Micro BCA Protein Assay Reagent Kit (Pierce).

Lysates were combined with 2X Laemmli Sample Buffer (BioRad) containing β-mercaptoethanol and were boiled for 5 min. Forty μg of protein were electrophoresed on a 4-20% Tris-HCl Criterion precast gel (Bio-Rad) and transferred to a Hybond-P membrane (Amersham Biosciences) by electroblotting. The blots were probed with antibodies against $p21^{waf1}$ (Neomarkers; Ab-11) and β-actin (Abcam, AC-15). Blots were washed three times with Tris-buffered saline supplemented with 0.1% TWEEN and then were probed with anti-mouse IgG horseradish peroxidase-linked whole antibody from sheep (Amersham). The blots were rewashed and detection was by enhanced chemiluminescence (SuperSignal West Pico Chemiluminescent Substrate, Pierce).

RESULTS

Cell-Type-Specific Transcriptional Responses *In Vitro*

To investigate the response of four distinct cell lines to treatment with chemotherapeutics, we used a mitochondrial dye conversion assay [3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT)] to measure the cytotoxicity of 5FU and DOX after 36h of treatment. We then treated these four cell lines (MCF-7, ZR-75-1, ME16C and HME-CC) with two mechanistically distinct chemotherapeutics (DOX and 5FU) at doses that produced similar levels of toxicity across all four lines (IC50). The IC50 concentrations and their 95% confidence intervals are shown in Table 2.1.

Our experimental design was aimed at identifying the steady-state and cell-type-specific transcriptional response of these cell lines and was not focused on defining chemotherapeutic-specific responses or temporal variation. By combining 12h, 24h, and 36h

Table 2.1. Estimated inhibitory concentration 50% for 5-fluorouracil and doxorubicin based on mitochondrial dye conversion assay

Cell Line	IC50[a]	Treatment Dose
5-Fluorouracil		
MCF-7	0.34 mM (0.13 - 0.55)	0.3 mM
ZR-75-1	3.3 mM (2.8 - 3.7)	3.0 mM
ME16C	0.064 mM (0.055 - 0.074)	0.06 mM
HME-CC	0.011 mM (0.009 - 0.013)	0.01 mM
Doxorubicin		
MCF-7	0.86 μM (0.74 - 0.97)	0.9 μM
ZR-75-1	0.43 μM (0.37 - 0.50)	0.4 μM
ME16C	0.52 μM (0.49 - 0.54)	0.5 μM
HME-CC	0.16 μM (0.14 - 0.18)	0.2 μM

[a]Values in parentheses represent 95% confidence intervals.

Table 2.2. Number of oligonucleotides significantly altered by treatment with chemotherapeutic as determined by Significance Analysis of Microarrays

Sample	No. of Oligos	No. of False Significant[a]
MCF7 treated	998	48.5 (4.9)
ZR-75-1 treated	783	38.5 (4.9)
ME16C treated	84	3.3 (3.9)
HME-CC treated	84	3.0 (3.4)
All tumors	28	0.7 (2.5)
Luminal tumors	15	0.8 (5.3)
Basal-like tumors	10	2 (20)

[a]Percent false significant indicated in parentheses.

treated experiments into a single class for supervised analyses, we avoided temporal artifacts and identified only those gene expression changes that were consistent over time.

Chemotherapeutic-Induced Gene Expression Patterns in Luminal Cell Lines.

Differences between basal-like and luminal cell lines responding to treatment were immediately evident given the absolute number of genes whose expression was altered when treated experiments were compared to sham experiments (Table 2.2). In each luminal cell line, approximately 10-fold more genes were altered in response to drug. To visualize these expression changes, we combined the SAM supervised lists for the two luminal cell lines and performed a hierarchical clustering analysis (Figure 2.1). Each cell line had a unique expression response to chemotherapy that was distinct enough to cause the two treated luminal lines to cluster into different dendrogram branches (Figure 2.1B).

Common features dominated the overall expression patterns in the two luminal cell lines despite some cell line specific responses. For example, a cluster of genes that reflect cell proliferation *in vitro* and *in vivo* (Perou et al., 1999; Perou et al., 2000; Ross et al., 2000; Sørlie et al., 2001; Whitfield et al., 2002) was identified (Figure 2.1C). These genes had slightly increased expression in the sham experiments because of feeding, but had greatly diminished expression during drug treatment. This cluster included well characterized cell cycle regulators (Whitfield et al., 2002) such as *cyclin A2, cyclin B1, cell division cycle 2*, and many genes involved in specific phases of the cell cycle such as *Ki-67, ribonucleotide reductase M2, polo-like kinase*, and *topoisomerase IIA*. This cluster also included *pituitary tumor-transforming 1*, a gene that is overexpressed in many cancers, is tumorigenic *in vivo*, and has been shown to bind p53 (Bernal et al., 2002). The gene product of *serine/threonine*

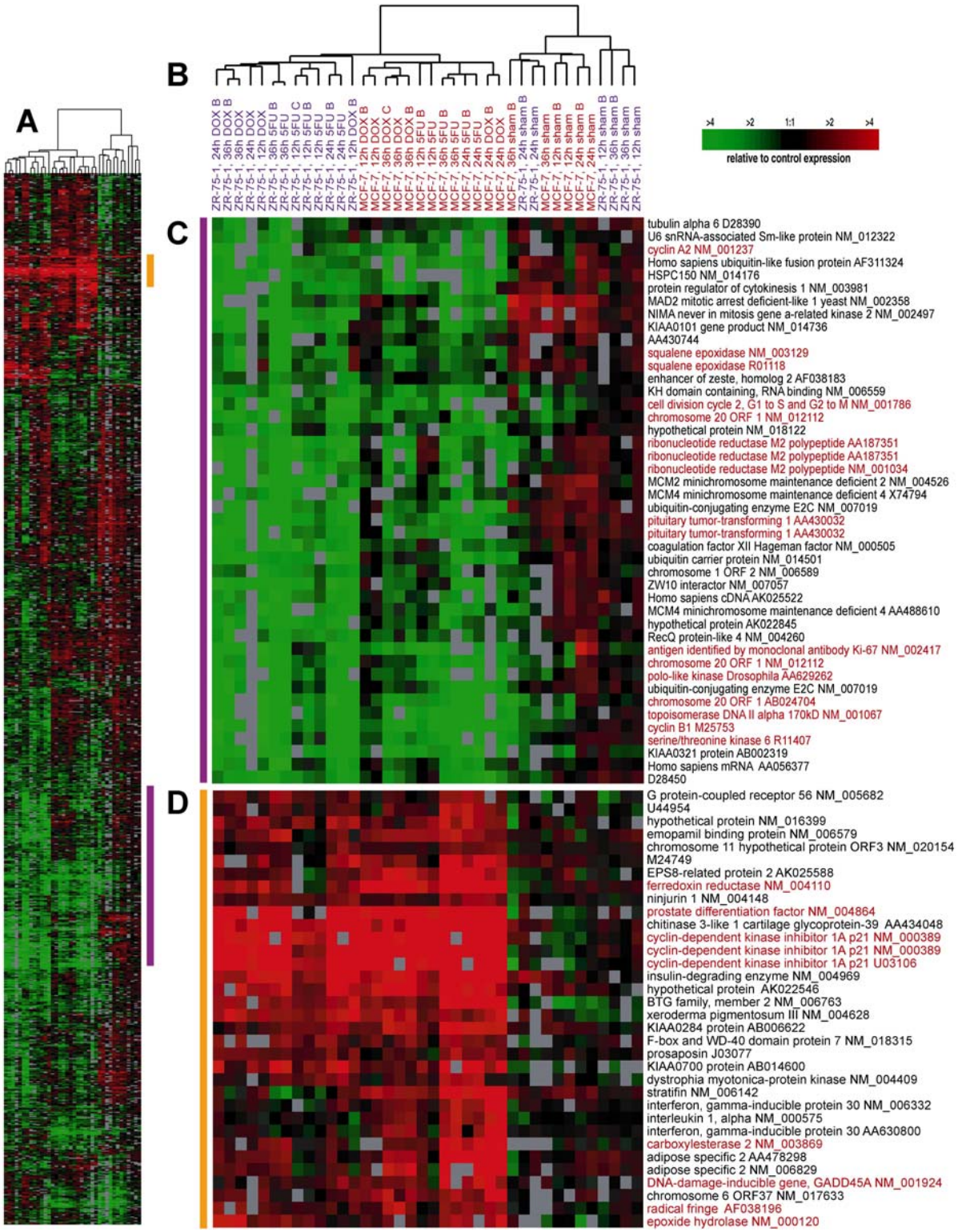
A
B
C
D
ZR-75-1, 24h DOX B
ZR-75-1, 36h DOX B
ZR-75-1, 36h DOX
ZR-75-1, 24h DOX
ZR-75-1, 12h DOX
ZR-75-1, 36h 5FU B
ZR-75-1, 36h 5FU
ZR-75-1, 12h 5FU C
ZR-75-1, 12h 5FU B
ZR-75-1, 12h 5FU
ZR-75-1, 24h 5FU B
ZR-75-1, 24h 5FU
ZR-75-1, 12h DOX B
MCF-7, 12h DOX B
MCF-7, 12h DOX
MCF-7, 36h DOX C
MCF-7, 36h DOX
MCF-7, 36h DOX B
MCF-7, 12h 5FU B
MCF-7, 12h 5FU
MCF-7, 36h 5FU B
MCF-7, 36h 5FU
MCF-7, 24h 5FU B
MCF-7, 24h 5FU
MCF-7, 24h DOX B
MCF-7, 24h DOX
MCF-7, 36h sham B
ZR-75-1, 24h sham B
ZR-75-1, 24h sham
MCF-7, 36h sham
MCF-7, 12h sham B
MCF-7, 12h sham
MCF-7, 24h sham B
MCF-7, 24h sham
ZR-75-1, 12h sham B
ZR-75-1, 36h sham B
ZR-75-1, 36h sham
ZR-75-1, 12h sham
>4
>2
1:1
>2
>4
relative to control expression
tubulin alpha 6 D28390
U6 snRNA-associated Sm-like protein NM_012322
cyclin A2 NM_001237
Homo sapiens ubiquitin-like fusion protein AF311324
HSPC150 NM_014176
protein regulator of cytokinesis 1 NM_003981
MAD2 mitotic arrest deficient-like 1 yeast NM_002358
NIMA never in mitosis gene a-related kinase 2 NM_002497
KIAA0101 gene product NM_014736
AA430744
squalene epoxidase NM_003129
squalene epoxidase R01118
enhancer of zeste, homolog 2 AF038183
KH domain containing, RNA binding NM_006559
cell division cycle 2, G1 to S and G2 to M NM_001786
chromosome 20 ORF 1 NM_012112
hypothetical protein NM_018122
ribonucleotide reductase M2 polypeptide AA187351
ribonucleotide reductase M2 polypeptide AA187351
ribonucleotide reductase M2 polypeptide NM_001034
MCM2 minichromosome maintenance deficient 2 NM_004526
MCM4 minichromosome maintenance deficient 4 X74794
ubiquitin-conjugating enzyme E2C NM_007019
pituitary tumor-transforming 1 AA430032
pituitary tumor-transforming 1 AA430032
coagulation factor XII Hageman factor NM_000505
ubiquitin carrier protein NM_014501
chromosome 1 ORF 2 NM_006589
ZW10 interactor NM_007057
Homo sapiens cDNA AK025522
MCM4 minichromosome maintenance deficient 4 AA488610
hypothetical protein AK022845
RecQ protein-like 4 NM_004260
antigen identified by monoclonal antibody Ki-67 NM_002417
chromosome 20 ORF 1 NM_012112
polo-like kinase Drosophila AA629262
ubiquitin-conjugating enzyme E2C NM_007019
chromosome 20 ORF 1 AB024704
topoisomerase DNA II alpha 170kD NM_001067
cyclin B1 M25753
serine/threonine kinase 6 R11407
KIAA0321 protein AB002319
Homo sapiens mRNA AA056377
D28450
G protein-coupled receptor 56 NM_005682
U44954
hypothetical protein NM_016399
emopamil binding protein NM_006579
chromosome 11 hypothetical protein ORF3 NM_020154
M24749
EPS8-related protein 2 AK025588
ferredoxin reductase NM_004110
ninjurin 1 NM_004148
prostate differentiation factor NM_004864
chitinase 3-like 1 cartilage glycoprotein-39 AA434048
cyclin-dependent kinase inhibitor 1A p21 NM_000389
cyclin-dependent kinase inhibitor 1A p21 NM_000389
cyclin-dependent kinase inhibitor 1A p21 U03106
insulin-degrading enzyme NM_004969
hypothetical protein AK022546
BTG family, member 2 NM_006763
xeroderma pigmentosum III NM_004628
KIAA0284 protein AB006622
F-box and WD-40 domain protein 7 NM_018315
prosaposin J03077
KIAA0700 protein AB014600
dystrophia myotonica-protein kinase NM_004409
stratifin NM_006142
interferon, gamma-inducible protein 30 NM_006332
interleukin 1, alpha NM_000575
interferon, gamma-inducible protein 30 AA630800
carboxylesterase 2 NM_003869
adipose specific 2 AA478298
adipose specific 2 NM_006829
DNA-damage-inducible gene, GADD45A NM_001924
chromosome 6 ORF37 NM_017633
radical fringe AF038196
epoxide hydrolase NM_000120

Figure 2.1

Figure 2.1. Gene expression pattern for genes significantly altered in MCF-7 and ZR-75-1 cell lines responding to chemotherapeutics. Cluster analysis was conducted using 26 treated and 12 sham experiments. Data from the union of the genes identified by SAM for MCF-7 and ZR-75-1 were identified, combined into a non-redundant list, and the compressed cluster diagram is shown in *A*. Colored bars in *A* illustrate the location of clusters *C* and *D*. The dendrogram in *B* shows that experiments were divided into two primary branches (treated and sham), and the treated branch was subdivided into two secondary branches: MCF-7 experiments (in red) and ZR-75-1 experiments (in blue). A large cluster enriched for genes involved or correlated with proliferation (*C*) and a cluster enriched for genes involved in responding to stress or DNA-damage (*D*) are shown. Genes discussed in the text are highlighted in red.

kinase 6 (STK6) is also present and has cell-cycle dependent expression, with maximum expression in G2-M (Kimura et al., 1997); in addition, STK6 has been shown to bind *chromosome 20 open reading frame 1* (Kufer et al., 2002), which is also repressed and in this cluster. *Squalene epoxidase* was down-regulated in the luminal cell lines and is a gene that was differentially expressed between luminal and basal-like tumors *in vivo* (Sørlie et al., 2001).

A large cluster of genes that include DNA-damage and stress response genes was up-regulated in response to treatment in the luminal lines (Figure 2.1D). $p21^{waf1}$ and the DNA-damage response gene *GADD45* were induced strongly in both lines. Also present in this cluster were a number of genes involved in xenobiotic metabolism including *carboyxlesterase 2, epoxide hydrolase*, and *ferredoxin reductase*. The latter two of these genes, along with $p21^{waf1}$ and *GADD45*, are all known to be p53-regulated (Liu and Chen, 2002; Park et al., 2002). Induction of xenobiotic metabolism genes may represent a stereotyped adaptive response of the cell to DNA damage.

Chemotherapeutic-Induced Gene Expression Patterns in Basal-Like Cell Lines

A much smaller list of genes showed significantly altered expression in the ME16C or HME-CC basal-like cell lines (Figure 2.2). Using the combined list of genes that were significantly altered in either basal-like cell line in a hierarchical clustering analysis showed that the basal-like lines did not cluster as distinctly as the luminal cell lines (Figure 2.2B). Within the treated branch, some time points for the HME-CC line clustered on separate branches, but the drug-treated ME16C experiments all grouped together. This suggests that the changes induced in basal-like cells treated with chemotherapeutics were subject to more

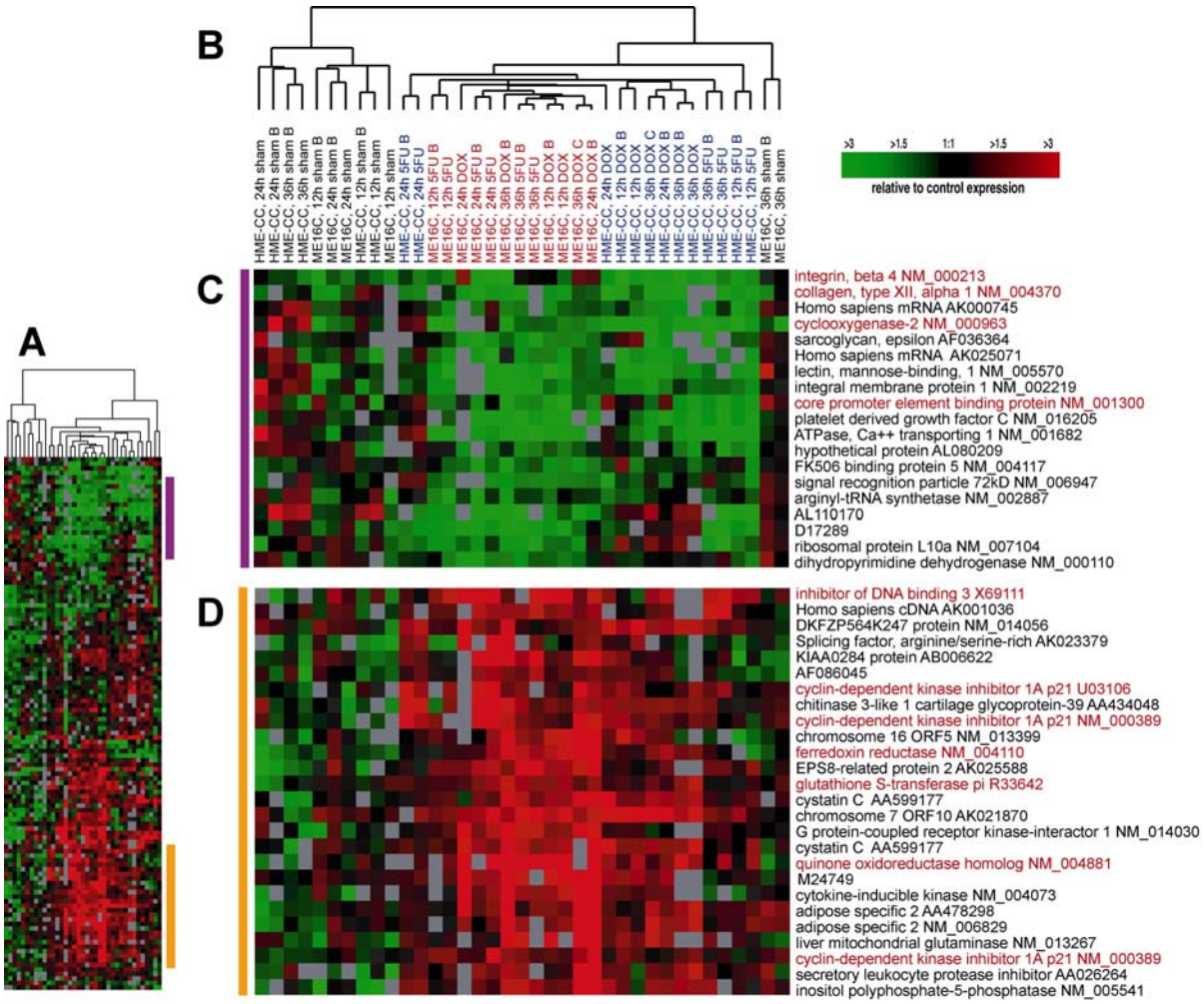

Figure 2.2. Gene expression pattern for genes significantly altered by chemotherapeutics in ME16C and HME-CC cell lines. Cluster analysis was conducted using 25 treated and 12 sham experiments. Data from the union of the genes identified by SAM for ME16C or HME-CC were identified, combined into a non-redundant list, and the scaled down cluster diagram is shown in *A*. Colored bars illustrate the location of clusters *C* and *D*. The dendrogram in *B* shows that the experiments were divided into two primary branches, one consisting primarily of treated experiments (ME16C in red, HME-CC in blue) and one consisting exclusively of shams. A large cluster enriched for genes involved in differentiation and possibly correlated with proliferation (*C*) and a cluster enriched for genes involved in responding to stress or DNA-damage (*D*) are shown. Genes discussed in the text are highlighted in red.

temporal variation. The changes also appeared more subtle; strong signatures like those observed in the luminal cell lines were not nearly as evident in these basal-like cell lines. We identified a small cluster of genes that was slightly induced in the sham experiments, but that was down-regulated in the treated experiments (Figure 2.2C). Many of these genes are involved in cellular differentiation including *integrin-β4*, *collagen type XII α1, COX2* and *core promoter element binding protein*. A proliferation signature (similar to Figure 2.1C) was not identified in the treated basal-like lines. However, a set of genes involved in the DNA damage and/or stress response was identified (Figure 2.2D) and similar to the luminal cell lines (Figure 2.1D), $p21^{waf1}$ was induced, although less dramatically. Several xenobiotic metabolism genes were also up-regulated including the p53-regulated genes *ferredoxin reductase* and *quinone oxidoreductase homolog*, as well as *glutathione-S-transferase pi (GST-π)*. *Inhibitor of DNA binding 3*, an inhibitor of differentiation (Sikder et al., 2003), was also induced in both basal-like cell lines.

Comparison of Basal-Like versus Luminal Cell Lines

In the analyses above, we identified genes that differed between shams and treated samples on a cell line by cell line basis. To assess differences between basal-like and luminal cell lines, we first compared the lists of chemotherapeutic-induced genes for the luminal (1000 genes) and the basal-like (100 genes) cell lines. There were 42 genes on both lists, but only two genes (*chitinase 3-like 1* and $p21^{waf1}$) were up-regulated in all four lines (no genes down-regulated). We then used SAM to directly identify the set of genes that distinguished the treated luminal lines from the treated basal-like lines. With a 5% false discovery rate, 920 genes were statistically different. The top 100 distinguishing genes were

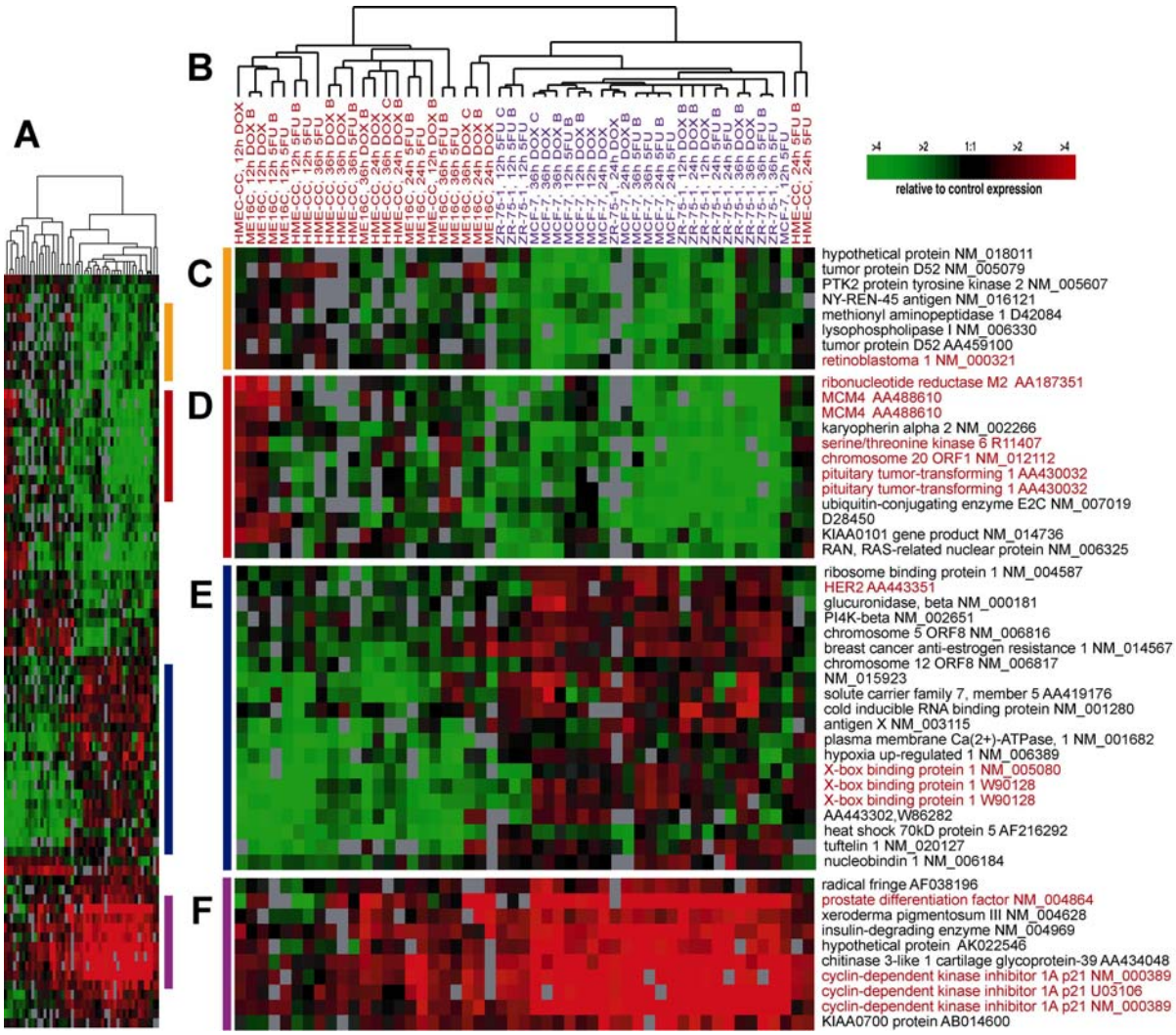

Figure 2.3. Gene expression pattern of top 100 genes that distinguished between basal-like and luminal chemotherapeutic-treated cell lines. Cluster analysis was conducted on the 51 chemotherapeutic-treated MCF-7, ZR-75-1, HME-CC and ME16C experiments. The scaled down cluster diagram is shown in *A*. Colored bars illustrate the location of clusters *C, D, E*, and *F*. The dendrogram in *B* shows that the 51 experiments were divided into two dendrogram branches based on gene expression. The luminal and basal-like cell lines are shown in blue and red, respectively. Clusters of genes are shown whose expression were more drastically down-regulated (*C* and *D*) or up-regulated (*E* and *F*) in luminal cell lines compared to basal-like cell lines. Genes discussed in the text are highlighted in red.

used to cluster the experiments (Figure 2.3). The grouping of the cell lines identified two primary dendrogram branches (Figure 2.3B), one representing only basal-like cell lines and one representing predominantly luminal cell lines. A few of the late time points for the basal-like cell lines fell within the luminal branch but remained distinct on their own secondary branches due to their unique expression profiles. This finding again illustrates the temporal variability across the basal-like cell lines time points.

Consistent with our previous analysis, the luminal cell lines showed a greatly reduced proliferation signature, which was relatively unchanged in the basal-like cell lines (Figures 2.3C and 2.3D). This gene set included *retinoblastoma 1, ribonucleotide reductase M2, MCM4, chromosome 20 open reading frame 1* and *pituitary tumor-transforming 1*, all of which regulate cell proliferation or have cell cycle-dependent expression (Whitfield et al., 2002). A cluster of genes whose expression was induced in luminal lines and repressed in basal-like lines is shown in Figure 2.3E, wheras the gene set in Figure 2.3F was induced in both cell types, but was more highly induced in luminal cells versus basal-like cells. Among these genes was *X-box binding protein 1 (XBP1)*, a gene whose expression was previously shown to be highly expressed in luminal tumors *in vivo* (Sørlie et al., 2001). *XBP1* is a transcription factor involved in mediating the unfolded protein response (Ma and Hendershot, 2001), which may represent a stress response that is more prominent in secretory luminal cells. *HER2* also appeared to be induced more distinctly in luminal cells treated with chemotherapeutics (Figure 2.3E). *HER2* has been extensively studied in breast cancer and it has been shown that MCF-7 cells that over-express *HER2* retained their proliferative advantage following DOX treatment in a human breast cancer xenograft model (Pegram et al., 1997). Figure 2.3F illustrates that the DNA damage response was much more dramatic in

the luminal cell lines, with expression of *p21*^{waf1} and *prostate differentiation factor* being highly up-regulated in luminal cells and less dramatically induced in basal-like cell lines. We also confirmed the cell-type-specific differences in basal-like versus luminal induction of p21^{waf1} on the protein level by Western blot (Figure 2.4).

In Vivo Responses to Chemotherapeutics

We have previously profiled 115 breast tumors and have identified clinically distinct subtypes using patterns of gene expression (Perou et al., 2000; Sørlie et al., 2001; Sørlie et al., 2003). Tumor biopsies were sampled before chemotherapy, and for 46 of these tumors, tumor biopsies were also sampled after chemotherapy (Geisler et al., 2001; Geisler et al., 2003). To allow comparisons with our *in vitro* work, we conducted a supervised analysis using SAM to identify gene expression differences between before and after chemotherapy samples. For the first analysis, we disregarded tumor subtype differences and treatment differences (patients were treated with either DOX or 5FU/mitomycin) and looked for consistent differences between 81 before samples and 50 after samples. All breast tumor subtypes identified in Sørlie *et al.* (Sørlie et al., 2003) were represented. The list of genes that differed between the before and after samples is shown in a cluster diagram (Figure 2.5A) in which the samples are arranged and ordered by tumor subtype as defined in Sørlie *et al.* (Sørlie et al., 2003); genes were clustered and all fold-changes are displayed relative to the median gene expression level. A total of 28 cDNA clones representing 23 genes were more highly expressed in the after samples relative to the before samples (no genes were significantly lower in the after samples). These findings agree (at least 13 genes in common) with a similar analysis performed on a subset of this data (Korn et al., 2002). Among these 23

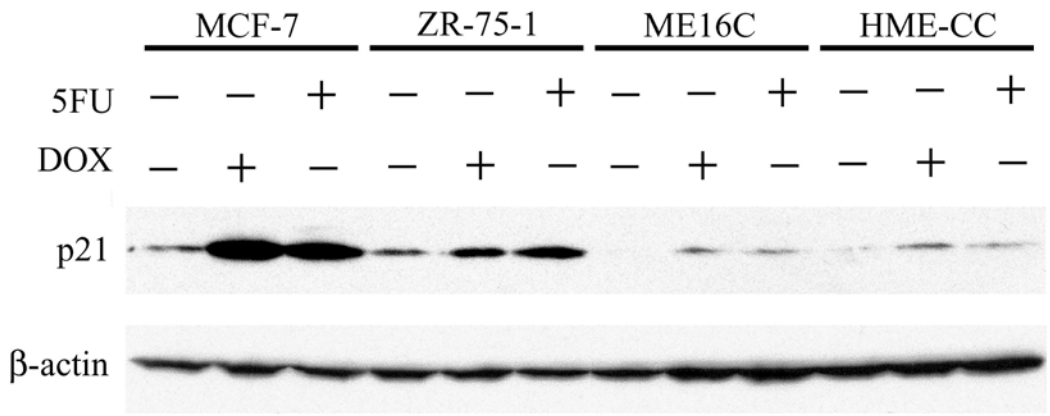

Figure 2.4. Protein levels of p21^{waf1} in chemotherapeutic-treated basal-like and luminal cell lines. Cell lines were treated with an IC50 dose of DOX or 5FU, and lysates were collected at 36h. p21^{waf1} levels were induced in all chemotherapeutic-treated samples relative to sham samples. β-actin was assayed as a loading control.

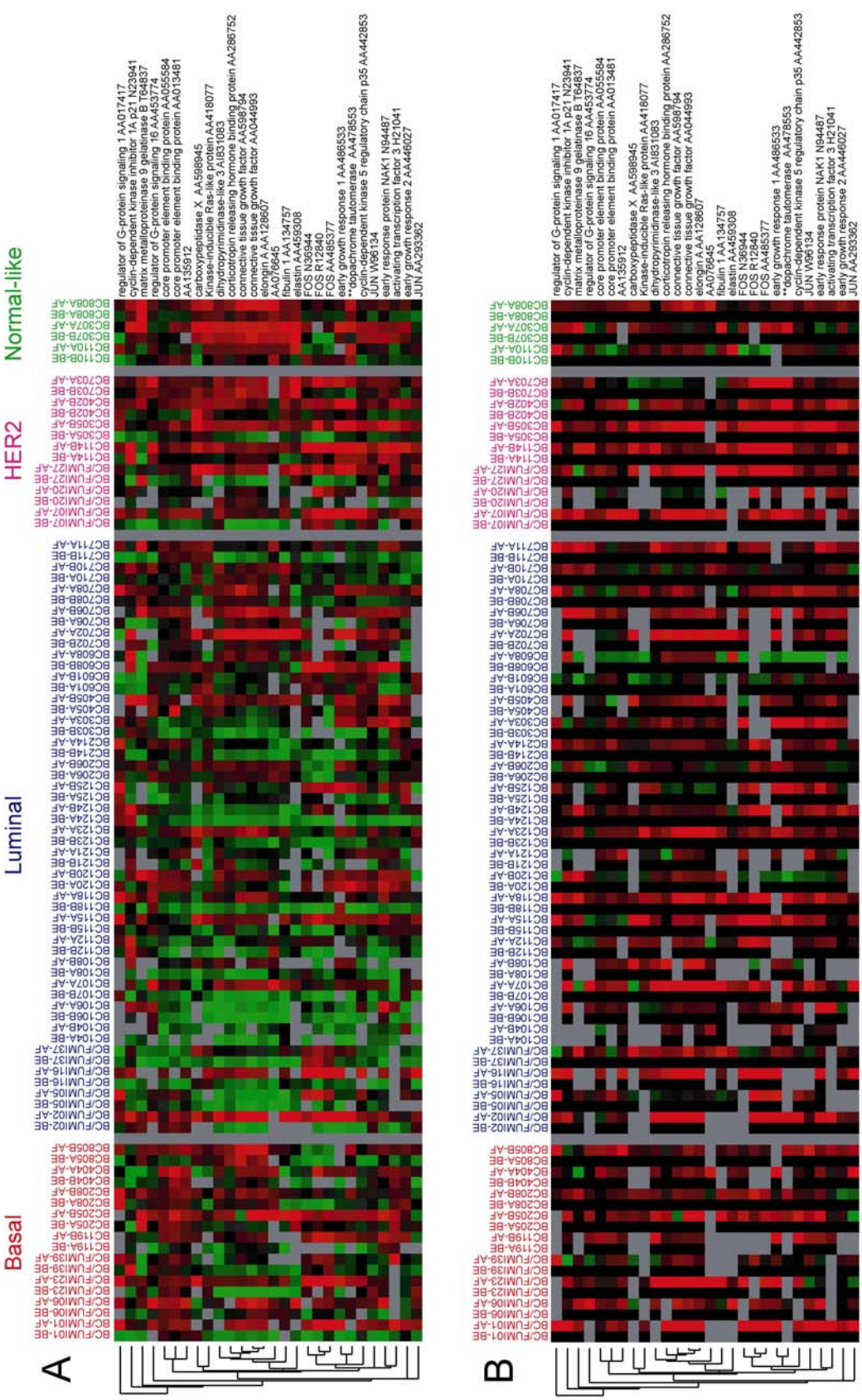
A
Basal
Luminal
HER2
Normal-like
regulator of G-protein signaling 1 AA017417
cyclin-dependent kinase inhibitor 1A p21 N23941
matrix metalloproteinase 9 gelatinase B T64837
regulator of G-protein signaling 16 AA453774
core promoter element binding protein AA055584
core promoter element binding protein AA013481
AA135912
carboxypeptidase X AA598945
Kinase-inducible Ras-like protein AA418077
dihydropyrimidinase-like 3 AI831083
corticotropin releasing hormone binding protein AA286752
connective tissue growth factor AA598794
connective tissue growth factor AA044993
elongin A AA128607
AA076645
fibulin 1 AA134757
elastin AA459308
FOS N36944
FOS R12840
FOS AA485377
early growth response 1 AA486533
*dopachrome tautomerase AA478553
cyclin-dependent kinase 5 regulatory chain p35 AA442853
JUN W96134
early response protein NAK1 N94487
activating transcription factor 3 H21041
early growth response 2 AA446027
JUN AA293362
B
regulator of G-protein signaling 1 AA017417
cyclin-dependent kinase inhibitor 1A p21 N23941
matrix metalloproteinase 9 gelatinase B T64837
regulator of G-protein signaling 16 AA453774
core promoter element binding protein AA055584
core promoter element binding protein AA013481
AA135912
carboxypeptidase X AA598945
Kinase-inducible Ras-like protein AA418077
dihydropyrimidinase-like 3 AI831083
corticotropin releasing hormone binding protein AA286752
connective tissue growth factor AA598794
connective tissue growth factor AA044993
elongin A AA128607
AA076645
fibulin 1 AA134757
elastin AA459308
FOS N36944
FOS R12840
FOS AA485377
early growth response 1 AA486533
*dopachrome tautomerase AA478553
cyclin-dependent kinase 5 regulatory chain p35 AA442853
JUN W96134
early response protein NAK1 N94487
activating transcription factor 3 H21041
early growth response 2 AA446027
JUN AA293362

Figure 2.5

Figure 2.5. Gene expression pattern for genes altered by chemotherapeutics in tumors. *A,* Median-centered expression data using the gene set determined by SAM to be significantly changed in the before versus the after DOX-treated and 5FU-treated tumors. Tumor pairs from all of the tumor subtypes characterized in Sørlie *et al.* (Sørlie et al., 2001) are included; tumor samples in red are basal-like; tumor samples in blue are luminal (subtype A and B); tumor samples in pink are ERBB2/HER2 positive; tumor samples in green are normal-like. *B*, Each after sample in 5A was normalized to the expression value for its corresponding before sample. A gray square was assigned to both samples if either the before or after sample had missing data. *A* and *B* exclude all tumor samples for which a complete before and after set was unavailable.

genes were the AP-1 co-activators *FOS* and *JUN, p21^{waf1}*, and a number of other genes involved in wound healing including *connective tissue growth factor* and *matrix metalloproteinase 9*. The genes are listed in Appendix IIA.

In Figure 2.5B, we normalized the expression ratio in each after sample to its paired before sample (displayed in black) on a gene-by-gene basis. This allowed visualization of the changes caused by chemotherapy in each patient. These changes are difficult to discern in Figure 2.5A due to the diversity of initial expression values in the before samples. Relative to the paired before samples, nearly all of the after samples in all four tumor subtypes showed induced expression of these genes, despite their diverse expression ranges in Figure 2.5A. A few of the tumors had anomalous behavior, underscoring the individuality of tumor responses even within a subtype.

Based upon our cell line data, we hypothesized that there might also be tumor subtype-specific responses, so we conducted analyses on the before samples versus the after samples for the basal-like and luminal subtypes separately. Using 81 luminal tumor samples and SAM analysis, we identified 14 genes that were changed in expression after treatment (Table 2.3 and Appendix IIB). Using 21 basal-like tumor samples, we identified nine genes that were induced after treatment (Table 2.3 and Appendix IIC). In this analysis, there was a five-gene overlap between the basal-like and luminal gene lists. A number of genes that were seen in the combined analysis were significantly altered in only one of the subtypes. For example, *p21^{waf1}* was only present on the luminal list and *core promoter element binding protein (COPEB)* was only present on the basal-like list. Next, we compared each cell type's *in vivo* list with its corresponding *in vitro* list and identified four genes that were altered in luminal tumors and luminal cell lines (*p21^{waf1}, elongin A, prostate differentiation factor and*

Table 2.3. Genes altered by chemotherapy in luminal and basal-like breast tumor subtypes.

Luminal Tumors	**Basal-Like Tumors**
connective tissue growth factor AA598794	connective tissue growth factor AA598794
connective tissue growth factor AA044993	connective tissue growth factor AA044993
Early growth response 1 AA486533	early growth response 1 AA486533
Early response protein NAK1 N94487	early response protein NAK1 N94487
§elongin A AA128607	elongin A AA128607
FOS N36944	FOS R12840
Corticotropin releasing hormone binding protein AA286752	§core promoter element binding protein AA013481
cyclin-dependent kinase 5, regulatory subunit 1 (p35) AA442853	dermatan sulfate proteoglycan 3 AA131238
§cyclin-dependent kinase inhibitor 1A, $p21^{waf1}$ N23941	Homo sapiens mRNA AA135912
dihydropyrimidinase-like 3 AI831083	RAB21 AA076645
**dopachrome tautomerase AA478553	
Kinase-inducible Ras-like protein AA418077	
§**prostate differentiation factor N26311	
spondin 1 H09099	
§thrombospondin 1 AA464532	

§ This gene was also altered by treatment in the corresponding cell line experiments.
**A potentially chimeric cDNA clone that maps to two different Unigene entries.

thrombospondin 1). *COPEB* was the only gene that was significantly altered in both the basal-like tumors and cell lines.

Finally, to identify additional similarities between the cell line and tumor data sets, we used the SAM-generated luminal and basal-like cell line-derived gene lists in a clustering analysis of the tumor samples. Basal-like and luminal gene expression signatures identified in the cell lines also appeared differentially expressed in tumor subtypes (data not shown). For example, when the basal-like cell line list was used to cluster all tumor samples, $p21^{waf1}$ and *MDM2* clustered together and showed higher expression in the luminal tumors, while *COPEB* and *GST-pi* showed higher expression in the basal-like tumors. When the luminal cell line list was used to cluster all tumor samples, subtype specific responses were also evident; for example, the basal-like tumors showed high expression of the proliferation signature both before and after chemotherapy. This is consistent with the *in vitro* findings because the proliferation signature was unchanged in the basal-like cell lines after chemotherapy treatment.

DISCUSSION

The mammary gland contains a heterogeneous population of epithelial cells in different stages of differentiation. Basal-like and luminal epithelium represent two cell populations that are thought to arise from a common progenitor, but they each express unique markers and perform unique functions (Pechoux et al., 1999; Smith and Chepko, 2001). Luminal epithelia are widely believed to give rise to the majority of breast cancers, but there is evidence that up to 15% of breast cancers show some characteristics of basal-like epithelium (Perou et al., 2000; Sørlie et al., 2001). MCF-7 cells and HME cell lines have

been extensively studied as models of breast cancer; however, they represent different types of breast cancers. MCF-7 and ZR-75-1 cells (data not shown) have expression similarities with ERα-positive breast tumors, while HME lines (finite life span or immortalized) have expression similarities with basal-like breast tumors (Ross and Perou, 2001).

In our model of breast cancer, basal-like and luminal epithelial cells have unique transcriptional responses to the chemotherapeutics DOX and 5FU. The two luminal cell lines showed similar response patterns to one another including the strong induction of DNA damage/stress response genes, notably *p21*^{waf1} (Figures 2.1D and 2.3F). The basal-like cell lines showed a much less dramatic induction of *p21*^{waf1} (Figure 2.2D and 2.3F). All four of our cell lines are wildtype for p53 by sequence analysis and express p53 protein (data not shown), so the differences in *p21*^{waf1} expression cannot be attributed to differences in p53 status. p21^{waf1} is involved in the G1 checkpoint response, and others have reported an impaired G1 checkpoint in HME cell lines (Meyer et al., 1999). Consistent with a strong cell cycle checkpoint in the luminal cell lines, MCF-7 and ZR-75-1 cells also repressed a large set of proliferation genes (Figure 2.1C). This suggests that their G1 checkpoint in response to DNA damage is intact (Whitfield et al., 2002). The two basal-like cell lines did not repress the proliferation signature, but they did down-regulate genes involved in differentiation (Figure 2.2C).

The basal-like cell lines we used for this study were hTERT immortalized HME cells, whereas the luminal cell lines were derived from human tumors. Telomerase expression is a hallmark of breast cancer (Herbert et al., 2001), but increased telomerase expression is one of many changes that are observed as cells progress toward a malignant state (Hanahan and Weinberg, 2000). Although breast tumor derived cell lines of luminal origin are widely

studied, analogous lines of basal-like origin have not yet been identified. We acknowledge that comparison of breast cancer lines versus immortalized breast lines represents a starting point for investigations of these cell types. Future comparisons using additional cell lines, and preferably cancer cell lines of basal-like origin, may yield more data of greater significance. We note that some of the expression differences observed between the basal-like and luminal cell lines could be due to differences in tumorigenicity. However, we found that our cell lines recapitulated some of the cell type differences seen *in vivo* in response to these same agents (Table 2.3). The overlap observed between the tumors and cell lines is significant, especially considering three differences between these data sets: (*a*) the tumor data was acquired using cDNA microarrays and a common reference sample whereas the cell lines were assayed using 60-mer oligonucleotide arrays and a cell-line specific reference (untreated pooled reference); (*b*) the cell lines were all *p53* wild type, while approximately 40% of the tumors were p53 mutant. Thus, the *in vivo* analysis is more likely to have excluded some p53-dependent responses to chemotherapy; and (*c*) the tumors represent a heterogeneous cell population and the cell lines represent only a single cell type.

A strength of tumor profiling studies is that they capture the heterogeneity of tumors in their natural environment. However, this heterogeneity makes it difficult to study the chemotherapy responses of specific cell types. The role of each cell type in a tumor can begin to be dissected using cell-line models, preferably with multiple cell lines representing each subtype. Cell lines are as unique as the tumors from which they were derived, but common response patterns can only become identifiable when looking at multiple cell lines in concert. This was illustrated in a recent study of 60 cell lines and 60,000 compounds (Scherf et al., 2000; Ross and Perou, 2001) in which relationships between sets of cell lines, sets of genes,

and toxicant sensitivity were identified. In the work presented here, we used four cell lines with two cell lines representing each of two tumor subtypes. Characterizing common responses and inter-individual variation in these cell lines will help to identify those responses that are stereotypical for each subtype.

Recent studies have demonstrated that DNA-damaging agents induce generic stress responses. In 2000, Gasch *et al.* (Gasch et al., 2000) showed that yeast displayed a stereotypic pattern of gene expression when exposed to a wide range of stresses including heat shock, growth factor deprivation, and treatment with hydrogen peroxide. These authors termed the stereotypic response the "environmental stress response" (ESR). The ESR included repression of growth-related genes and genes encoding ribosomal proteins and induction of genes involved in DNA damage response and metabolism. These results are in agreement with our finding that a major response to treatment included repression of genes involved in cell growth and induction of DNA damage response genes. Our work with breast cell lines corroborates other recent human cell line studies that have demonstrated common stress responses following DNA damaging treatments (Morgan et al., 2002; Park et al., 2002; Sesto et al., 2002; Weigel et al., 2002; Heinloth et al., 2003). In this chapter, we have demonstrated that some of the changes seen *in vitro* were also observed *in vivo*.

Finally, we note that DOX and 5FU have distinct mechanisms of action (Gewirtz, 1999; Longley et al., 2003). For example, DOX is thought to target topoisomerase IIA blocking the G2 to M transition and 5FU targets thymidylate synthase blocking S-phase progress. In our experiments with luminal cell lines, both drugs affected gene expression in all phases of the cell cycle (Figure 2.1C). These cell cycle genes serve as proliferation markers and are not specific to a single mode of action. The specific mechanisms of action of

DOX and 5FU may be evident in a subset of genes expressed in our experiments and subsequent analyses will attempt to identify this gene set. However, to fully validate toxicant-specific gene sets, it must also be demonstrated that the gene set predicts mode of action for independent data sets on mechanistically similar drugs. Our primary objective for this work was to understand how cell types differed in their stress response patterns, which are the dominant gene expression responses to DNA damage. The identification of cell-type specific stress responses *in vitro* and *in vivo* has implications for understanding the biological response to therapy.

ACKNOWLEDGMENTS

We thank Elizabeth Livanos of the University of North Carolina at Chapel Hill Chromosome Imaging Core Facility for karyotyping the HME cell lines. We acknowledge J. I. Herschkowitz for helpful comments on this manuscript. This work was supported by the National Institute of Environmental Health Sciences Toxicogenomics Research Consortium grant (5-U19-ES11391-03). M.A.T. was supported by the National Institute of Environmental Health Sciences through an Individual National Research Service Award (NRSA) 5F32ES012374 and an Institutional NRSA in Environmental Pathology 2T32ES07017. K.A.H. was supported by the National Institutes of Health Institutional NRSA in Genetics T32GM07092.

REFERENCES

Aas, T., Borresen, A. L., Geisler, S., Smith-Sorensen, B., Johnsen, H., Varhaug, J. E., Akslen, L. A. and Lonning, P. E. (1996). Specific P53 mutations are associated with de novo resistance to doxorubicin in breast cancer patients. Nat Med 2(7): 811-814.

Bernal, J. A., Luna, R., Espina, A., Lazaro, I., Ramos-Morales, F., Romero, F., Arias, C., Silva, A., Tortolero, M. and Pintor-Toro, J. A. (2002). Human securin interacts with p53 and modulates p53-mediated transcriptional activity and apoptosis. Nat Genet 32(2): 306-311.

Eisen, M. B. and Brown, P. O. (1999). DNA arrays for analysis of gene expression. Methods Enzymol 303: 179-205.

Eisen, M. B., Spellman, P. T., Brown, P. O. and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 95(25): 14863-14868.

Gasch, A. P., Spellman, P. T., Kao, C. M., Carmel-Harel, O., Eisen, M. B., Storz, G., Botstein, D. and Brown, P. O. (2000). Genomic expression programs in the response of yeast cells to environmental changes [In Process Citation]. Mol Biol Cell 11(12): 4241-4257.

Geisler, S., Borresen-Dale, A. L., Johnsen, H., Aas, T., Geisler, J., Akslen, L. A., Anker, G. and Lonning, P. E. (2003). TP53 gene mutations predict the response to neoadjuvant treatment with 5-fluorouracil and mitomycin in locally advanced breast cancer. Clin Cancer Res 9(15): 5582-5588.

Geisler, S., Lonning, P. E., Aas, T., Johnsen, H., Fluge, O., Haugen, D. F., Lillehaug, J. R., Akslen, L. A. and Borresen-Dale, A. L. (2001). 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(6): 2505-2512.

Gewirtz, D. A. (1999). A critical evaluation of the mechanisms of action proposed for the antitumor effects of the anthracycline antibiotics adriamycin and daunorubicin. Biochem Pharmacol 57(7): 727-741.

Hanahan, D. and Weinberg, R. A. (2000). The hallmarks of cancer. Cell 100(1): 57-70.

Heinloth, A. N., Shackelford, R. E., Innes, C. L., Bennett, L., Li, L., Amin, R. P., Sieber, S. O., Flores, K. G., Bushel, P. R. and Paules, R. S. (2003). Identification of distinct and common gene expression changes after oxidative stress and gamma and ultraviolet radiation. Mol Carcinog 37(2): 65-82.

Herbert, B. S., Wright, W. E. and Shay, J. W. (2001). Telomerase and breast cancer. Breast Cancer Res 3(3): 146-149.

Hughes, T. R., Marton, M. J., Jones, A. R., Roberts, C. J., Stoughton, R., Armour, C. D., Bennett, H. A., Coffey, E., Dai, H., He, Y. D., Kidd, M. J., King, A. M., Meyer, M. R., Slade, D., Lum, P. Y., Stepaniants, S. B., Shoemaker, D. D., Gachotte, D., Chakraburtty, K., Simon, J., Bard, M. and Friend, S. H. (2000). Functional discovery via a compendium of expression profiles. Cell 102(1): 109-126.

Jarvinen, T. A., Tanner, M., Rantanen, V., Barlund, M., Borg, A., Grenman, S. and Isola, J. (2000). Amplification and deletion of topoisomerase IIalpha associate with ErbB-2 amplification and affect sensitivity to topoisomerase II inhibitor doxorubicin in breast cancer. Am J Pathol 156(3): 839-847.

Kimura, M., Kotani, S., Hattori, T., Sumi, N., Yoshioka, T., Todokoro, K. and Okano, Y. (1997). Cell cycle-dependent expression and spindle pole localization of a novel human protein kinase, Aik, related to Aurora of Drosophila and yeast Ipl1. J Biol Chem 272(21): 13766-13771.

Korn, E. L., McShane, L. M., Troendle, J. F., Rosenwald, A. and Simon, R. (2002). Identifying pre-post chemotherapy differences in gene expression in breast tumours: a statistical method appropriate for this aim. Br J Cancer 86(7): 1093-1096.

Kufer, T. A., Sillje, H. H., Korner, R., Gruss, O. J., Meraldi, P. and Nigg, E. A. (2002). Human TPX2 is required for targeting Aurora-A kinase to the spindle. J Cell Biol 158(4): 617-623.

Liu, G. and Chen, X. (2002). The ferredoxin reductase gene is regulated by the p53 family and sensitizes cells to oxidative stress-induced apoptosis. Oncogene 21(47): 7195-7204.

Longley, D. B., Harkin, D. P. and Johnston, P. G. (2003). 5-fluorouracil: mechanisms of action and clinical strategies. Nat Rev Cancer 3(5): 330-338.

Ma, Y. and Hendershot, L. M. (2001). The unfolding tale of the unfolded protein response. Cell 107(7): 827-830.

Meyer, K. M., Hess, S. M., Tlsty, T. D. and Leadon, S. A. (1999). Human mammary epithelial cells exhibit a differential p53-mediated response following exposure to ionizing radiation or UV light. Oncogene 18(42): 5795-5805.

Morgan, K. T., Ni, H., Brown, H. R., Yoon, L., Qualls, C. W., Jr., Crosby, L. M., Reynolds, R., Gaskill, B., Anderson, S. P., Kepler, T. B., Brainard, T., Liv, N., Easton, M., Merrill, C., Creech, D., Sprenger, D., Conner, G., Johnson, P. R., Fox, T., Sartor, M., Richard, E., Kuruvilla, S., Casey, W. and Benavides, G. (2002). Application of cDNA microarray technology to in vitro toxicology and the selection of genes for a real-time

RT-PCR-based screen for oxidative stress in Hep-G2 cells. Toxicol Pathol 30(4): 435-451.

Park, W. Y., Hwang, C. I., Im, C. N., Kang, M. J., Woo, J. H., Kim, J. H., Kim, Y. S., Kim, H., Kim, K. A., Yu, H. J., Lee, S. J., Lee, Y. S. and Seo, J. S. (2002). Identification of radiation-specific responses from gene expression profile. Oncogene 21(55): 8521-8528.

Pechoux, C., Gudjonsson, T., Ronnov-Jessen, L., Bissell, M. J. and Petersen, O. W. (1999). Human mammary luminal epithelial cells contain progenitors to myoepithelial cells. Dev Biol 206(1): 88-99.

Pegram, M. D., Finn, R. S., Arzoo, K., Beryt, M., Pietras, R. J. and Slamon, D. J. (1997). The effect of HER-2/neu overexpression on chemotherapeutic drug sensitivity in human breast and ovarian cancer cells. Oncogene 15(5): 537-547.

Perou, C. M., Jeffrey, S. S., van de Rijn, M., Rees, C. A., Eisen, M. B., Ross, D. T., Pergamenschikov, A., Williams, C. F., Zhu, S. X., Lee, J. C., Lashkari, D., Shalon, D., Brown, P. O. and Botstein, D. (1999). Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. Proc Natl Acad Sci U S A 96(16): 9212-9217.

Perou, C. M., Sørlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., Fluge, O., Pergamenschikov, A., Williams, C., Zhu, S. X., Lonning, P. E., Borresen-Dale, A. L., Brown, P. O. and Botstein, D. (2000). Molecular portraits of human breast tumours. Nature 406(6797): 747-752.

Ross, D. T. and Perou, C. M. (2001). A comparison of gene expression signatures from breast tumors and breast tissue derived cell lines. Dis Markers 17(2): 99-109.

Ross, D. T., Scherf, U., Eisen, M. B., Perou, C. M., Rees, C., Spellman, P., Iyer, V., Jeffrey, S. S., Van de Rijn, M., Waltham, M., Pergamenschikov, A., Lee, J. C., Lashkari, D., Shalon, D., Myers, T. G., Weinstein, J. N., Botstein, D. and Brown, P. O. (2000). Systematic variation in gene expression patterns in human cancer cell lines. Nat Genet 24(3): 227-235.

Scherf, U., Ross, D. T., Waltham, M., Smith, L. H., Lee, J. K., Tanabe, L., Kohn, K. W., Reinhold, W. C., Myers, T. G., Andrews, D. T., Scudiero, D. A., Eisen, M. B., Sausville, E. A., Pommier, Y., Botstein, D., Brown, P. O. and Weinstein, J. N. (2000). A gene expression database for the molecular pharmacology of cancer [see comments]. Nat Genet 24(3): 236-244.

Sesto, A., Navarro, M., Burslem, F. and Jorcano, J. L. (2002). Analysis of the ultraviolet B response in primary human keratinocytes using oligonucleotide microarrays. Proc Natl Acad Sci U S A 99(5): 2965-2970.

Sikder, H. A., Devlin, M. K., Dunlap, S., Ryu, B. and Alani, R. M. (2003). Id proteins in cell growth and tumorigenesis. Cancer Cell 3(6): 525-530.

Smith, G. H. and Chepko, G. (2001). Mammary epithelial stem cells. Microsc Res Tech 52(2): 190-203.

Sørlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Thorsen, T., Quist, H., Matese, J. C., Brown, P. O., Botstein, D., Eystein Lonning, P. and Borresen-Dale, A. L. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 98(19): 10869-10874.

Sørlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J. S., Nobel, A., Deng, S., Johnsen, H., Pesich, R., Geisler, S., Demeter, J., Perou, C. M., Lonning, P. E., Brown, P. O., Borresen-Dale, A. L. and Botstein, D. (2003). Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A 100(14): 8418-8423.

Sotiriou, C., Neo, S. Y., McShane, L. M., Korn, E. L., Long, P. M., Jazaeri, A., Martiat, P., Fox, S. B., Harris, A. L. and Liu, E. T. (2003). Breast cancer classification and prognosis based on gene expression profiles from a population-based study. Proc Natl Acad Sci U S A 100(18): 10393-10398.

Thor, A. D., Berry, D. A., Budman, D. R., Muss, H. B., Kute, T., Henderson, I. C., Barcos, M., Cirrincione, C., Edgerton, S., Allred, C., Norton, L. and Liu, E. T. (1998). erbB-2, p53, and efficacy of adjuvant therapy in lymph node-positive breast cancer. J Natl Cancer Inst 90(18): 1346-1360.

Tusher, V., Tibshirani, R. and Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A 98(9): 5116-5121.

Van Ewijk, P. H. and Hoekstra, J. A. (1993). Calculation of the EC50 and its confidence interval when subtoxic stimulus is present. Ecotoxicol Environ Saf 25(1): 25-32.

van 't Veer, L. J., Dai, H., van de Vijver, M. J., He, Y. D., Hart, A. A., Mao, M., Peterse, H. L., van der Kooy, K., Marton, M. J., Witteveen, A. T., Schreiber, G. J., Kerkhoven, R. M., Roberts, C., Linsley, P. S., Bernards, R. and Friend, S. H. (2002). Gene expression profiling predicts clinical outcome of breast cancer. Nature 415(6871): 530-536.

Wang, H. C. and Fedoroff, S. (1972). Banding in human chromosomes treated with trypsin. Nat New Biol 235(54): 52-54.

Weigel, A. L., Handa, J. T. and Hjelmeland, L. M. (2002). Microarray analysis of H2O2-, HNE-, or tBH-treated ARPE-19 cells. Free Radic Biol Med 33(10): 1419-1432.

Whitfield, M. L., Sherlock, G., Saldanha, A. J., Murray, J. I., Ball, C. A., Alexander, K. E., Matese, J. C., Perou, C. M., Hurt, M. M., Brown, P. O. and Botstein, D. (2002). Identification of genes periodically expressed in the human cell cycle and their expression in tumors. Mol Biol Cell 13(6): 1977-2000.

CHAPTER III

PREDICTION OF TOXICANT-SPECIFIC GENE EXPRESSION SIGNATURES AFTER CHEMOTHERAPEUTIC TREATMENT OF BREAST CELL LINES

PREFACE

This work was previously published and represents a co-first author effort between Melissa Troester and myself. I performed the experiments with 5-fluorouracil treatment and performed data analysis. Melissa Troester performed the experiments with doxorubicin treatment and performed data analysis. Melissa Troester and I equally contributed to the writing of the manuscript. Joel Parker performed the PAM and KNN analyses. Charles Perou conceived the project.

Melissa A. Troester, Katherine A. Hoadley, Joel S. Parker, and Charles M. Perou. (2004). Prediction of toxicant-specific gene expression signatures after chemotherapeutic treatment of breast cell lines. Environmental Health Perspectives 112(16): 1607-1613.

ABSTRACT

Global gene expression profiling has demonstrated that the predominant cellular response to a range of toxicants is a general stress response. This stereotyped environmental stress response commonly includes repression of protein synthesis and cell-cycle-regulated genes and induction of DNA damage and oxidative stress-responsive genes. Our laboratory

has recently characterized the general stress response of breast cell lines derived from basal-like and luminal epithelium following treatment with doxorubicin (DOX) or 5-fluorouracil (5FU) and showed that each cell type has a distinct response. However, we expected that some of the expression changes induced by DOX and 5FU would be unique to each compound and might reflect the underlying mechanisms of action of these agents. Therefore, we employed supervised analyses (Significance Analysis of Microarrays) to identify genes that showed differential expression between DOX-treated and 5FU-treated cell lines. We then used cross-validation analyses and identified genes that afforded high predictive accuracy in classifying samples into the two treatment classes. To test whether these gene lists had good predictive accuracy in an independent data set, we treated our panel of cell lines with etoposide, a compound mechanistically similar to DOX. We demonstrated that using expression patterns of 100 genes we were able to obtain 100% predictive accuracy in classifying the etoposide samples as being more similar in expression to DOX-treated than 5FU-treated samples. These analyses also showed that toxicant-specific gene expression patterns, similar to general stress responses, vary according to cell type.

INTRODUCTION

A stereotyped environmental stress response to a wide range of stressors and toxicants was first demonstrated in yeast (Gasch et al., 2000) and has subsequently been observed in a variety of mammalian cell models (Morgan et al., 2002; Park et al., 2002; Sesto et al., 2002; Weigel et al., 2002; Heinloth et al., 2003; Murray et al., 2004). We have previously used DNA microarray experiments to characterize the transcriptional responses of four breast cell lines to the chemotherapeutics doxorubicin (DOX) and 5-fluorouracil (5FU);

these cell lines included two hTERT-immortalized human mammary epithelial (HME) cell lines and two tumor-derived cell lines of luminal epithelial origin (MCF-7 and ZR-75-1). A general stress response was shown to predominate when these cells were treated with DOX and 5FU (Troester et al., 2004). All four cell lines repressed genes involved in cell growth and induced DNA-damage response and xenobiotic metabolism genes, but there were differences in the general stress responses depending upon the cell type or origin of the cell line.

The mechanisms of action of DOX and 5FU are distinct. DOX is a topoisomerase IIA (TOP2A) poison. TOP2A is a nuclear enzyme that transiently breaks and rejoins the phosphodiester backbone of both strands of the double helix. As such, it is vital for DNA replication, chromosome segregation, and maintenance of chromosome structure. In previous studies (Tewey et al., 1984), DOX formed a stable ternary complex with DNA and TOP2A, thereby inhibiting the normal function of the enzyme. The complexed enzyme is unable to re-ligate DNA so complex formation increases DNA strand breaks. TOP2A is highly expressed during S-phase, but TOP2A poisoning causes cell cycle arrest in G2-M. The commonly used chemotherapeutic 5FU has several known mechanisms of action that distinguish it from DOX. 5FU covalently binds to thymidylate synthase, preventing *de novo* production of thymidine. It also incorporates into DNA and RNA (Pizzorno et al., 2000; Longley et al., 2003). The importance of each of these 5FU-mediated disruptions in cellular metabolism varies across cell lines and patients, but current studies emphasize the role of thymidylate synthase inhibition (Banerjee et al., 2002; Peters et al., 2002; Longley et al., 2003). Thymidylate synthase is highly expressed during S-phase and its inhibition is thought to cause cell cycle arrest in S-phase.

Using microarrays, it is often possible to identify unique patterns associated with specific toxicants in addition to common patterns of response. We used our panel of treated breast cell lines (Troester et al., 2004) to identify toxicant-specific expression signatures for DOX and 5FU. Cell lines derived from breast basal-like and luminal epithelium had distinct toxicant-specific patterns of response. Using two statistical methods for class prediction, we then identified lists of genes that distinguish DOX-TREATED and 5FU-treated cells and used these lists to predict the mechanism of etoposide (ETOP), a drug that is mechanistically similar to DOX.

MATERIALS AND METHODS

Cells and Cell Culture Conditions

ME16C and HME-CC cells, two basal-like hTERT-immortalized HME cell lines described by Troester et al. (Troester et al., 2004), were gifts from Jerry W. Shay at the University of Texas Southwestern Medical Center and Christopher Counter at Duke Medical Center, respectively. ME16C cells and HME-CC cells were maintained in Mammary Epithelium Growth Media (Cambrex Bio Science Walkersville). MCF-7 cells (a gift from F. Tamanoi at University of California at Los Angeles) and ZR-75-1 cells (American Type Culture Collection) were maintained in RPMI 1640 with L-glutamine (GIBCO) supplemented with 10% Fetal Bovine Serum (Sigma), and 50 unit/mL penicillin, 50 unit/mL streptomycin (GIBCO). All cell lines were tested for *Mycoplasma* by the University of North Carolina at Chapel Hill Tissue Culture Facility prior to conducting experiments and at regular intervals thereafter. Cells were maintained at 37°C and 5% carbon dioxide.

Cytotoxicity Assay

A mitochondrial dye conversion assay (Cell Titer 96, Promega) was used to measure cell viability following treatment. This assay was conducted according to manufacturer's instructions, with modification as follows. Briefly, 5,000 cells were seeded per well of a 96-well plate. Cells were allowed to adhere overnight and then media were replaced with fresh media containing a range of drug doses (DOX: 0 – 1 μM, ETOP: 0 – 500 μM, 5FU: 0 – 10 mM). After 36 h of drug treatment, 15 μL of tetrazolium dye solution were added and incubated for 1 h at 37 °C before adding stop solution. Dye conversion products were solubilized in a humidified chamber overnight, and absorbance was measured at 570 nm (minus background absorbance at 650 nm). The IC50 for 36h of treatment with each drug in each cell line was estimated using nonlinear regression (SAS Statistical Software, version 8; SAS Institute Inc., Cary, NC) as described previously (Troester et al., 2004).

Collection of mRNA for Microarray Experiments

Cell lines were grown in 150-mm dishes to 70-80% confluence and then treated for 12h, 24h, or 36h with toxicant at the IC50 concentration. The cells were harvested by scraping and cell lysates were enriched for mRNA using a Micro-FastTrack kit (Invitrogen). The reference RNA was generated by harvesting mRNA from each cell line at 80% confluence and pooling four such harvests (i.e. four MCF-7 harvests were pooled and served as reference mRNA for all MCF-7 experiments).

Microarray Experiments

To synthesize labeled cDNA, reverse transcription reactions were carried out using 3 μg of mRNA as described previously (Perou et al., 2000; Troester et al., 2004). Briefly, 5FU, DOX, ETOP, and vehicle controls were labeled with Cy5-dUTP and the pooled cell line control was labeled with Cy3-dUTP. The Cy3- and Cy5-labeled sampled were combined and hybridized overnight at 65 ºC to a custom oligonucleotide microarray created in the University of North Carolina at Chapel Hill Genomics Core Facility. Arrays were spotted with Compugen Human oligos representing approximately 22,000 genes. Two replicate arrays for each sample were selected for subsequent analysis. All microarray raw data tables are available at the UNC Microarray Database [https://genome.unc.edu/] and have been deposited in the Gene Expression Omnibus under the accession number GSE1647 (submitter C. Perou).

Significance Analysis of Microarrays

Genes that were significantly up- or down-regulated were identified using the Significance Analysis of Microarrays (SAM) Add-In for Microsoft Excel (Tusher et al., 2001). For SAM analysis, data were excluded for genes that did not have mean intensity greater than twice the median background for both the red and green channel in at least 70% of the experiments. The log-base-2 of the median red intensity over median green intensity was calculated for each gene. Missing data were imputed using the SAM for Excel plug-in with 100 permutations and *k*-nearest neighbors (KNN) with k=10. For each cell line, 12h, 24h, and 3h DOX-treated arrays were coded as one class and were compared the 12h, 24h, and 36h 5FU-treated arrays using a two-class, unpaired SAM analysis. Delta values were

adjusted to obtain the largest gene list with a false discovery rate less than 5%. The effects of adding media would be present in the signatures of both compounds and would not be identified as significantly associated with either toxicant. However, because DOX was solubilized in water and ETOP and 5FU were solubilized in dimethylsulfoxide (DMSO), we also collected mRNA from each cell line treated with DMSO only for 12h, 24h, or 36h hours (data not shown). We compared these DMSO-treated samples to sham (media only) samples for the same time points using SAM. The lowest false discovery rate obtained was 15.3% (15 genes with 2.29 false significant). The toxicant-specific changes we detected using these analyses are therefore unlikely to reflect changes induced by vehicle.

Class Prediction

The number of genes needed to distinguish DOX and 5FU samples were identified using 10-fold cross validation (CV) analysis using Prediction Analysis of Microarrays (PAM) and a KNN classifier. The KNN metric uses the Euclidian distance to determine the similarity of a sample to its *k* nearest sample neighbors. To select genes for the KNN method, we used a gene selection method that was first described by Dudoit and Fridlyand (2002); the KNN genes were identified in the training set according to the ratio of between-group to within-group sums of squares (Dudoit and Fridlyand, 2002). The top *n* ranked genes were used for each round of cross-validation. The size of the gene subset was increased for subsequent rounds of CV. The *n* top ranked genes that gave the highest average prediction accuracy during CV was also determined and reported. Gene selection using PAM was completed as described previously (Tibshirani et al., 2002). Genes were selected that yielded the greatest predictive accuracy in classifying DOX versus 5FU using a 10-fold CV analysis.

For class prediction, we performed a 10-fold CV analysis to iteratively optimize the list of genes and to determine prediction accuracies. Each round of CV would begin by splitting the samples into a training set (90% of the samples) and a test set (10% left-out samples), with gene selection and training being performed on the 90% and then used to predict the status of the withheld 10%. This was repeated 10 times, each time using a different 10% subset and a different gene set. Our reported prediction accuracies are the average of these iterative cycles of prediction for the optimized model. To independently assess the validity of these gene lists, we used them to predict class for ETOP samples; this analysis is independent because the ETOP samples were not used to train the predictor.

Clustering of Toxicant-Specific Responses

Once gene lists had been identified for the toxicant-specific responses of each cell line, cluster analysis was conducted using Cluster to perform uncentered, average-linkage clustering; the data were visualized using Treeview (Eisen et al., 1998; Eisen and Brown, 1999). The gene lists generated with SAM for the luminal lines (MCF-7 and ZR-75-1) were combined into a non-redundant list, and data for these genes were compiled for all MCF-7 and ZR-75-1 samples. Likewise, the gene lists for the two basal-like lines (ME16C and HME-CC) were combined into a non-redundant list and data for these genes were compiled for all HME-hTERT samples. For clustering and displaying results, data were excluded for genes that did not have mean intensity greater than twice the median background for both the red and green channel in at least 80% (Figures 3.1 and 3.2) or 70% (Figure 3.3) of the experiments.

RESULTS

Toxicant-Specific Transcriptional Responses

To investigate the toxicant-specific responses of four breast cell lines treated with chemotherapeutics, we collected mRNA from MCF-7, ZR-75-1, ME16C and HME-CC cell lines after treating with DOX and 5FU at doses that produced similar levels of toxicity (IC50) across all four lines.

The IC50 concentration was estimated from mitochondrial dye conversion assay results after 36h treatments with 5FU and DOX. The IC50 concentrations and their 95% confidence intervals are shown in Table 3.1. For DOX and 5FU, the doses selected are consistent with physiological doses expected in patients receiving treatment with DOX (Gewirtz, 1999) or 5FU (Peters et al., 1993; Terret et al., 2000). This experimental design was aimed at defining the steady-state transcriptional response of these cell lines to toxicants and on defining chemotherapeutic-specific responses that were consistent over time. By combining 12h, 24h, and 36h treated experiments into a single class for all supervised analyses; we identified genes that had a consistent pattern of expression across all three time points. These genes are most likely to be consistent with *in vivo* experiments or patient samples, where it is difficult to assess how long a tissue sample has been exposed to a toxic agent. While we did not specifically study temporal variation in our SAM analyses, some of the temporal variation in gene expression can be observed in the clusters.

Toxicant-Specific Responses in Luminal Cell Lines

A large list of genes was identified for MCF-7 (974 genes with 44.7 false significant) and for ZR-75-1 (883 genes with 41.6 false significant) when supervised analyses were

Table 3.1: Estimated inhibitory concentration 50% (IC50) for 5-fluorouracil, doxorubicin, and etoposide based on mitochondrial dye conversion assay[a]

	Cell Line	IC50[b]	Treatment Dose[b]
5-Fluorouracil	MCF-7	0.034 (0.13, 0.55)	0.3
	ZR-75-1	3.3 (2.8, 3.7)	3.0
	ME16C	0.064 (0.055, 0.074)	0.06
	HME-CC	0.011 (0.009, 0.013)	0.01
Doxorubicin	MCF-7	0.86 (0.74, 0.97)	0.9
	ZR-75-1	0.43 (0.37, 0.50)	0.4
	ME16C	0.52 (0.49, 0.54)	0.5
	HME-CC	0.16 (0.14, 0.18)	0.2
Etoposide	MCF-7	35 (30, 40)	40
	ZR-75-1	26 (8.6, 43)	30
	ME16C	21 (18,23)	20
	HME-CC	6.1 (5.6, 6.7)	10

[a]IC50 values for 5-fluorouracil and doxorubicin were previously reported (Troester et al., 2004).

[b]Values in parentheses represent 95% confidence intervals. Doses for 5-fluorouracil in mM; those for doxorubicin and etoposide in μM.

conducted to compare DOX-treated versus 5FU-treated samples. Hierarchical clustering analysis of the MCF-7 and ZR-75-1 experiments using the combined and non-redundant gene lists showed distinct responses for each toxicant (Figure 3.1). The primary dendrogram branches for DOX-treated and 5FU-treated experiments were subdivided into MCF-7 and ZR-75-1 branches (Figure 3.1B); this suggests that the majority of variation in these genes is attributable to the toxicant, but that cell lines also contribute to the variation. A total of 191 genes (77 down-regulated and 114 up-regulated) appeared on the SAM lists for both MCF-7 and ZR-75-1. However, there are many more genes that show qualitative similarity in the toxicant-specific responses of MCF-7 and ZR-75-1 cells (Figure 3.1) than is captured using the strict SAM analysis. Figure 3.1D shows a cluster of genes that is up-regulated in MCF-7 cells following DOX treatment, but which is down-regulated in ZR-75-1 cells following both treatments; *thymidylate synthase* is included in this cluster. Recent studies have shown that thymidylate synthase protein, the target of 5FU, binds *p53* mRNA and regulates the expression of *p53* at the translational level (Chu et al., 1999; Ju et al., 1999). This is interesting because *p53* expression is slightly induced by DOX in MCF-7 cells, but not in ZR-75-1 cells nor by 5FU treatment in either cell line (Figure 3.1E).

The gene set in Figure 3.1E also shows that several other genes had slightly higher expression in MCF-7 cells treated with DOX, and that these genes were typically repressed in ZR-75-1 cells. For example, the mismatch repair gene *mutL homolog 1* (*MLH1*) was unchanged by DOX, and *N-methyl-purine-DNA glycosylase*, a base excision repair gene was repressed by 5FU. Both DOX and 5FU can cause DNA damage, but differences in the profiles of damage induced by each compound may account for differently regulated repair enzymes. *Cyclin E1* was also slightly induced in DOX-treated MCF-7 cells, as has been

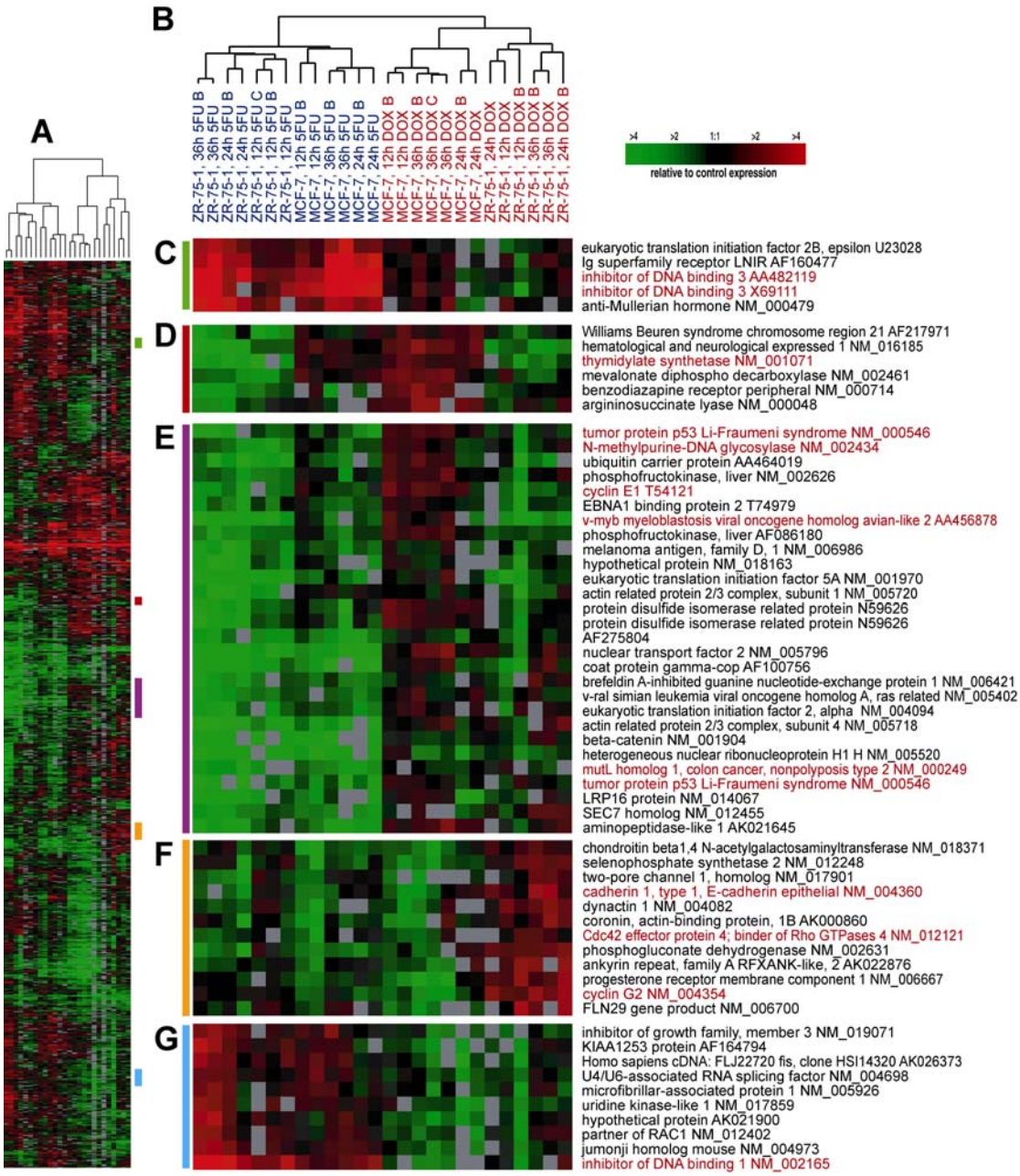

Figure 3.1: Gene expression patterns for genes that distinguish between DOX-treated and 5FU-treated luminal cells (MCF-7 and ZR-75-1). Hierarchical clustering analysis was conducted using 13 DOX-treated and 13 5FU-treated samples. Data from the union of the genes identified by SAM for MCF-7 and ZR-75-1 were identified and combined into a non-redundant list, and the compressed cluster is shown in *A*. Colored bars in A illustrate the location of clusters shown in *C*-*G*. The dendrogram in *B* shows that the samples clustered into two groups according to treatment (DOX experiments labeled in red, 5FU experiments labeled in blue), but within each treatment branch, cell-line-specific branches are also identifiable. Gene names highlighted in red are discussed in the text.

shown in previous studies (Arooz et al., 2000). *Cyclin E1*, along with *v-myb myeloblastosis viral oncogene homolog avian-like 2* (*MYBL2*), are important genes involved in the G1-S transition and are transcriptional targets of E2F (Yasui et al., 2003).

Figure 3.1F shows that ZR-75-1 cells have a unique response to DOX compared to MCF-7 cells and 5FU-treated cells. In concordance with increased *E-cadherin* expression shown in this cluster, an increase in *E-cadherin* mRNA (and E-cadherin-mediated cell-cell adhesion) has been shown previously in another breast cancer cell line following treatment with DOX (Yang et al., 1999). *Cyclin G2* was also induced in ZR-75-1 cells treated with DOX. This cyclin is inducible by DNA damage in a p53-independent manner (Bates et al., 1996).

Figures 3.1C and G show clusters of genes that are induced by 5FU in both cell lines and either unchanged or only modestly changed in DOX-treated lines. For example, *inhibitor of DNA binding 3* (*ID3*) (Figure 3.1C) and *inhibitor of DNA binding 1* (*ID1*) (Figure 3.1G) were strongly induced only in the 5FU-treated samples. The Id proteins control cellular differentiation and cell-cycle progression by preventing transcription factors from binding DNA (Norton et al., 1998). These proteins target basic helix-loop-helix proteins that regulate cell-type-specific and cell-cycle-regulatory gene expression (Lassar et al., 1994); however, the role of these proteins in the response to 5FU is not known.

Toxicant-Specific Responses in Basal-Like Cell Lines

A smaller list of toxicant-specific genes was identified for ME16C (76 genes with 3.7 false significant) and HME-CC (193 genes with 8.6 false significant) cells when SAM was used to compare DOX-treated with 5FU-treated samples. Hierarchical clustering using the

combined and non-redundant gene lists for these two cell lines showed that there were distinct responses by toxicant (Figure 3.2). However, the primary dendrogram branch for 5FU-treated basal-like cell lines also included two early time points for DOX-treated ME16C (Figure 3.2B). The 12h ME16C time point has many gene expression changes in response to treatment (Troester et al., 2004), but this time point does not exhibit the same toxicant-specific signature as the 24h and 36h time points. These temporal differences likely account for the grouping of toxicant-specific signatures in Figure 3.2. As we have also seen in our previous study of the general stress response of these cell lines, the temporal response to these two toxicants varies by cell line.

Figure 3.2C shows a cluster of genes that is up-regulated in DOX-treated basal-like cell lines, but down-regulated in 5FU-treated basal-like cells. These genes differ in both magnitude and direction of change. A number of these genes play a role in mediating DNA repair, including the *ubiquitin conjugating enzyme E2A*, which is a member of the RAD6 pathway that uses ubiquitin conjugation to control DNA damage-induced mutagenesis (Stelter and Ulrich, 2003). Similarly, DNA polymerase delta is known to repair single strand DNA interruptions produced during the process of base excision repair (Ho and Satoh, 2003). *Cell division cycle 25B*, an important regulator of mitosis, is also found in this cluster.

The cluster in Figure 3.2D contains several mitochondrial genes (indicated in red). The altered expression of mitochondrial genes might be expected based on a recent study that demonstrated that anthracyclines, such as DOX, impair cellular respiration (Souid et al., 2003). Figure 3.2E consists of a set of genes that are clearly enriched for ribosomal proteins. Disruption of protein biosynthesis has been associated with alterations in the cell cycle and cell growth (Ruggero and Pandolfi, 2003). Five ribosomal proteins are highlighted in red and

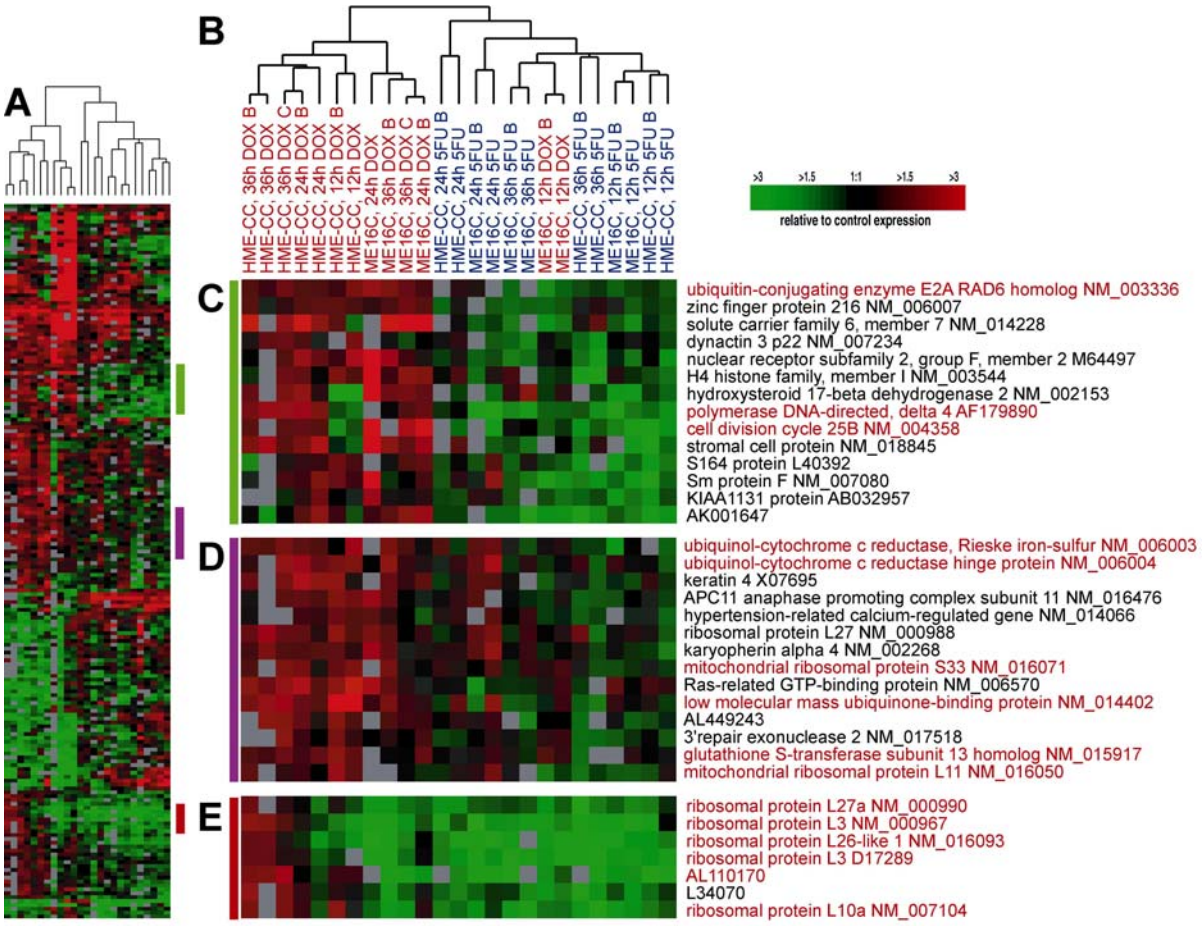

Figure 3.2: Gene expression patterns for genes that distinguish between DOX-treated and 5FU-treated basal-like cells (ME16C and HME-CC). Hierarchical clustering analysis was conducted using 13 DOX-treated and 12 5FU-treated samples. Data from the union of the genes identified by SAM for ME16C and HME-CC were identified and combined into a non-redundant list, and the compressed cluster is shown in *A*. Colored bars in *A* illustrate the location of clusters shown in *C-E*. The dendrogram in *B* shows that the samples clustered into two groups according to treatment (DOX experiments labeled in red, 5FU experiments labeled in blue); however, the early time points for DOX-treated ME16C samples clustered with the 5FU treated samples. Gene names highlighted in red are discussed in the text.

AL110170 is a hypothetical protein with 65% homology to ribosomal protein L22. The genes for these proteins are induced in the DOX-treated HME-CC cell line after 36 hours, but are repressed in the ME16C cells at this and all other time points assayed.

Class Prediction and Sample Classification for Etoposide Treated Samples

Having identified a number of genes that distinguish DOX-treated from 5FU-treated breast cell lines using SAM, we next performed class prediction analyses to assess whether these differences could be used to classify an independent data set collected using the same four cell lines. Because SAM does not perform classification, we used ten-fold CV with PAM (Tibshirani et al., 2002) and a KNN metric based upon the work of Dudoit and Fridlyand (Dudoit and Fridlyand, 2002). CV was implemented to optimize the number of neighbors (k) and the number of genes for KNN and to optimize the shrinkage parameter (delta) for PAM. Parameters were selected that generated the highest CV accuracy (internal validation) when distinguishing the DOX-treated and 5FU-treated samples. Then, using the optimized models, we made predictions on a test set of ETOP-treated samples (external validation). [Note that because CV excludes samples and the final model using the optimized parameters does not, the delta value selected during CV with PAM may correspond to a different number of genes during prediction. However, the number of genes selected in cross validation is held constant for a KNN-based prediction.] The results of all CV analyses using PAM and KNN are presented in Appendices IIIA-D.

We expected that because ETOP and DOX both inhibit TOP2A, their resulting transcriptional profiles should be similar. Therefore, we consider ETOP samples correctly classified if they were classified as DOX. In the two-class analysis (DOX versus 5FU), we

obtained a high degree of cross validation accuracy (80-98%) during training and a high degree of predictive accuracy (100%) in assigning the ETOP experiments as more similar to DOX than 5FU (Table 3.2). However, when we attempted to further subclassify the DOX and 5FU samples according to cell-type (basal-like-DOX versus basal-like-5FU versus luminal-DOX versus luminal-5FU), our cross validation (76-80%) and prediction (75%) accuracies were diminished (Table 3.3). The errors in four-class prediction occurred in the 12h basal-like samples. This is not surprising based on our SAM results in Figure 3.2, where the early time points in one of the basal-like cell lines appeared distinct from later points.

To visualize the expression differences from the two-class DOX versus 5FU predictor using Euclidian-KNN, we took these samples and the 100 gene set shown to be 98% accurate in prediction and performed hierarchical clustering analysis (Figure 3.3). The similarities between the ETOP and DOX samples are observable across this gene set. This analysis showed two separate dendrogram branches in Figure 3.3B with one branch containing all of the 5FU samples and the other containing all of the ETOP and DOX samples. Some of the genes identified in the earlier supervised analysis were recapitulated in this predictive gene set. Notably, *ID3* appears in Figure 3.3C and *p53* appears in Figure 3.3E. An interesting cluster of genes that was more strongly induced in DOX and ETOP samples appears in Figure 3.3D, which includes the genes *cathepsin L* and *cystatin C*. The activity of cysteine protease cathepsin L is regulated by the cystatins (a family of cysteine proteinase inhibitors) and their imbalance is associated with increased invasiveness and development of the malignant cell phenotype (Kos and Lah, 1998).

Table 3.2: Two-class cross-validation and prediction accuracy for etoposide samples.

	Cross-Validation Accuracy		Prediction Accuracy	
Method	PAM	KNN[a]	PAM	KNN[a]
No. genes	2460 (2.75)[b]	100	279 (2.75)[a]	100
Accuracy	80%	96%	100%	100%

[a]k=11; [b]Delta value is shown in parentheses.

Table 3.3: Four-class cross-validation and prediction accuracy for etoposide samples.

	Cross-Validation Accuracy		Prediction Accuracy	
Method	PAM	KNN[a]	PAM	KNN[a]
No. genes	652 (3.5)[b]	100	465 (3.5)[a]	100
Accuracy	76%	86%	75%	75%

[a]k=9; [b]Delta value is shown in parentheses.

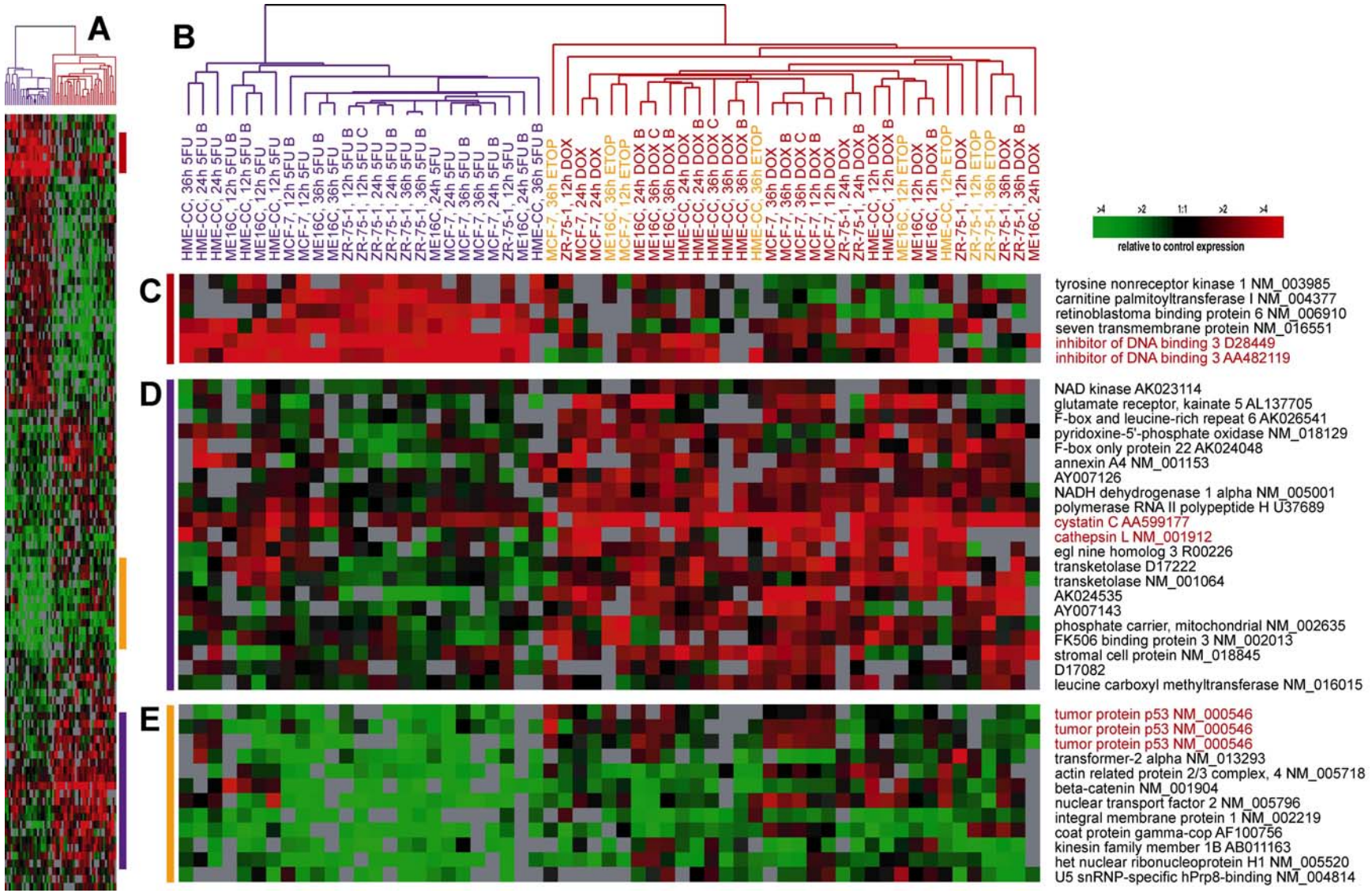
A
B
HME-CC, 36h 5FU
HME-CC, 24h 5FU B
HME-CC, 24h 5FU
ME16C, 12h 5FU B
HME-CC, 12h 5FU
ME16C, 12h 5FU
HME-CC, 12h 5FU B
MCF-7, 12h 5FU
MCF-7, 12h 5FU B
ME16C, 36h 5FU B
ME16C, 36h 5FU
ZR-75-1, 12h 5FU B
ZR-75-1, 12h 5FU C
ZR-75-1, 24h 5FU
ZR-75-1, 24h 5FU B
ZR-75-1, 36h 5FU
ZR-75-1, 36h 5FU B
ME16C, 24h 5FU
MCF-7, 24h 5FU
MCF-7, 36h 5FU B
MCF-7, 36h 5FU
MCF-7, 24h 5FU B
ZR-75-1, 12h 5FU
ME16C, 24h 5FU B
HME-CC, 36h 5FU B
MCF-7, 36h ETOP
ZR-75-1, 12h DOX
MCF-7, 24h DOX
MCF-7, 24h DOX
ME16C, 36h ETOP
MCF-7, 12h ETOP
ME16C, 24h DOX B
ME16C, 36h DOX C
ME16C, 36h DOX B
HME-CC, 24h DOX
HME-CC, 24h DOX B
HME-CC, 36h DOX
HME-CC, 36h DOX B
HME-CC, 36h ETOP
MCF-7, 36h DOX
MCF-7, 36h DOX B
MCF-7, 36h DOX C
MCF-7, 12h DOX B
MCF-7, 12h DOX
ZR-75-1, 24h DOX
ZR-75-1, 24h DOX B
HME-CC, 12h DOX
HME-CC, 12h DOX B
ME16C, 12h ETOP
ME16C, 12h DOX
ME16C, 12h DOX B
HME-CC, 12h ETOP
ZR-75-1, 12h DOX
ZR-75-1, 12h ETOP
ZR-75-1, 36h ETOP
ZR-75-1, 36h DOX
ZR-75-1, 36h DOX B
ME16C, 24h DOX
>4
>2
1:1
>2
>4
relative to control expression
C
tyrosine nonreceptor kinase 1 NM_003985
carnitine palmitoyltransferase I NM_004377
retinoblastoma binding protein 6 NM_006910
seven transmembrane protein NM_016551
inhibitor of DNA binding 3 D28449
inhibitor of DNA binding 3 AA482119
D
NAD kinase AK023114
glutamate receptor, kainate 5 AL137705
F-box and leucine-rich repeat 6 AK026541
pyridoxine-5'-phosphate oxidase NM_018129
F-box only protein 22 AK024048
annexin A4 NM_001153
AY007126
NADH dehydrogenase 1 alpha NM_005001
polymerase RNA II polypeptide H U37689
cystatin C AA599177
cathepsin L NM_001912
egl nine homolog 3 R00226
transketolase D17222
transketolase NM_001064
AK024535
AY007143
phosphate carrier, mitochondrial NM_002635
FK506 binding protein 3 NM_002013
stromal cell protein NM_018845
D17082
leucine carboxyl methyltransferase NM_016015
E
tumor protein p53 NM_000546
tumor protein p53 NM_000546
tumor protein p53 NM_000546
transformer-2 alpha NM_013293
actin related protein 2/3 complex, 4 NM_005718
beta-catenin NM_001904
nuclear transport factor 2 NM_005796
integral membrane protein 1 NM_002219
coat protein gamma-cop AF100756
kinesin family member 1B AB011163
het nuclear ribonucleoprotein H1 NM_005520
U5 snRNP-specific hPrp8-binding NM_004814

Figure 3.3

Figure 3.3: Gene expression patterns for genes selected for a two-class (DOX vs. 5FU) predictive model. Hierarchical clustering analysis was conducted using 26 DOX-treated, 25 5FU-treated samples and 8 ETOP-treated samples. Data from the genes identified using a KNN classifier for DOX-treated versus 5FU-treated experiments are displayed in the compressed cluster shown in *A*. Colored bars in *A* illustrate the location of clusters shown in *C-E*. The dendrogram in *B* shows that the samples clustered into two groups according to treatment (DOX experiments labeled in red, 5FU experiments labeled in blue, and ETOP experiments labeled in orange). Gene names highlighted in red are discussed in the text.

DISCUSSION

Most changes that occur in gene expression after treatment with either DOX or 5FU are indicative of a general stress response (Troester et al., 2004). However, in the work presented here, we were interested in identifying the toxicant-specific transcriptional responses to DOX and 5FU in breast epithelial cell lines. We conducted several different supervised analyses to find genes that distinguished between DOX and 5FU and were able to define toxicant-specific profiles. Using SAM, we found that each cell type (basal-like or luminal-derived) and each cell line had unique responses to DOX and 5FU. Similar to our previous observations for general stress responses (Troester et al., 2004), we found that the luminal cell lines tended to respond to treatment by differentially regulating a large number of genes, while the basal-like cell lines had fewer gene changes in response to treatment. In addition, the basal-like cell lines showed greater temporal variation in expression than the luminal lines. Some of the genes that comprised the general stress signature for each cell type were also found to have toxicant-specific expression in our supervised analyses. This occurred in cases where both DOX and 5FU induced or repressed gene expression relative to shams, but where one treatment induced a change with greater magnitude. For example, the expression of *cystatin C* was induced more strongly by TOP2A inhibitors than by 5FU (Figure 3.3D), but was induced in both treatments relative to sham (Troester et al., 2004). Thus, *cystatin C* is a general stress response gene with a toxicant-specific gene expression signature.

Toxicant-specific expression responses in our data were corroborated in many cases by published reports with these drugs in the same or similar cell lines. For example, impaired cellular respiration following DOX treatment has been previously reported (Souid

et al., 2003), and in our data, mitochondrial gene expression was altered (Figure 3.2D). Earlier studies have shown that 5FU's target protein thymidylate synthase can bind p53 (Chu et al., 1999; Ju et al., 1999), and we show that *p53* mRNA levels are reduced in our 5FU-treated cells. Thus, many of the gene expression changes that we identified recapitulated previous findings. However, a number of significant changes that were not anticipated based upon the literature were identified and may have functional importance in the response to these drugs. For example, the induction of *ID1 and ID3* has not previously been reported for 5FU. The importance of the Id proteins has only recently begun to be investigated (Norton et al., 1998); our findings suggest that these pathways may be responsive to toxicant treatment and warrant further investigation.

In addition to characterizing the toxicant-specific changes by cell line and cell type, we used toxicant-specific gene lists to make predictions on a third toxicant (ETOP) that is believed to have a similar mechanism of action as one of the training toxicants (DOX). Successfully classifying similar compounds establishes that observed transcriptional responses reflect underlying mode of action. Using as few as 100 genes we were able to classify ETOP samples as being similar to DOX treated samples with 100% predictive accuracy. This predictive accuracy was reduced to 75% when we attempted to further subclassify the DOX and 5FU samples according to cell type of origin. However, considering that with a four-class model, the likelihood of correctly classifying samples by chance is only 25% (compared with 50% for a two-class model), the four-class model still performs very well. The samples that were misclassified were the early time points in basal-like cell lines, which is consistent with our previous findings that the basal-like cell lines have distinct expression profiles at 12h compared to 24h and 36h (Troester et al., 2004).

We have used computational analyses to demonstrate that distinct transcriptional patterns can be identified for mechanistically dissimilar compounds and that toxicants with similar mechanisms can be classified accordingly. We selected two compounds with distinct mechanisms to train our model and a test compound with a mechanism similar to one of the training compounds. These kinds of mechanistic analyses are critical for predictive toxicology using gene arrays. Many studies in the field of toxicogenomics are aimed at populating databases with expression data for diverse toxicants with known mechanisms of action (Hamadeh et al., 2002). These databases can then be used to infer mechanism of action for new compounds. Our data show that this approach is feasible and identifies many new genes and pathways that are important in the response to these toxicants.

ACKNOWLEDGEMENTS

This work was supported by the National Institute of Environmental Health Sciences (NIEHS) grant (5-U19-ES11391-03). M.A.T. was supported by NIEHS Individual National Research Service Award (NRSA) 5F32ES012374 and Institutional NRSA in Environmental Pathology 2T32ES07017. K.A.H. was supported by National Institutes of Health Institutional NRSA in Genetics T32GM07092.

REFERENCES

Arooz, T., Yam, C. H., Siu, W. Y., Lau, A., Li, K. K. and Poon, R. Y. (2000). On the concentrations of cyclins and cyclin-dependent kinases in extracts of cultured human cells. Biochemistry 39(31): 9494-9501.

Banerjee, D., Mayer-Kuckuk, P., Capiaux, G., Budak-Alpdogan, T., Gorlick, R. and Bertino, J. R. (2002). Novel aspects of resistance to drugs targeted to dihydrofolate reductase and thymidylate synthase. Biochim Biophys Acta 1587(2-3): 164-173.

Bates, S., Rowan, S. and Vousden, K. H. (1996). Characterisation of human cyclin G1 and G2: DNA damage inducible genes. Oncogene 13(5): 1103-1109.

Chu, E., Copur, S. M., Ju, J., Chen, T. M., Khleif, S., Voeller, D. M., Mizunuma, N., Patel, M., Maley, G. F., Maley, F. and Allegra, C. J. (1999). Thymidylate synthase protein and p53 mRNA form an in vivo ribonucleoprotein complex. Mol Cell Biol 19(2): 1582-1594.

Dudoit, S. and Fridlyand, J. (2002). A prediction-based resampling method for estimating the number of clusters in a dataset. Genome Biol 3(7): RESEARCH0036.

Eisen, M. B. and Brown, P. O. (1999). DNA arrays for analysis of gene expression. Methods Enzymol 303: 179-205.

Eisen, M. B., Spellman, P. T., Brown, P. O. and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 95(25): 14863-14868.

Gasch, A. P., Spellman, P. T., Kao, C. M., Carmel-Harel, O., Eisen, M. B., Storz, G., Botstein, D. and Brown, P. O. (2000). Genomic expression programs in the response of yeast cells to environmental changes [In Process Citation]. Mol Biol Cell 11(12): 4241-4257.

Gewirtz, D. A. (1999). A critical evaluation of the mechanisms of action proposed for the antitumor effects of the anthracycline antibiotics adriamycin and daunorubicin. Biochem Pharmacol 57(7): 727-741.

Hamadeh, H. K., Bushel, P. R., Jayadev, S., Martin, K., DiSorbo, O., Sieber, S., Bennett, L., Tennant, R., Stoll, R., Barrett, J. C., Blanchard, K., Paules, R. S. and Afshari, C. A. (2002). Gene expression analysis reveals chemical-specific profiles. Toxicol Sci 67(2): 219-231.

Heinloth, A. N., Shackelford, R. E., Innes, C. L., Bennett, L., Li, L., Amin, R. P., Sieber, S. O., Flores, K. G., Bushel, P. R. and Paules, R. S. (2003). Identification of distinct and

common gene expression changes after oxidative stress and gamma and ultraviolet radiation. Mol Carcinog 37(2): 65-82.

Ho, E. L. and Satoh, M. S. (2003). Repair of single-strand DNA interruptions by redundant pathways and its implication in cellular sensitivity to DNA-damaging agents. Nucleic Acids Res 31(23): 7032-7040.

Ju, J., Pedersen-Lane, J., Maley, F. and Chu, E. (1999). Regulation of p53 expression by thymidylate synthase. Proc Natl Acad Sci U S A 96(7): 3769-3774.

Kos, J. and Lah, T. T. (1998). Cysteine proteinases and their endogenous inhibitors: target proteins for prognosis, diagnosis and therapy in cancer (review). Oncol Rep 5(6): 1349-1361.

Lassar, A. B., Skapek, S. X. and Novitch, B. (1994). Regulatory mechanisms that coordinate skeletal muscle differentiation and cell cycle withdrawal. Curr Opin Cell Biol 6(6): 788-794.

Longley, D. B., Harkin, D. P. and Johnston, P. G. (2003). 5-fluorouracil: mechanisms of action and clinical strategies. Nat Rev Cancer 3(5): 330-338.

Morgan, K. T., Ni, H., Brown, H. R., Yoon, L., Qualls, C. W., Jr., Crosby, L. M., Reynolds, R., Gaskill, B., Anderson, S. P., Kepler, T. B., Brainard, T., Liv, N., Easton, M., Merrill, C., Creech, D., Sprenger, D., Conner, G., Johnson, P. R., Fox, T., Sartor, M., Richard, E., Kuruvilla, S., Casey, W. and Benavides, G. (2002). Application of cDNA microarray technology to in vitro toxicology and the selection of genes for a real-time RT-PCR-based screen for oxidative stress in Hep-G2 cells. Toxicol Pathol 30(4): 435-451.

Murray, J. I., Whitfield, M. L., Trinklein, N. D., Myers, R. M., Brown, P. O. and Botstein, D. (2004). Diverse and specific gene expression responses to stresses in cultured human cells. Mol Biol Cell 15(5): 2361-2374.

Norton, J. D., Deed, R. W., Craggs, G. and Sablitzky, F. (1998). Id helix-loop-helix proteins in cell growth and differentiation. Trends Cell Biol 8(2): 58-65.

Park, W. Y., Hwang, C. I., Im, C. N., Kang, M. J., Woo, J. H., Kim, J. H., Kim, Y. S., Kim, H., Kim, K. A., Yu, H. J., Lee, S. J., Lee, Y. S. and Seo, J. S. (2002). Identification of radiation-specific responses from gene expression profile. Oncogene 21(55): 8521-8528.

Perou, C. M., Sørlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., Fluge, O., Pergamenschikov, A., Williams, C., Zhu, S. X., Lonning, P. E., Borresen-Dale, A. L., Brown, P. O. and Botstein, D. (2000). Molecular portraits of human breast tumours. Nature 406(6797): 747-752.

Peters, G. J., Backus, H. H., Freemantle, S., van Triest, B., Codacci-Pisanelli, G., van der Wilt, C. L., Smid, K., Lunec, J., Calvert, A. H., Marsh, S., McLeod, H. L., Bloemena, E., Meijer, S., Jansen, G., van Groeningen, C. J. and Pinedo, H. M. (2002). Induction of thymidylate synthase as a 5-fluorouracil resistance mechanism. Biochim Biophys Acta 1587(2-3): 194-205.

Peters, G. J., Lankelma, J., Kok, R. M., Noordhuis, P., van Groeningen, C. J., van der Wilt, C. L., Meyer, S. and Pinedo, H. M. (1993). Prolonged retention of high concentrations of 5-fluorouracil in human and murine tumors as compared with plasma. Cancer Chemother Pharmacol 31(4): 269-276.

Pizzorno, G., Handschumacher, R. and Cheng, Y.-C. (2000). Pyrimidine and Purine Antimetabolites. Cancer Medicine. R. C. Bast, D. W. Kufe, R. E. Pollocket al. Ontario, Canada, BC Decker Inc: 625-647.

Ruggero, D. and Pandolfi, P. P. (2003). Does the ribosome translate cancer? Nat Rev Cancer 3(3): 179-192.

Sesto, A., Navarro, M., Burslem, F. and Jorcano, J. L. (2002). Analysis of the ultraviolet B response in primary human keratinocytes using oligonucleotide microarrays. Proc Natl Acad Sci U S A 99(5): 2965-2970.

Souid, A. K., Tacka, K. A., Galvan, K. A. and Penefsky, H. S. (2003). Immediate effects of anticancer drugs on mitochondrial oxygen consumption. Biochem Pharmacol 66(6): 977-987.

Stelter, P. and Ulrich, H. D. (2003). Control of spontaneous and damage-induced mutagenesis by SUMO and ubiquitin conjugation. Nature 425(6954): 188-191.

Terret, C., Erdociain, E., Guimbaud, R., Boisdron-Celle, M., McLeod, H. L., Fety-Deporte, R., Lafont, T., Gamelin, E., Bugat, R., Canal, P. and Chatelut, E. (2000). Dose and time dependencies of 5-fluorouracil pharmacokinetics. Clin Pharmacol Ther 68(3): 270-279.

Tewey, K. M., Rowe, T. C., Yang, L., Halligan, B. D. and Liu, L. F. (1984). Adriamycin-induced DNA damage mediated by mammalian DNA topoisomerase II. Science 226(4673): 466-468.

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

Troester, M. A., Hoadley, K. A., Sorlie, T., Herbert, B.-S., Borresen-Dale, A.-L., Lonning, P. E., Shay, J. W., Kaufmann, W. K. and Perou, C. M. (2004). Cell-Type-Specific Responses to Chemotherapeutics in Breast Cancer. Cancer Res 64(12): 4218-4226.

Tusher, V., Tibshirani, R. and Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A 98(9): 5116-5121.

Weigel, A. L., Handa, J. T. and Hjelmeland, L. M. (2002). Microarray analysis of H2O2-, HNE-, or tBH-treated ARPE-19 cells. Free Radic Biol Med 33(10): 1419-1432.

Yang, S. Z., Kohno, N., Kondo, K., Yokoyama, A., Hamada, H., Hiwada, K. and Miyake, M. (1999). Adriamycin activates E-cadherin-mediated cell-cell adhesion in human breast cancer cells. Int J Oncol 15(6): 1109-1115.

Yasui, K., Okamoto, H., Arii, S. and Inazawa, J. (2003). Association of over-expressed TFDP1 with progression of hepatocellular carcinomas. J Hum Genet 48(12): 609-613.

CHAPTER IV

EGFR SIGNALING VARIES WITH BREAST TUMOR SUBTYPE

PREFACE

This chapter represents a manuscript that is in the submission process. I performed the cell line experiments, cell line and tumor data analysis, drafted the paper and helped with the design of the study. Victor Weigman, Cheng Fan, and Melissa Troester assisted with data analysis. Carolyn Sartor made the initial observations of EGFR dependency of SUM102 cells and assisted with the discussion. Lisa Carey, Lynda Sawyer, Thais Rieger-House, and Philip Bernard accomplished tumor sample collection, clinical data acquisition and interpretations of new tumor samples. Xiaping He performed tumor RNA preparation and microarray experiments for tumor samples. Charles Perou was the Principal Investigator, instigated and designed the study, and helped draft the paper.

Katherine A. Hoadley, Victor J. Weigman, Cheng Fan, Lynda R. Sawyer, Xiaping He, Melissa A. Troester, Carolyn I. Sartor, Thais Rieger-House, Philip S. Bernard, Lisa A. Carey, and Charles M. Perou. Breast Cancer Research [in submission].

ABSTRACT

The epidermal growth factor receptor (EGFR/HER1) protein and its downstream signaling events are important for regulating cell growth and behavior in many epithelial tumors including lung and colon. In breast cancers, the role of EGFR is complex and may vary with estrogen receptor (ER) status. To investigate the role of EGFR signaling in breast cancer, several breast basal-like and luminal epithelial cell lines were examined for sensitivity to EGFR inhibitors gefitinib and cetuximab. We identified an EGFR-activation profile in the basal-like breast cancer cell line SUM102 and analyzed expression of these genes in human breast tumors. The breast basal-like cell lines were generally more sensitive to gefitinib compared to the luminal lines. The basal-like cell line SUM102 was the only cell line sensitive to cetuximab. The basal-like tumor derived lines were also the most sensitive to carboplatin, which acted synergistically with cetuximab. Using SUM102 cells, we identified an EGFR-activation profile that included a strong MEK-dependent signature. The EGFR-activation signature was used to cluster a large panel of breast tumors. Three distinct clusters of genes were evident *in vivo*, two of which were strongly predictive of poor patient outcomes. These two poor prognostic signatures were highly expressed in most basal-like and in approximately half of HER2+/ER- and luminal B tumors. Ninety percent of the basal-like tumors that showed high expression the EGFR-activation profiles also showed high expression of *CRYAB* and/or a *KRAS*-amplicon signature. These data suggest that most basal-like tumors have an EGFR-activation profile, however, few are likely to respond to the direct inhibition of EGFR due to ligand independent activation of the EGFR-RAS-MEK pathway via *CRYAB* or *KRAS* function. Thus, for those tumors that show a ligand

independent EGFR-activation profile, alternative strategies that target downstream components like MEK may prove to be viable alternatives.

INTRODUCTION

The epidermal growth factor receptor (EGFR/HER1) is a member of the human epidermal growth factor receptor (HER) family of transmembrane receptor tyrosine kinases that is linked to growth control, cell adhesion, mobility, and apoptosis (Yarden and Sliwkowski, 2001). As such, EGFR is an important regulator of epithelial cell biology, but its function in breast tumors may vary according to other clinical features like estrogen receptor (ER) status. Microarray studies have identified several subtypes of breast cancer arising from at least two different epithelial cell types (Perou et al., 2000; Sørlie et al., 2001; Sørlie et al., 2003; Hu et al., 2005). One of the molecular subtypes of breast cancer is partly defined by the expression of ER while another is partly defined by the genomic DNA amplification and high expression of HER2 (i.e. HER2+/ER-, see (Hu et al., 2005)). The basal-like subtype has low expression of both ER and HER2. EGFR was found to have high expression in many of the basal-like tumors as assessed by both gene and protein expression (Nielsen et al., 2004).

EGFR overexpression has been reported in a variety of epithelial tumors (Salomon et al., 1995), leading to the development of drugs directed against this receptor (Baselga, 2002; Mass, 2004). One of these targeting strategies employs monoclonal antibodies (cetuximab) that bind the extracellular ligand-binding domain, while other strategies include small molecule inhibitors (gefitinib and erlotinib) that compete with ATP for binding to the intracellular tyrosine kinase domain (Culy and Faulds, 2002; Graham et al., 2004; Dowell et

al., 2005). In non-small cell lung cancer and breast cancer cell lines, it has been shown that some small molecule EGFR inhibitors increase cell killing when used in combination with chemotherapeutics (Ciardiello et al., 2000; Tamura and Fukuoka, 2003); therefore, the interactions between HER1 inhibitors and cytotoxic agents represent a promising combination for the future treatment of epithelial tumors that are dependent upon EGFR-signaling.

The lack of clinical response in breast cancers to gefitinib *in vivo* has been partially attributed to activation of this pathway downstream of EGFR. Several studies have implicated the PI3K/AKT and MEK/ERK pathways as being responsible for EGFR inhibitor resistance due to downstream activation. EGFR-independent activation of the PI3K/AKT pathway may occur through either loss of PTEN or mutation/activation of PI3K, both of which have been linked to gefitinib (Moasser et al., 2001; Bianco et al., 2003; She et al., 2003). Others have suggested that the MEK/ERK pathway may play a more important role in resistance to EGFR inhibitors (Lev et al., 2004; Janmaat et al., 2006; Normanno et al., 2006). Recently, Moyano *et al.* identified αB-Crystallin (CRYAB) as a protein that can constitutively activate the MEK/ERK pathway in breast cancer cells and cause a breast epithelial cell line to become EGF independent (Moyano et al., 2006).

In this study, we used basal-like breast cell lines to examine the EGFR signaling pathway and its interactions with cytotoxic chemotherapy. Using an EGFR-activation profile derived from a basal-like tumor derived cell line, we determined that most basal-like and approximately 50% of Luminal B and HER2+/ER- tumors showed an EGFR-activation profile. EGFR-activation signatures may be useful in selecting patients for therapeutics that target the EGFR-RAS-MEK pathway.

MATERIALS AND METHODS

Cell Culture

SUM102 and SUM149 cells were a gift from Steve Ethier of Wayne State University [http://www.asterand.com/Asterand/BIOREPOSITORY/hbreastcancercelllines.aspx] and represent cell lines derived from ER- and HER2- basal-like breast tumors. The SUM lines were maintained in an Epithelial Growth Medium developed by the Tissue Culture Facility at the University of North Carolina at Chapel Hill [http://www.unc.edu/depts/tcf/info.html], and the SUM149 line was further supplemented with 5% FBS. The MCF-7, ZR-75-1, HME-CC and ME16C cell lines were obtained and maintained as previously described (Troester et al., 2004; Troester et al., 2004).

Cytotoxicity Assay

Cell line sensitivities to drugs were assessed using a mitochondrial dye conversion assay (MTT, Cell Titer 96, Promega) as described previously with the following modifications (Troester et al., 2004). Cells were seeded into triplicate 96-well plates (SUM102, HME-CC, and ME16C – 5,000 cells/well, SUM149 – 10,000 cells/well, MCF-7 and ZR-75-1 – 7,000 cells/well) and allowed to adhere overnight. Cells were treated for 72h with a range of doses of individual drugs. Carboplatin, doxorubicin, 5-fluorouracil, paclitaxel, and LY294002 were purchased from Sigma. Gefitinib was a gift from AstraZeneca and cetuximab was purchased from the UNC Hospitals Pharmacy Storeroom. U0126 was purchased from Cell Signaling. The inhibitory concentration that caused a 50%

reduction in MTT dye conversion (IC50) dose was determined as previously described (Troester et al., 2004).

Drug combination interactions were analyzed using methods developed by Chou and Talalay (Chou and Talalay, 1984). Using cell lines plated as described above, seven treatment combinations consisting of constant ratios of IC50 doses (ranging from one-eighth of each dose to eight times the IC50) were applied to cells and growth compared to untreated controls using the MTT assay. Four treatment schedules were tested: 72h concurrent, 72h inhibitor followed by 72h chemotherapeutic, 72h chemotherapeutic followed by 72h inhibitor, and a 144h concurrent dose with a media change at 72h (similar to the sequential treatments). CalcuSyn (BioSoft) was used to determine the combination index, which is a measurement of the type of drug interactions. A combination index (CI) of one indicates an additive response, less than one indicates a synergistic response (greater than additive), and greater than one indicates an antagonistic response (less than additive).

Collection of mRNA for Cell Line Experiments

For each treatment, the SUM102 cells were grown in 15-cm dishes until 50-60% confluence. SUM102 cells were treated for 48h with a dose equivalent to two times the 72h-IC50 dose of each inhibitor (treated samples). To identify EGFR, MEK, and PI3K activation signatures, medium was removed after 48h of inhibitor treatment and replaced with fresh medium without inhibitor. mRNA was harvested at 4h, 8h, and 24h (post treatment samples). Cells were harvested by scraping, quickly placed into RNA lysis buffer, and mRNA was isolated using the Micro-FastTrack kit (Invitrogen).

Collection of RNA for Human Tumor Samples

248 breast tissue samples represented by 241 fresh frozen breast tumor samples and 7 normal breast tissue samples were obtained from four different sources using IRB approved protocols from each participating institution: the University of North Carolina at Chapel Hill, The University of Utah, Thomas Jefferson University and the University of Chicago; many of these samples have appeared in previous publications (Weigelt et al., 2005; Hu et al., 2006; Oh et al., 2006; Perreard et al., 2006), and 117 are new to this study. Patients were heterogeneously treated in accordance with the standard of care dictated by their disease stage, ER, and HER2 status.

Tumor Sequence Analysis

Tumor genomic DNA samples were isolated from 96 tumors using Qiagen DNeasy Kits according to the manufacturers protocol. Gene resequencing analyses were performed at Polymorphic DNA Technologies (Alameda, CA) using an ABI 3730xl DNA sequencer and cycle sequencing, according to the manufacturers protocol. A two-step "boost/nested" PCR strategy was used where first a PCR reaction is performed to generate a larger DNA fragment, which is then used as a template for the nested reaction with a second set of PCR primers. Double stranded sequencing is performed on the nested product using the nested PCR primers as the sequencing primers. Exons 19 and 21 of *EGFR* were sequenced across all 96 patients, while exons 1 and 2 of *KRAS2*, 1 and 2 of *HRAS*, and 11 and 15 of *BRAF* were sequenced across 54 patients. No somatic alterations were detected in any gene in any sample.

Microarray Experiments

For the human tumor samples, the total RNA isolation and microarray protocols were performed as described in Hu *et al.* (Hu et al., 2005); in this study, a number of tumor samples from previous studies were retested using a new custom Agilent microarray enriched for breast cancer genes. For cell lines experiments, labeled cRNA was generated from the mRNA using Agilent's Low RNA Input Linear Amplification Kit as described in Hu *et al.* (Hu et al., 2005). The 48h treated samples were compared to an untreated cell line reference to look for effects of an inhibitor and to post treatment samples to identify the activation signature for that drug/pathway. Labeled experimental sample (Cy5 CTP) and reference (Cy3 CTP) were mixed and co-hybridized overnight on the same Custom 22K Agilent Human Whole Genome Oligonucleotide Microarray described above. Two to four microarrays per experimental condition were performed, including a dye-flip replicate for gefitinib- and cetuximab-treated samples. Microarrays were scanned on an Axon GenePix 4000B microarray scanner and analyzed using GenePix Pro 5.1 software. Microarray raw data were uploaded into the UNC Microarray Database and Lowess normalization was performed on the Cy3 and Cy5 channels. The microarray and patient clinical data are available at UNC Microarray Database [https://genome.unc.edu/cgi-bin/SMD/publication/viewPublication.pl?pub_no=63] and have been deposited in the Gene Expression Omnibus under the accession number GSE6128.

Statistical Analyses

Intraclass correlations between cell line microarray experiments were performed to judge agreement between replicate experiments as described in Hu *et al.* (Hu et al., 2005).

Unsupervised analyses of the cell line samples were performed by selecting genes with an absolute signal intensity of at least 30 units in both channels in at least 70% of the samples tested and that also showed a Log_2 R/G Lowess normalized ratio of two on at least two arrays. The program Cluster was used to hierarchically cluster samples and genes, and Treeview was used to view the data (Eisen et al., 1998; Eisen and Brown, 1999). Using the SUM102 treated cells, a one-class Significance Analysis of Microarrays (SAM) was used to identify significantly induced genes in all the post treatment experiments (two to three arrays for each experimental timepoint) (Tusher et al., 2001). Gene ontology enrichment was assessed using EASE (Hosack et al., 2003).

Analyses of the primary tumor data used top 500 induced genes from the cell line SAM analysis described above, after filtering for 30 units in both channels in at least 70% of the samples. These genes were examined in a two-way hierarchical clustering analysis with the 248 UNC tumor sample set. Three distinct expression patterns were observed and labeled as Clusters #1-3. Next, the genes in each of these three tumor-defined clusters were identified in the NKI295 patient data set (van de Vijver et al., 2002; Chang et al., 2005), and a mean expression value for each cluster for each patient was determined. The NKI295 patients were then rank-ordered and separated into (a) two equal groups representing low and high, or (b) three equal groups representing low, medium, and high average expression for each cluster. In addition, similar gene-based rank order patient stratifications were performed for individual genes that included *EGFR*, *HER2*, *HER4*, *EGF*, *TGFA*, *AREG*, *CRYAB, KRAS, KRAS*-amplicon profile, *HRAS, NRAS, PIK3CA, PIK3R1, AKT1, AKT2, AKT3, MEK1*, *MEK2, ERK1,* and *ERK2*. Survival analyses were performed using Cox-Mantel log-rank test in Winstat for Excel (R. Fitch Software). Multivariate Cox proportional

hazards analysis was performed in SAS v9.0 (SAS Statistical Software, Cary, NC) to estimate the hazard ratio associated with cluster expression in the three groups after controlling for standard clinical predictors (age, ER status, size, grade, and node status). Chi Square tests (SAS v9.0) was used to examine correlations between cluster groups, individual genes, and tumor subtype.

Gene expression relative levels were visualized in relation to the EGFR signaling pathway using Cytoscape [www.cytoscape.org] (Shannon et al., 2003). The pathway was built *de novo* based on information from KEGG [http://www.genome.ad.jp/kegg/] (Ogata et al., 1999), BioCarta [http://www.biocarta.com], and a review by Yarden and Silowkoski (Yarden and Sliwkowski, 2001) with a focus on the RAS-MEK and PI3K/AKT. Using the UNC breast tumor microarray dataset, an average gene expression profile is displayed for the luminal A, luminal B, basal-like, and HER2+/ER- tumors. Tumor "intrinsic" subtype was determined for each sample using the 306 gene Centroid Predictor described in Hu *et al.* (Hu et al., 2006); the subtype classifications used for the NKI295 sample set were also derived from this same centroid predictor and are described in Fan *et al.* (Fan et al., 2006).

RESULTS

Cell Line Models of Breast Cancer

Breast cancer is a heterogeneous disease arising from at least two distinct epithelial cell populations; therefore, we selected cell lines models of basal-like and luminal cells to begin our investigations of the EGFR-pathway. The MCF-7 and ZR-75-1 cell lines were derived from breast tumors of luminal origin and have expression of CK8/18 and ER. Our previous studies examining cell lines of basal-like origin used immortalized human

mammary epithelial cell lines (HMECs) (Troester et al., 2004; Troester et al., 2004); however, these lines are derived from normal rather than tumor tissue. Two ER-negative and HER2-non-amplified tumor-derived cell lines, SUM149 and SUM102, have been previously shown to express EGFR (Sartor et al., 1997; Lev et al., 2004) and show basal-like profiles (Bertucci et al., 2005). The SUM102 and SUM149 lines share many characteristics with the basal-like tumors including expression of CK5/6; therefore, we used these two tumor-derived lines as *in vitro* models of basal-like breast cancers. By microarray analysis, EGFR gene expression was very low in the luminal cell lines and higher in the basal-like lines. EGFR protein expression by Western blot analysis was detectable in basal-like, but not in the luminal lines (data not shown).

Drug Sensitivity Assays

To assess EGFR inhibitor sensitivity, the six cell lines were treated for 72h with a range of doses of gefitinib or cetuximab and an MTT assay was used to determine IC50 doses (Table 4.1). In response to gefitinib, the basal-like tumor-derived cell lines (SUM149 and SUM102) were two- to 100-fold more sensitive than the luminal lines. The two immortalized HMEC lines were also 33- and 50-fold more sensitive to gefitinib than the luminal lines, suggesting that the basal-like cell type as a whole is more sensitive to gefitinib than the luminal cell type. Cetuximab sensitivity was observed only in a single cell line (SUM102, IC50=2ug/ml), with IC50 doses for MCF-7, ZR-75-1, SUM149, ME16C2, and HME-CC not achievable even with cetuximab doses as high as 100ug/ml. These cell lines were also treated with inhibitors that affect targets downstream of EGFR in its pathway including U1026 (MEK1/2 inhibitor) and LY294002 (PI3K inhibitor). Most of the cell lines

Table 4.1. Estimated IC50 doses of six breast cell lines for the EGFR inhibitors gefitinib, cetuximab, the MEK1/2 inhibitor U0126, and the PI3K inhibitor LY294002

Cell Line	Gefitinib (μM)	Cetuximab (μg/mL)	U0126 (uM)	LY294002 (uM)
ME16C	0.3 (0.02)	>100[a]	19.7 (0.66)	21.2 (0.63)
HME-CC	0.2 (0.01)	>100[a]	12.7 (0.33)	7.3 (0.17)
SUM102	0.1 (0.002)	2.3 (0.15)	4.3 (0.20)	3.4 (0.10)
SUM149	4.7 (0.14)	>100[a]	21.8 (0.80)	18.4 (0.48)
MCF-7	21.1 (0.29)	>100[a]	17.0 (1.15)	3.9 (0.13)
ZR-75-1	11.1 (0.12)	>100[a]	25.0 (0.74)	2.4 (0.05)

Note that the standard errors are presented within ()
[a]No achievable IC50 dose with doses up to 100μg/mL

Table 4.2. Estimated IC50 doses of six breast cell lines treated with chemotherapeutics

Cell Line	5-Florouracil (uM)	Doxorubicin (nM)	Carboplatin (uM)	Paclitaxel (nM)
ME16C	6.0 (0.29)	32.8 (1.89)	37.5 (0.63)	0.052 (0.004)
HME-CC	1.1 (0.07)	35.5 (3.26)	48.3 (1.41)	0.025 (0.003)
SUM102	16.8 (0.82)	5.1 (0.27)	11.7 (0.26)	0.00057 (0.00001)
SUM149	28.6 (1.33)	45.0 (3.06)	7.7 (0.24)	0.71 (0.006)
MCF-7	1.2 (0.15)	56.9 (4.26)	89.4 (3.79)	0.23 (0.02)
ZR-75-1	8.4 (1.06)	26.5 (1.39)	62.6 (1.98)	0.99 (3.34)

Note that the standard errors are presented within ().

had a similar level of sensitivity to U0126 with the exception that SUM102 was approximately 5-fold more sensitive. IC50 doses for LY294002 were similar for most lines with the exception of ME16C and SUM149 cells, which were approximately 5-fold more resistant than the other lines. The SUM102 line was the only cell line that was sensitive to all four inhibitors and has previously been shown to be EGFR-dependent (Sartor et al., 1997), and thus was chosen for further analyses of the EGFR-pathway.

Drug Combination Analyses

A phase II clinical trial is currently recruiting breast cancer patients who are ER-negative, Progesterone Receptor-negative, and HER2 non-amplified (i.e. basal-like patients) to assess treatment responses to cetuximab alone or in combination with carboplatin [http://www.clinicaltrials.gov/ct/show/NCT00232505]. A second phase II trial in a nonselected population of metastastic breast cancer patients is also evaluating cetuximab in combination with carboplatin and irinotecan [http://www.clinicaltrials.gov/ct/show/NCT00248287]. Therefore, we examined the combined effects of cetuximab and carboplatin, as well as three additional chemotherapeutics (doxorubicin, 5-fluorouracil, and paclitaxel), in SUM102 cells. We also tested the combined effects of gefitinib, U0126, and LY294002 with chemotherapeutic agents. Individual drug sensitivity (IC50 doses) for each chemotherapeutic was determined for all six cell lines (Table 4.2). The relative sensitivities varied across the cell lines and did not appear to correlate with subtype, with the exception of two basal-like tumor-derived cell lines (SUM102 and SUM149) that were at least three-fold more sensitive to carboplatin and at

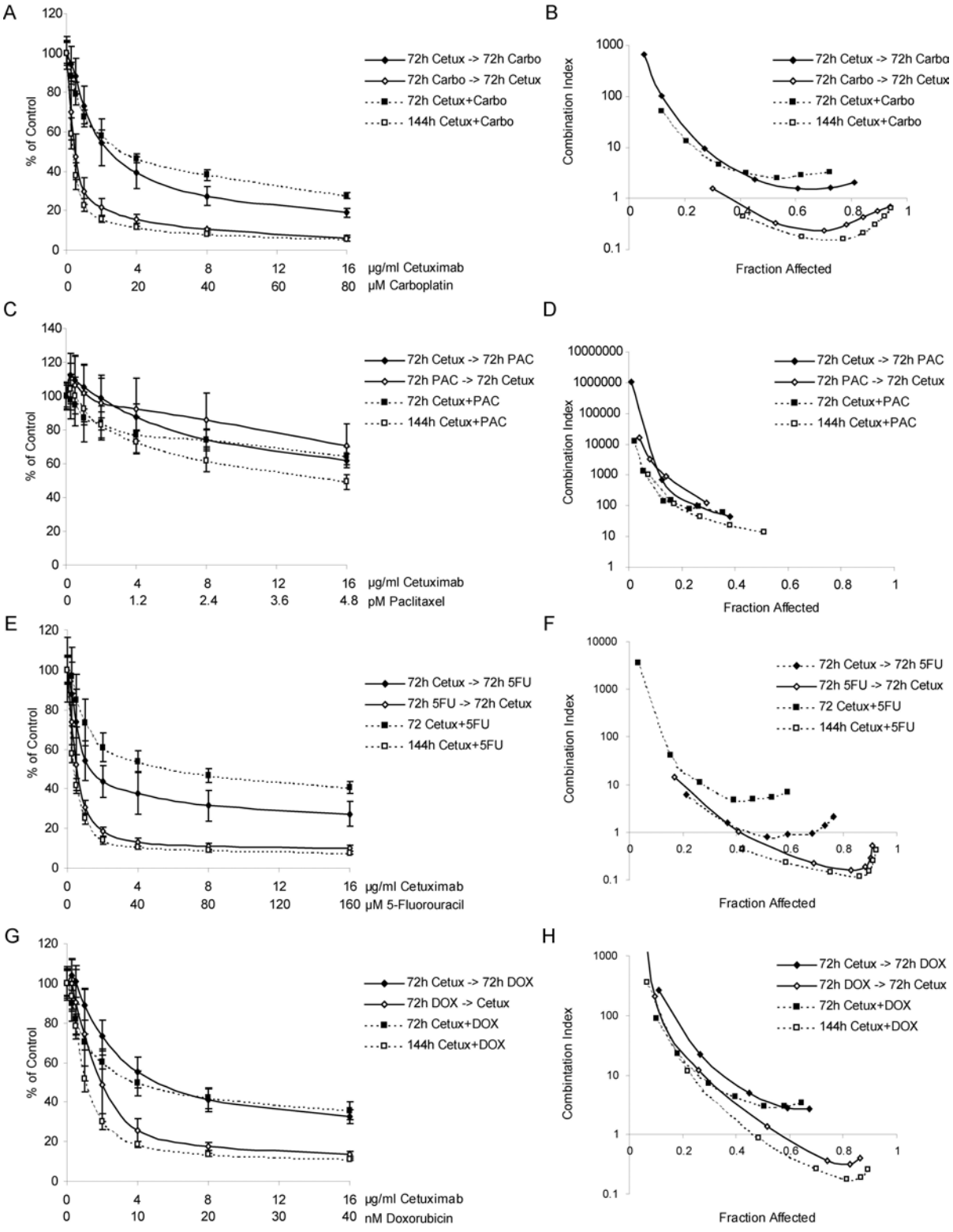
A
% of Control
72h Cetux -> 72h Carbo
72h Carbo -> 72h Cetux
72h Cetux+Carbo
144h Cetux+Carbo
µg/ml Cetuximab
µM Carboplatin
B
Combination Index
Fraction Affected
C
% of Control
72h Cetux -> 72h PAC
72h PAC -> 72h Cetux
72h Cetux+PAC
144h Cetux+PAC
µg/ml Cetuximab
pM Paclitaxel
D
Combination Index
Fraction Affected
E
% of Control
72h Cetux -> 72h 5FU
72h 5FU -> 72h Cetux
72 Cetux+5FU
144h Cetux+5FU
µg/ml Cetuximab
µM 5-Fluorouracil
F
Combination Index
Fraction Affected
G
% of Control
72h Cetux -> 72h DOX
72h DOX -> Cetux
72h Cetux+DOX
144h Cetux+DOX
µg/ml Cetuximab
nM Doxorubicin
H
Combintation Index
Fraction Affected

Figure 4.1

Figure 4.1. Effects of different combination schedules of cetuximab with chemotherapeutics in SUM102 cells. Cells were treated with four different combination schedules: 1) 72h cetuximab followed by 72h chemotherapy, 2) 72h chemotherapy followed by 72h cetuximab, 3) 72h concurrent chemotherapy and cetuximab, and 4) 144h concurrent chemotherapy and cetuximab. *A*) Growth inhibitory effects of cetuximab and carboplatin combinations. *B*) Combination analysis of cetuximab and carboplatin treatments. *C*) Growth inhibitory effects of cetuximab and paclitaxel combinations. *D*) Combination analysis of cetuximab and paclitaxel treatments. *E*) Growth inhibitory effects of cetuximab and 5-fluorouracil combinations. *F*) Combination analysis of cetuximab and 5-fluorouracil treatments. *G*) Growth inhibitory effects of cetuximab and doxorubicin combinations. *H*) Combination analysis of cetuximab and doxorubicin treatments. Combination Index (CI) values below one are synergistic, equal to one are additive, and greater than one are antagonistic.

least two-fold more resistant to 5-fluorouracil when compared to their "normal" HMEC counterparts or the luminal cell lines.

The interaction of cetuximab with a chemotherapeutic in combination was examined solely in the SUM102 line because this was the only cetuximab sensitive line. As a starting point, we treated SUM102 cells for 72h with cetuximab and a chemotherapeutic simultaneously. Synergistic interactions were not evident in any combination; all combinations were highly antagonistic as assessed by the method of Chou and Talalay in CalcuSyn (Chou and Talalay, 1984) (Figure 4.1). We next analyzed the effect of sequential treatment: cells were treated for (a) for 72h with cetuximab followed by 72h with chemotherapy, (b) for 72h with chemotherapy followed by 72h with cetuximab, or (c) with cetuximab and chemotherapy simultaneously for 144h. Chemotherapy followed by cetuximab was generally more growth inhibitory than cetuximab followed by chemotherapy (Figure 4.1A-H). The one exception was cetuximab with paclitaxel, where all sequence combinations were antagonistic (Figure 4.1C and D). However, this antagonism may result from the high sensitivity to paclitaxel already observed in the SUM102 line. Carboplatin followed by cetuximab and the 144h concurrent treatments were synergistic even at low doses of both drugs. 5-fluorouracil followed a similar trend to that of carboplatin, while in the doxorubicin combinations synergy was only evident at doses higher than the IC50 dose for doxorubicin first or the 144h concurrent (Figure 4.1F and H). Similar results were observed for combinations with gefitinib and LY294002 (a PI3K inhibitor) where chemotherapy followed by each inhibitor treatment and the 144h concurrent treatments were more effective than the inhibitor first. U0126 (a MEK inhibitor) combinations were different with chemotherapy first followed by U0126 being slightly less synergistic than the U0126

first or concurrent treatment; however, for U0126, all combinations except doxorubicin first, or paclitaxel first, were synergistic.

EGFR-Pathway Gene Expression Patterns

To identify EGFR-dependent transcriptional patterns, we analyzed the gene expression data of the SUM102 cell line treated with EGFR inhibitors and then released from this inhibition. Using an unsupervised analysis, we hierarchically clustered all time points from the cetuximab and gefitinib treatment experiments and identified over 500 genes that changed in expression at least 4-fold (Figure 4.2). Even though the two EGFR inhibitors have different mechanisms of inhibition, SUM102 cells treated for 48h with gefitinib or cetuximab showed very similar gene expression changes. Intraclass correlation (ICC) values between the gefitinib and cetuximab treated samples ranged from 0.627 to 0.934, and this level of similarity was also evident in the short dendrogram branches from the cluster analysis (Figure 4.2B). The post treatment samples that represent the reactivation of the EGFR-pathway were even more similar (intraclass correlations within each time point ranged from 0.862 to 0.962). A two-class SAM analysis to look for differences between gefitinib-post treatment samples versus cetuximab-post treatment samples identified only 58 significantly different genes with a false discovery rate (FDR) of 5%; thus, from a transcription standpoint, gefitinib and cetuximab elicited very similar results.

In response to gefitinib and cetuximab, the SUM102 cell line exhibited decreased expression of many proliferation genes (Figure 4.2G). There was also a large cluster of transcripts that were induced by the inhibitors, consisting predominately of hypothetical genes with unknown functions. We were more interested in the genes induced by removal of

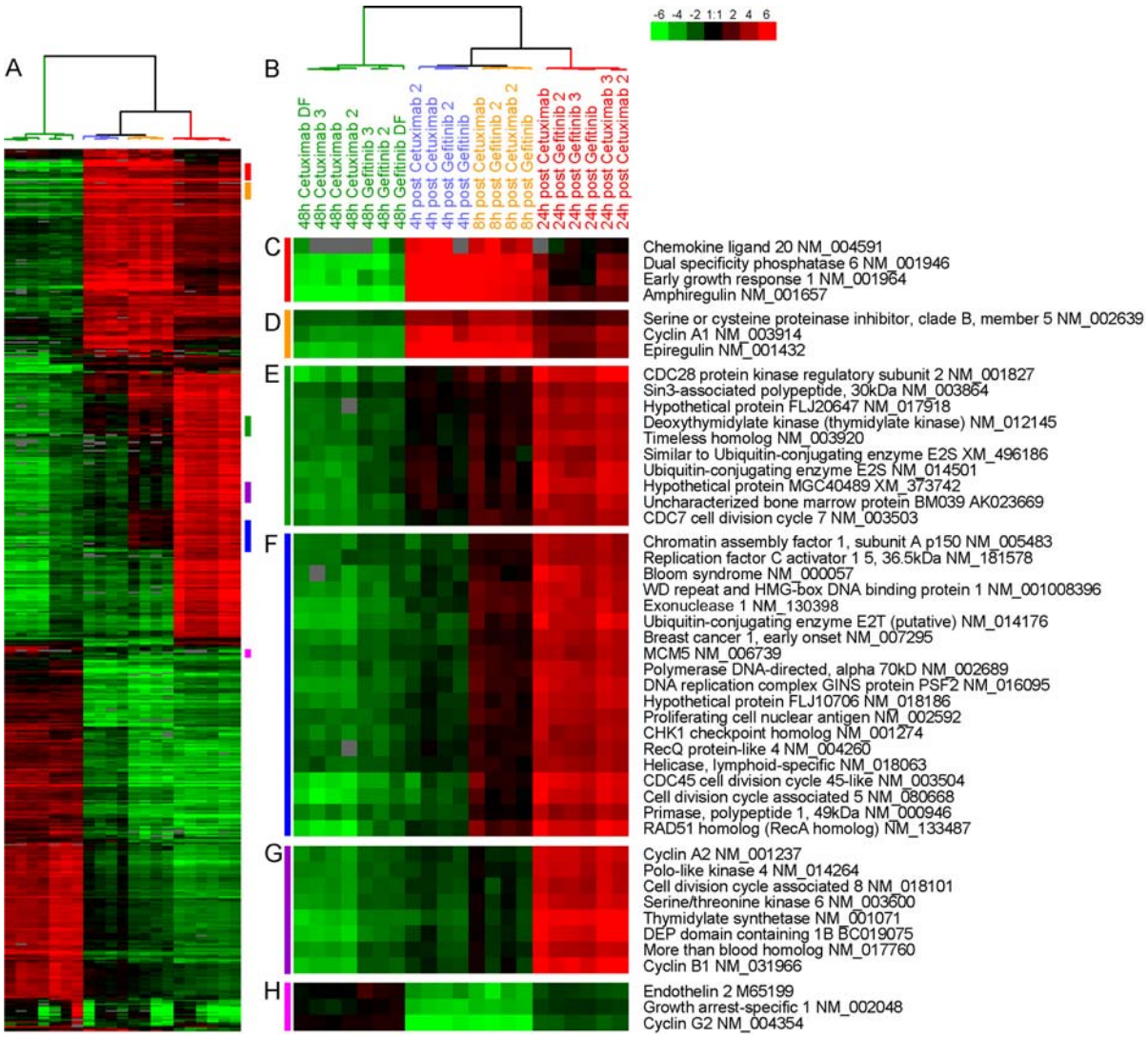

Figure 4.2. Gene expression patterns for SUM102 cells treated with gefitinib or cetuximab. Unsupervised hierarchical cluster analysis was performed on 48h inhibitor treated and 4h, 8h, and 24h post inhibitor treated samples. *A*) The complete cluster overview with the colored bars indicating the location of the clusters shown in *C*-*G*. *B*) Close up of the experimental sample associated dendrogram. *C*+*D*) 4h and 8h post treatment induced genes including the EGFR ligands *Amphiregulin* and *Epiregulin*. *E*) Genes involved with the G1/S phase transition induced beginning in the 4h post inhibitor and continuing though 24h. *F*) Genes involved in DNA synthesis induced at 8h post inhibitor and continuing through 24h. *G*) Proliferation genes typically observed in tumor derived profiles including *STK6* and *Cyclin B1*. *H*) Negative regulators of growth.

the inhibitor as this reflects the gene expression associated with *de novo* activation of the EGFR-pathway. As early as 4h and 8h after inhibitor removal there was a substantial increase in expression for two ligands in the EGFR pathway, *amphiregulin* and *epiregulin*. *Cyclin A1* was also substantially increased (Figure 4.2C and D). Starting at 4h and continuing through 8h and 24h, genes with known roles in G1/S phase such as *CDC6*, *CDC*7, *TIMELESS*, and *ORCL6* were increased (Figure 4.2E). By 8h and 24h, DNA synthesis and DNA damage checkpoint genes were induced (Figure 4.2F). Classical gene expression-defined proliferation genes including *STK6* and *Cyclin B1* were highly induced by 24h (Figure 4.2G). There was also a repression of negative regulators of growth such as *Growth arrest-specific 1* and *Cyclin G2* (Figure 4.2H).

To objectively identify an EGFR-activation signature from the SUM102 cells, a one-class SAM analysis was used to identify genes that were statistically induced in the post treatment samples. Adjusting the SAM delta value to obtain the largest gene list with less than 5% FDR resulted in a gene list that was extremely large (10,017 genes, 4.97% FDR), therefore, the top 500 induced genes were selected for further analysis (0.02% FDR). This gene list was used to cluster 248 UNC breast tumor samples representing all five breast tumor subtypes (Figure 4.3). The list of induced genes from the *in vitro* experiments were not homogenously expressed across the tumor samples, and therefore to study these multiple expression patterns in the tumors we defined "clusters" as any gene set that contained a minimum of 20 genes and a Pearson node correlation greater than 0.55. Using this criteria, we identified three clusters: Cluster #1 was high in a mix of breast tumor samples that contained all five breast cancer subtypes: luminal B, luminal A, basal-like, HER2+/ER- and normal-like samples (Figure 4.3C, far right dendrogram branch, 35 genes); Cluster #2

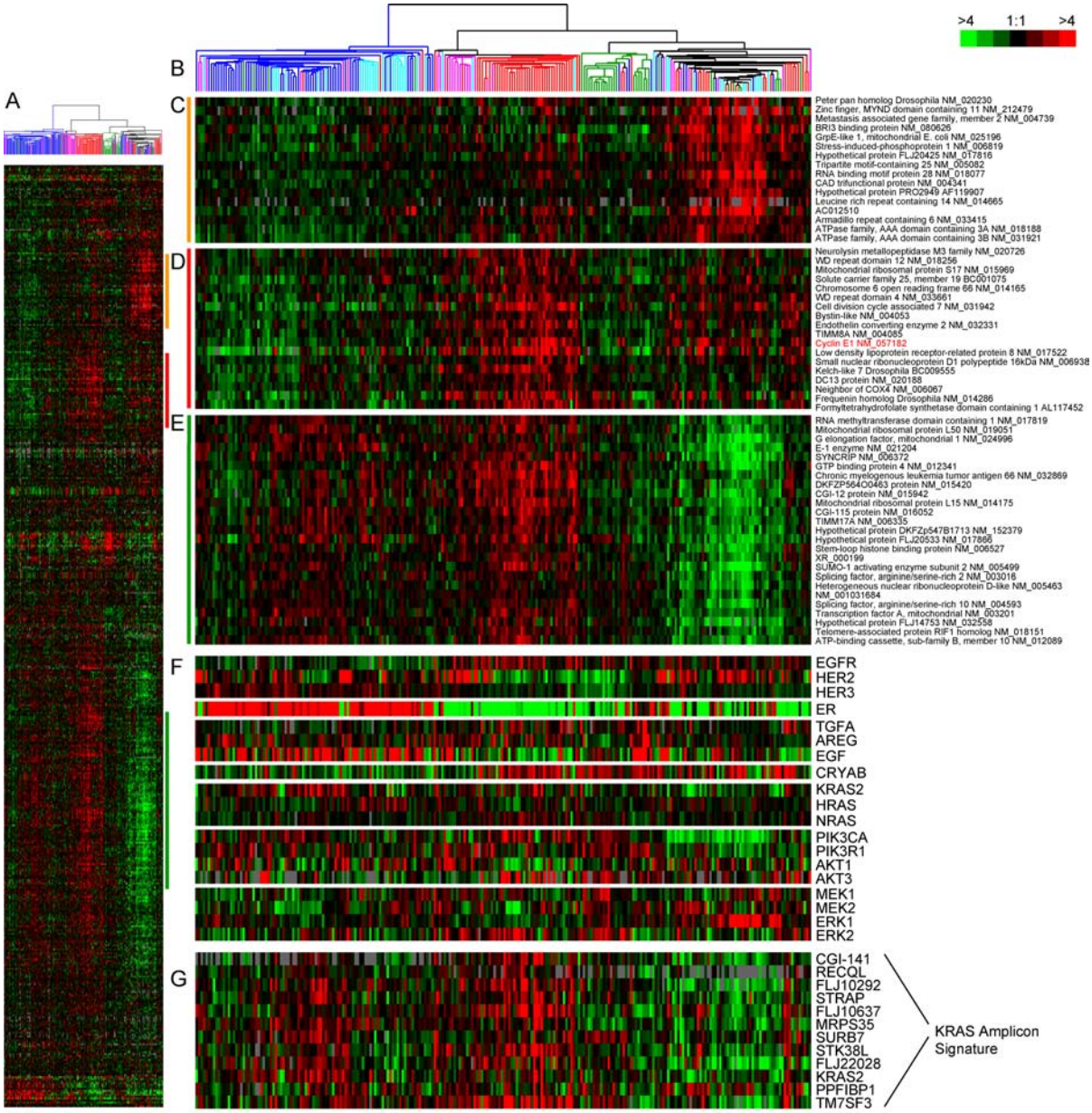

Figure 4.3. *In Vivo* EGFR-activation profiles and additional genes implicated in the EGFR-RAS-MEK pathway clustered on the UNC tumor data set. *A*) The top 500 induced genes from the SUM102 post treatment experiments were hierarchical clustered using the 248 UNC tumors. Colored bars indicate the location of the three clusters in *D-E*. *B*) Tumor associated dendrogram color coded according to tumor subtype: luminal A – dark blue, luminal B – light blue, true normals and normal-like – green, HER2+/ER- - pink, and basal-like – red. *C*) Cluster #1 that identified a mixed group of tumors. *D*) Selected genes from the center of Cluster #2 that is high in most basal-like tumors. *E*) Selected genes from the center of Cluster #3 that is high in the luminal tumors. *F*) Data for genes with suggested roles in HER pathway ordered according to hierarchical clustering with the 500 induced genes. *G*) Data for the *KRAS*-amplicon signature identified in Herschkowitz *et al.* (Herschkowitz et al.).

identified a set of tumors that was highly enriched for basal-like tumors and contained 58% of all basal-like tumors, 48% of all HER2+/ER- tumors and 3 Luminal B tumors (Figure 4.3D, center dendrogram branch, 27 genes); Cluster #3 was highly enriched for Luminal A and B tumors , as well as the HER2+/ER- and basal-like tumors that were also high for Cluster #2 (Figure 4.3E, left dendrogram branch– luminal A and B tumors, and center dendrogram branch – HER2+/ER- and basal tumors, 139 genes). Thus each gene cluster represents a stereotyped EGFR-activation signature that is enriched in a different subset of tumors. Full gene lists for each cluster are in Appendix IVA. Gene Ontology (GO) analysis using EASE was performed on each gene cluster but only Cluster #3 had any significant GO terms, which were RNA processing, metabolism, binding, splicing, and modification (EASE scores < 0.05). However, *Cyclin E1* was present within Cluster #2 and is a known prognostic marker for breast cancer patients (Schraml et al., 2003); *Cyclin E1* is also associated with basal-like breast cancers (Foulkes et al., 2004; Sieuwerts et al., 2006), which was recapitulated here, and known to be regulated by EGFR-signaling (Lu et al., 2003).

To examine the biological importance of these three gene sets, we individually applied them to a true test set of breast tumors (i.e. the NKI295 sample set described in (van de Vijver et al., 2002; Chang et al., 2005)) to determine whether they predicted patient outcomes. We first determined a mean expression value of all genes within a cluster for each patient. The patients were next rank-ordered based upon the mean expression values and divided into either two groups or three groups based upon their rank-order mean expression values. Kaplan-Meier survival analyses for Relapse-Free Survival (RFS) and Overall Survival (OS) were performed and all three clusters were statistically significant predictors of outcomes where the high expression always predicted a poor outcome (Figure 4.4 – OS;

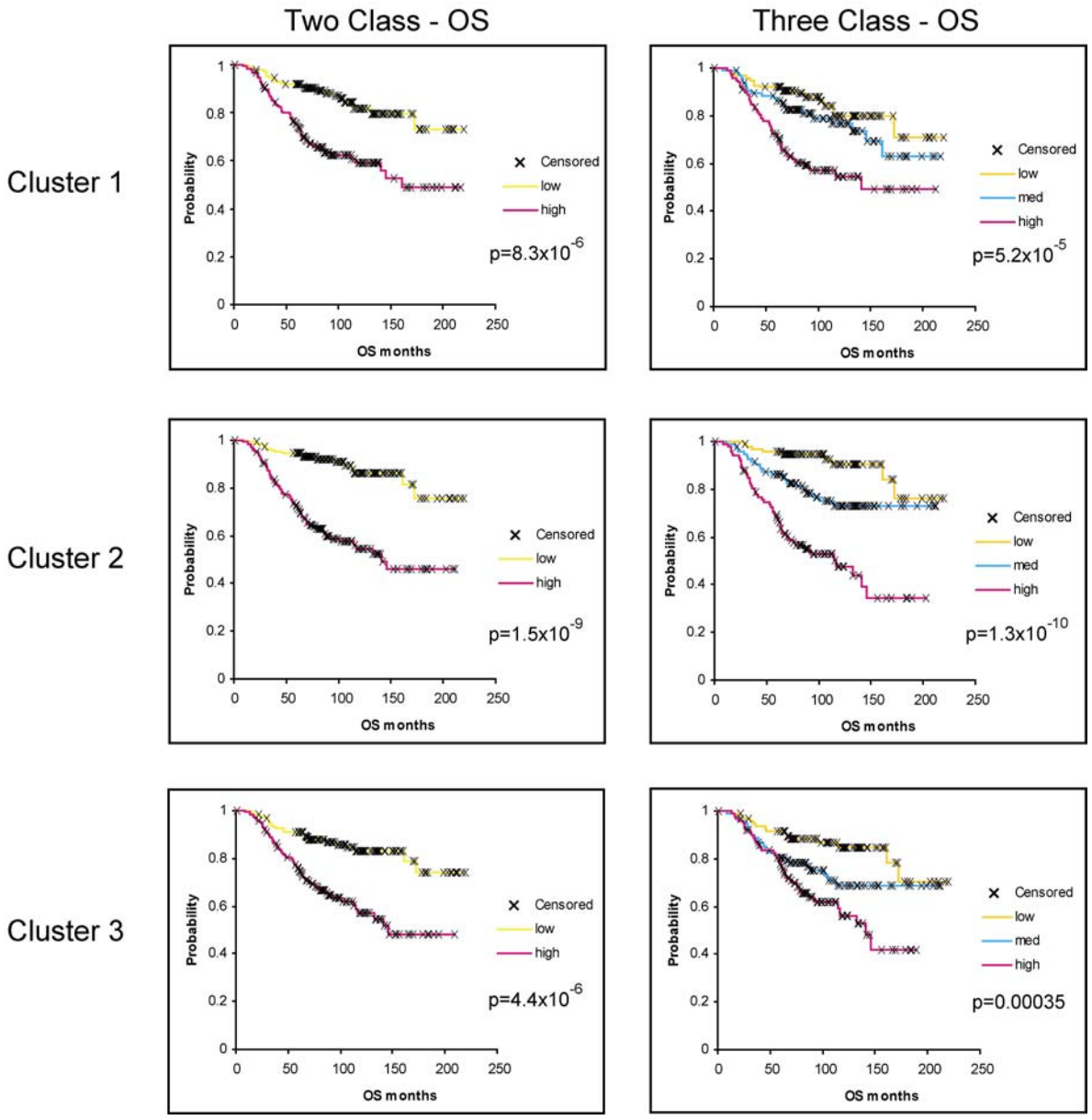

Figure 4.4. Kaplan-Meier survival plots for the 295 NKI tumors/patients using expression from the three different *in vivo* defined EGFR-activation profiles. The average expression value for each cluster in each patient was determined and the patients then put into rank-order and divided into two equal groups or three equal groups. Overall survival analysis was performed for each cluster. X indicates censored data due loss to follow-up or to information at last checkup. Note that Clusters #2 and #3 were also similarly prognostic for the UNC training data set presented in Figure 4.3.

similar data observed for RFS). Using a Cox multivariate analysis, we tested each group with the standard clinical parameters and determined that the high expression (top third) of Cluster #2 compared to the lowest expression (bottom third) significantly predicted a worse outcome for both RFS and OS (HR 2.63, 95% CI 1.44-4.79, p=0.0016 and HR 3.46, 95% CI 1.58-7.59, p=0.0019, respectively) after controlling for age, ER status, size, grade, and node status. Chi-squared analyses were performed to identify relations between tumor subtypes and Clusters #1-3. Consistent with observations from Figure 4.3, the basal-like, luminal B, and HER2+/ER- tumors were associated with the high expression of all three clusters while the luminal A and normal-like samples rarely showed high expression (Table 4.3, p=<0.0001); in particular, the basal-like tumors were almost all high for Cluster #2 (89% in top 1/3).

Role of MEK and PI3K in the EGFR-Profile

Activation of EGFR leads to the downstream activation of numerous signaling pathways including the MEK/ERK and PIK3/AKT pathways (Yarden and Sliwkowski, 2001). To examine the role of these effectors, we treated the SUM102 cell line with the MEK1/2 inhibitor U0126 and the PI3K inhibitor LY294002 alone and in combination. Microarray time course experiments using inhibitor treated cells followed by inhibitor removal were conducted for U0126 and LY294002 using the experimental protocol as were done for cetuximab and gefitinib. The observed gene expression profiles for the MEK and the PI3K experiments were similar in both gene identity and direction when compared to the EGFR-profile, but gene expression changes were reduced in magnitude. The MEK and PI3K signatures were very similar to each other at the 4h and 8h time points (average ICC = 0.83), but gave diverged at 24h (average ICC = 0.59). Gene expression signatures of LY294002 and U0126 samples were also correlated with gefitinib and cetuximab gene expression signatures at 4h and 8h post treatment (LY294002 compared to gefitinib/cetuximab ICC = 0.83, U0126 compared to gefitinib/cetuximab ICC = 0.77). The LY294002 and U0126 24h post treatment samples were less correlated with gefitinib and cetuximab 24h post samples (LY294002 compared to gefitinib/cetuximab ICC = 0.51, U0126 Compared to

gefitinib/cetuximab ICC = 0.41). We also treated cells with LY294002 and U0126 together to determine if the combined treatment would more completely recapitulate the EGFR activation profile; the 24h post combined treatment samples showed a higher correlation value to the gefitinib and cetuximab samples (average ICC = 0.73), but still did not account for the entire gene expression pattern of the 24h post cetuximab and gefitinib treatments. These results suggest that the EGFR-profile could not be simply attributed to either the MEK or PIK3 pathway, but the combination of these two pathways or other downstream signaling pathways such as STATs (Sartor et al., 1997) are more representative of the EGFR-signature than either pathway alone.

Potential Mechanisms for Activation of EGFR Signaling *In Vivo*

Activation of the EGFR-RAS-MEK pathway is known to occur via both ligand dependent and independent mechanisms. The empirically derived signatures in Clusters #1-3 (above) are likely to include both. Thus, to distinguish between ligand dependent and independent mechanisms, we tested the gene expression patterns of the HER family of receptors (*EGFR, HER2, HER4*), some of their ligands (*TGFA, EGF, AREG*), as well as other pathway components including *MEK1, MEK2, PIK3CA, PIK3R1, CRYAB, AKT1-3*, the Ras proteins (H, K and N), *ERK1*, *ERK2*, and the *KRAS*-amplicon signature (identified and defined by gene expression in Herschkowitz et al. (Herschkowitz et al.)) for their ability to predict patient outcomes, for correlations with tumor subtype (Table 4.3), and for correlations with Clusters #1-3 (Table 4.4). In this case, the 'ligand independent' gene list was not empirically defined, but based on existing literature for the activation of the EGFR signaling pathway. Gene expression for individual genes was rank-ordered and divided into

thirds as was done for Clusters #1-3 above. Each gene was tested for its ability to affect survival outcome in our tumor data. No genes significantly predicted RFS and OS in both the UNC and NKI data sets. Associations between genes, or Clusters #1-3, with subtype were determined by Chi-square analysis and identified many significant associations (Table 4.3). For example, high *HER2* expression, as expected, was significantly correlated with the HER2+/ER- subtype and *ER* expression was associated with both luminal subtypes (data not shown). *EGFR* expression was correlated with the basal-like subtype while high *HER4* and *PIK3R1* expression was associated with the luminal A subtype. Many other associations with the basal-like subtype were also evident that included the high expression of Clusters #1-3, *TGFA, AKT3*, *CRYAB, MEK1, NRAS* and the *KRAS*-amplicon signature and gene (Table 4.3). Other potentially biologically relevant associations included the high expression of Clusters #2 and #3, *HRAS, MEK1*, and *AKT1* in the HER2+/ER- subtype, and high expression of Clusters #1-3 and *HRAS* with the luminal B subtype.

We also tested for associations between the high expression of Clusters #1-3 with the high expression (i.e. top 1/3 group) of each of the above-mentioned genes in both the UNC and NKI datasets (Table 4.4). In both datasets, the high expression of *MEK2* and *HRAS* was associated with Cluster 1, while the high expression of many other genes correlated with Clusters 2 and 3; of note was the high expression of the *KRAS*-amplicon, *HRAS, NRAS,* and

Table 4.3. Chi-square analysis for association of gene expression with subtypes. Samples were rank ordered into three equal groups and the percentage of each subtype in the highest expression group is reported for the NKI patient data set.

	Basal-like	HER2+/ER-	Luminal A	Luminal B	Normal-like	p-value
# tumors	53	35	123	55	29	
Cluster 1[a]	68%	37%	12%	56%	14%	<0.0001
Cluster 2[a]	89%	49%	5%	49%	7%	<0.0001
Cluster 3[a]	77%	51%	11%	47%	0%	<0.0001
EGFR[a]	68%	20%	27%	18%	41%	<0.0001
HER2[a]	15%	100%	28%	26%	24%	<0.0001
HER4*	9%	3%	50%	38%	31%	<0.0001
TGFA[b]	74%	37%	17%	25%	38%	<0.0001
AREG[a]	3%	34%	43%	35%	41%	<0.0001
EGF	17%	40%	37%	36%	31%	0.23
CRYAB[a]	70%	11%	33%	4%	48%	<0.0001
KRAS amplicon[a]	68%	40%	24%	35%	0%	<0.0001
KRAS gene[c]	32%	37%	33%	38%	21%	0.3555
HRAS[d]	32%	66%	17%	64%	7%	<0.0001
NRAS[a]	70%	28%	17%	44%	21%	<0.0001
PIK3CA	30%	17%	36%	36%	41%	0.28
PIK3R1[a]	21%	14%	42%	25%	55%	0.0012
AKT1[a]	26%	63%	27%	40%	24%	<0.0001
AKT2*	26%	40%	27%	47%	38%	0.26
AKT3[a]	51%	14%	39%	9%	45%	<0.0001
MEK1	53%	46%	25%	29%	24%	0.0232
MEK2[e]	42%	43%	25%	42%	24%	0.0683
ERK1[f]	30%	26%	31%	42%	41%	0.4924
ERK2[g]	40%	31%	26%	45%	31%	0.0482

*Note: HER4 could not be assessed in UNC data due to too many missing values; HER3 was not present in the NKI data set; AKT2 was not present in the UNC data set

[a] associations were also similarly significant in the UNC sample set

[b] nominally significant in UNC data (p-value=0.0046)

[c] nominally significant association in the UNC data (p-value= 0.0051)

[d] nominally significant in the UNC data (p-value = 0.003)

[e] nominally significant in the UNC data (p-value = 0.0023)

[f] significant in the UNC data (p-value = 0.0003)

[g] significant in the UNC data (p-value = <0.0001)

Bonferroni corrected level of significance α=0.0022

Table 4.4. Associations between Clusters #1-3 and individual genes using the NKI295 sample set. Chi-squared analyses were used to identify associations between the high expression of the individual EGFR-activation profiles for each cluster (top 1/3) and the expression of individual genes categorized as high (top 1/3). The % of tumors with the high expression of each cluster and that show the high expression of the individual gene is shown.

	Cluster 1		Cluster 2		Cluster 3	
	%	p-val	%	p-val	%	p-val
EGFR	39%	0.1783	43%	0.0091[b]	38%	0.1484
HER2	26%	0.0017	25%	<0.0001[c]	24%	<0.0001[a]
HER4*	21%	<0.0001	12%	<0.0001	18%	<0.0001
TGFA	40%	0.0665	48%	0.0002	47%	0.0021
AREG	22%	0.0007[c]	23%	<0.0001[a]	28%	0.0644[f]
EGF	35%	0.1380	25%	0.0691	27%	0.0332[d]
CRYAB	35%	0.3214[f]	38%	0.0524	38%	0.0013
KRAS amplicon	38%	0.1973[e]	52%	<0.0001[c]	63%	<0.0001[a]
KRAS gene	27%	0.0022[a]	31%	0.8795	36%	0.1377[e]
HRAS	48%	<0.0001[c]	51%	<0.0001	47%	0.0018
NRAS	45%	0.0362	56%	<0.0001[c]	59%	<0.0001[a]
PIK3ca	22%	0.0032[b]	27%	0.1415[e]	30%	0.3304[e]
PIK3R1	24%	0.0009[a]	20%	<0.0001[a]	19%	<0.0001
AKT1	41%	0.0112	39%	0.0899	34%	0.3615
AKT2*	40%	0.0519	37%	0.3524	33%	0.9378
AKT3	26%	0.0004	33%	0.1569	35%	0.6377[f]
MEK1	39%	0.0335	47%	0.0032[d]	48%	<0.0001
MEK2	58%	<0.0001[a]	44%	0.0113[d]	36%	0.5519[f]
ERK1	37%	0.0718[e]	23%	0.0009[c]	19%	<0.0001[a]
ERK2	39%	0.0238	37%	0.3457[e]	36%	0.4601[e]

*Note: HER4 could not be assessed in UNC data due to too many missing values; HER3 was not present in the NKI data set; AKT2 was not present in the UNC dataset.

[a] the statistically significant association was also significant in the UNC data set ($p<0.0025$).

[b] the association was nominally significant in the NKI dataset ($p<0.05$), but significant in the UNC dataset ($p<0.0025$).

[c] the association was significant in the NKI dataset ($p<0.0025$), but nominally significant in the UNC dataset ($p<0.05$).

[d] the association was nominally significant in both datasets ($p<0.05$).

[e] the association was significant in UNC dataset ($p<0.0025$).

[f] the association was nominally significant in the UNC dataset ($p<0.05$).

Bonferroni corrected level of significance $\alpha=0.0025$

MEK1 with both Clusters #2 and #3, and the high expression of *EGFR* with only Cluster #2. The association of different genes with the three EGFR-activation signatures represents the heterogeneity of this signaling pathway in breast cancer.

Lastly, an obvious mechanism for activation of the EGFR-RAS-MEK pathway is the somatic mutation of a *RAS* gene, *BRAF*, or *EGFR* itself, which are relatively frequent events in non-small cell lung carcinomas. We performed resequencing analyses on a subset of the breast tumors analyzed here by microarray for *EGFR* Exons 19 and 20, which contain the ATP binding domain of EGFR, and for the common sites of mutation in *HRAS, KRAS* and *BRAF*. No somatic sequence variants were detected in 96 tumors including 54 tumors that were oversampled for basal-like and HER2+/ER- respectively, thus suggesting that the somatic mutation of these genes in not likely to be responsible for the activation of the EGFR-RAS-MEK pathway in breast tumors.

DISCUSSION

The epidermal growth factor receptor family is of tremendous biological and clinical importance for many solid epithelial tumors. Numerous drugs have been developed that target this pathway including trastuzumab that targets HER2 (Slamon et al., 2001), cetuximab that targets EGFR (Adams and Weiner, 2005), and various small molecular inhibitors like gefitinib and erlotinib that have high specificity for EGFR1 and may also inhibit HER2 with varying efficiencies (Akita and Sliwkowski, 2003; Anido et al., 2003) as well as small molecular inhibitors such as lapatinib that specifically target both EGFR and HER2 (Johnston and Leary, 2006). In non-small cell lung cancers, the somatic mutation of EGFR/HER1 leads to a receptor that is hypersensitive to the small molecule EGFR inhibitors

gefitinib and erlotinib (Lynch et al., 2004; Paez et al., 2004), thus providing a biological target that is present in some lung cancer patients. In breast cancer patients, the response rate to single agent EGFR inhibitors has been low, however, these trials were performed on unselected patient populations (Agrawal et al., 2005; Normanno et al., 2005).

The EGFR-pathway has recently become a potential target in the basal-like subtype because it was shown that at least 50% of basal-like tumors express EGFR as assessed by IHC (Nielsen et al., 2004). These results, in part, lead to the initiation of a clinical trial for ER-negative, PR-negative, and HER2-nonamplified (i.e. basal-like) breast cancers where these patients will receive cetuximab alone versus cetuximab plus carboplatin. Our *in vitro* analyses show that all four basal-like cell lines were more sensitive to EGFR inhibitors compared to luminal cell lines. In addition, only a single cell line (SUM102) was sensitive to cetuximab when EGF was present within the media, which is the condition that best mimics the *in vivo* environment (Singh and Harris, 2005). This finding led us to evaluate the combination of cetuximab and various chemotherapeutics in SUM102 cells, where we observed that the combination of cetuximab and carboplatin was highly synergistic at low doses of each drug. Even though the short-term co-treatment of cetuximab and carboplatin was antagonistic, the increased synergism observed in the long-term co-treatment justifies the choice of cetuximab plus carboplatin in the "triple-negative" patient trial (NCT00232505).

Carboplatin, as well as other platinum derivatives, may be good chemotherapeutic agents for basal-like breast cancers due to the implicated important function of the BRCA1-pathway in this subtype. Namely, *BRCA1* mutation carriers are predisposed to develop tumors of the basal-like subtype (Foulkes et al., 2003; Sørlie et al., 2003; Arnes et al., 2005). Indeed, in our basal-like cell lines, it has been reported that the SUM149 line has a *BRCA1*

mutation and SUM102 line has barely detectable transcript levels of *BRCA1* (Elstrodt et al., 2006). From a mechanistic standpoint, *BRCA1* is required for repair of cisplatin induced DNA damage by recruiting *RAD51* to the site of damage (Bhattacharyya et al., 2000; Zhou et al., 2005) and *BRCA1*-deficient cells exhibit increased sensitivity to cisplatin compared to wild type cells (Husain et al., 1998; Quinn et al., 2003; Tassone et al., 2003; Kennedy et al., 2004). The combination of an EGFR inhibitor and a platinum drug has also been found to be synergistic in several other cell types (Ciardiello et al., 2000; Hambek et al., 2005; Morelli et al., 2005). In our experiments, we showed that not only are the basal-like tumor derived cell lines the most sensitive to platinum drugs and the EGFR inhibitors when applied individually, but also that the combination was synergistic.

Given the biological importance of the EGFR pathway, we wished to identify an EGFR-activation profile and examine its interplay with other biological features. We first evaluated EGFR signaling using the SUM102 cell line, which was the only cell line tested that was sensitive to both cetuximab and gefitinib and has previously been shown to be EGFR-dependent (Sartor et al., 1997). The SUM102 analysis identified a very large signature that contained known "immediate early" target genes (Figure 4.2C) and a stereotypical proliferation response (Whitfield et al., 2002; Whitfield et al., 2006). To further investigate the biological relevance of EGFR-activation response *in vivo*, we took the top 500 induced genes from the SUM102 post inhibitor time point experiments and used them to analyze a large set of primary breast tumors. As was seen when performing a similar analysis that used *in vitro* defined estrogen-regulated genes (Oh et al., 2006), the homogenous expression pattern obtained from the cell line experiment corresponded to a heterogeneous patterns of expression in the primary tumors, which split into three distinct expression patterns (Figure

4.3), of which two predicted patient outcomes in both the training and test data sets (i.e. Cluster #2 and #3) and each cluster high in different subtype/subsets of tumors. While clusters #2 and #3 may be of prognostic value, their greatest value may be as an assay to identify tumors/patients that may benefit from therapeutic intervention of the EGFR-RAS-MEK pathway. These signatures represent a more dynamic descriptor of pathway activity compared to EGFR protein status alone, which is already known to not predict responsiveness to EGFR inhibitors (Fountzilas et al., 2005; Gasparini et al., 2005; von Minckwitz et al., 2005). Microarray studies of breast cancer patients treated with EGFR inhibitors will be needed to address this hypothesis.

One of the most critical questions is what are the molecular events that cause activation of the EGFR-RAS-MEK pathway, and do these activation events vary with subtype? To address these questions, we queried our data and found that known critical signaling components of this pathway correlated with tumor subtype or with the EGFR *in vivo* defined profiles. These data are summarized in Figure 4.5 where the average expression value for each gene, within each of the four major subtypes, is shown relative to the average expression value of that gene across all samples within the context of a diagram of the HER1-RAS-MEK pathway. Many relationships were identified including previously known associations as well as new associations.

The luminal A versus luminal B distinction showed numerous EGFR-pathway specific differences that included the low expression of all three EGFR-activation signatures within the luminal A subtype and the high expression of the growth inhibitory *HER4* protein, with the average expression of two of its ligands (*HB-EGF* and *NRG1*, Figure 4.5A); this subtype also shows the low expression of *EGFR* and *CRYAB*. Conversely, the luminal B

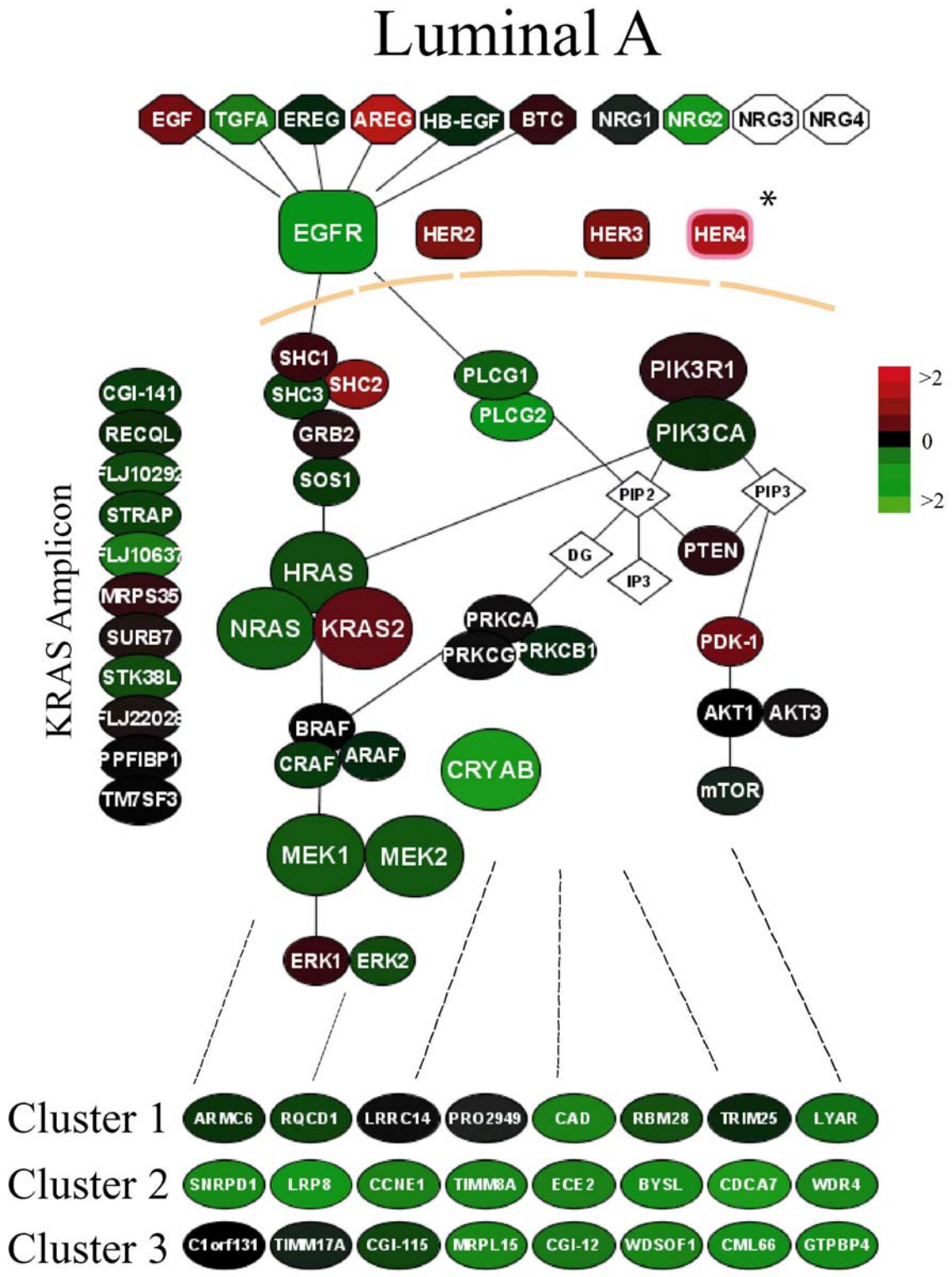
Luminal A
EGF
TGFA
EREG
AREG
HB-EGF
BTC
NRG1
NRG2
NRG3
NRG4
EGFR
HER2
HER3
HER4
*
SHC1
SHC3
SHC2
GRB2
SOS1
PLCG1
PLCG2
PIK3R1
PIK3CA
PIP2
PIP3
DG
IP3
PTEN
HRAS
NRAS
KRAS2
PRKCA
PRKCG
PRKCB1
PDK-1
AKT1
AKT3
mTOR
BRAF
CRAF
ARAF
CRYAB
MEK1
MEK2
ERK1
ERK2
KRAS Amplicon
CGI-141
RECQL
FLJ10292
STRAP
FLJ10637
MRPS35
SURB7
STK38L
FLJ22028
PPFIBP1
TM7SF3
>2
0
>2
Cluster 1
ARMC6
RQCD1
LRRC14
PRO2949
CAD
RBM28
TRIM25
LYAR
Cluster 2
SNRPD1
LRP8
CCNE1
TIMM8A
ECE2
BYSL
CDCA7
WDR4
Cluster 3
C1orf131
TIMM17A
CGI-115
MRPL15
CGI-12
WDSOF1
CML66
GTPBP4

Figure 4.5A

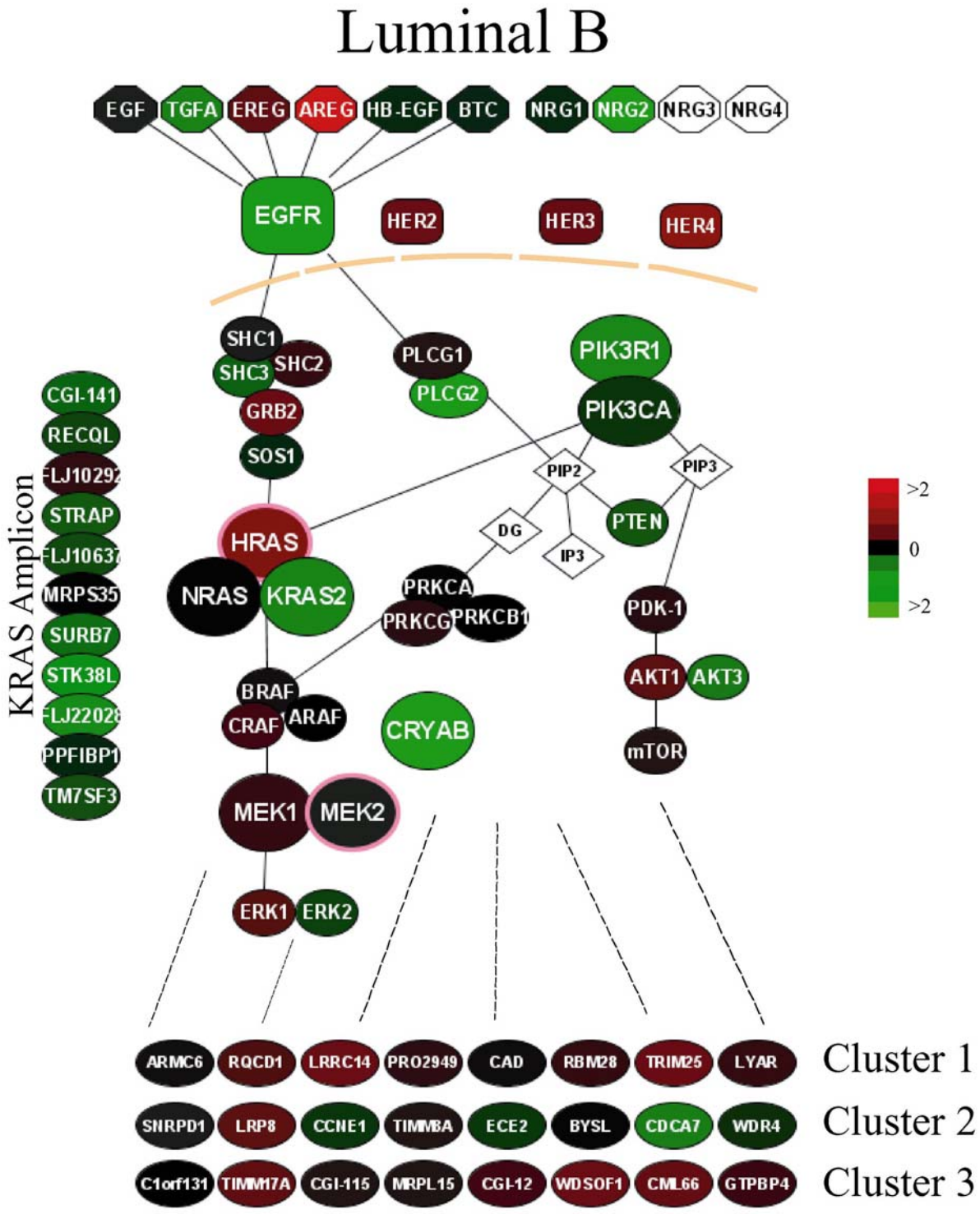
Luminal B
EGF
TGFA
EREG
AREG
HB-EGF
BTC
NRG1
NRG2
NRG3
NRG4
EGFR
HER2
HER3
HER4
SHC1
SHC3
SHC2
GRB2
SOS1
PLCG1
PLCG2
PIK3R1
PIK3CA
PIP2
PIP3
DG
IP3
PTEN
HRAS
NRAS
KRAS2
PRKCA
PRKCG
PRKCB1
PDK-1
AKT1
AKT3
mTOR
BRAF
CRAF
ARAF
CRYAB
MEK1
MEK2
ERK1
ERK2
KRAS Amplicon
CGI-141
RECQL
FLJ10292
STRAP
FLJ10637
MRPS35
SURB7
STK38L
FLJ22028
PPFIBP1
TM7SF3
>2
0
>2
ARMC6
RQCD1
LRRC14
PRO2949
CAD
RBM28
TRIM25
LYAR
Cluster 1
SNRPD1
LRP8
CCNE1
TIMM8A
ECE2
BYSL
CDCA7
WDR4
Cluster 2
C1orf131
TIMM17A
CGI-115
MRPL15
CGI-12
WDSOF1
CML66
GTPBP4
Cluster 3

Figure 4.5B

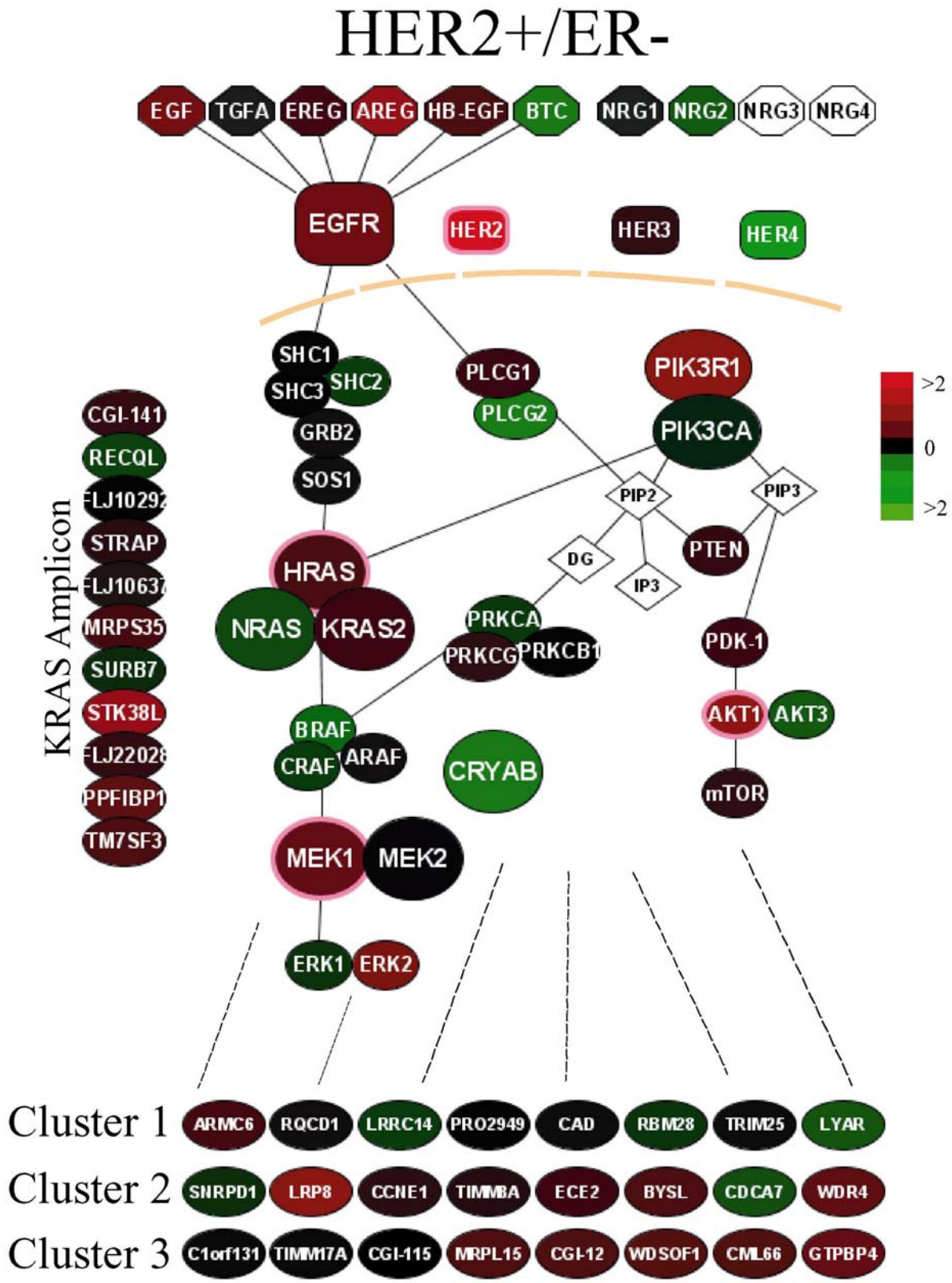
HER2+/ER-
EGF
TGFA
EREG
AREG
HB-EGF
BTC
NRG1
NRG2
NRG3
NRG4
EGFR
HER2
HER3
HER4
SHC1
SHC3
SHC2
GRB2
SOS1
PLCG1
PLCG2
PIK3R1
PIK3CA
PIP2
PIP3
DG
IP3
PTEN
HRAS
NRAS
KRAS2
PRKCA
PRKCG
PRKCB1
PDK-1
AKT1
AKT3
mTOR
BRAF
CRAF
ARAF
CRYAB
MEK1
MEK2
ERK1
ERK2
KRAS Amplicon
CGI-141
RECQL
FLJ10292
STRAP
FLJ10637
MRPS35
SURB7
STK38L
FLJ22028
PPFIBP1
TM7SF3
>2
0
>2
Cluster 1
ARMC6
RQCD1
LRRC14
PRO2949
CAD
RBM28
TRIM25
LYAR
Cluster 2
SNRPD1
LRP8
CCNE1
TIMMBA
ECE2
BYSL
CDCA7
WDR4
Cluster 3
C1orf131
TIMM17A
CGI-115
MRPL15
CGI-12
WDSOF1
CML66
GTPBP4

Figure 4.5C

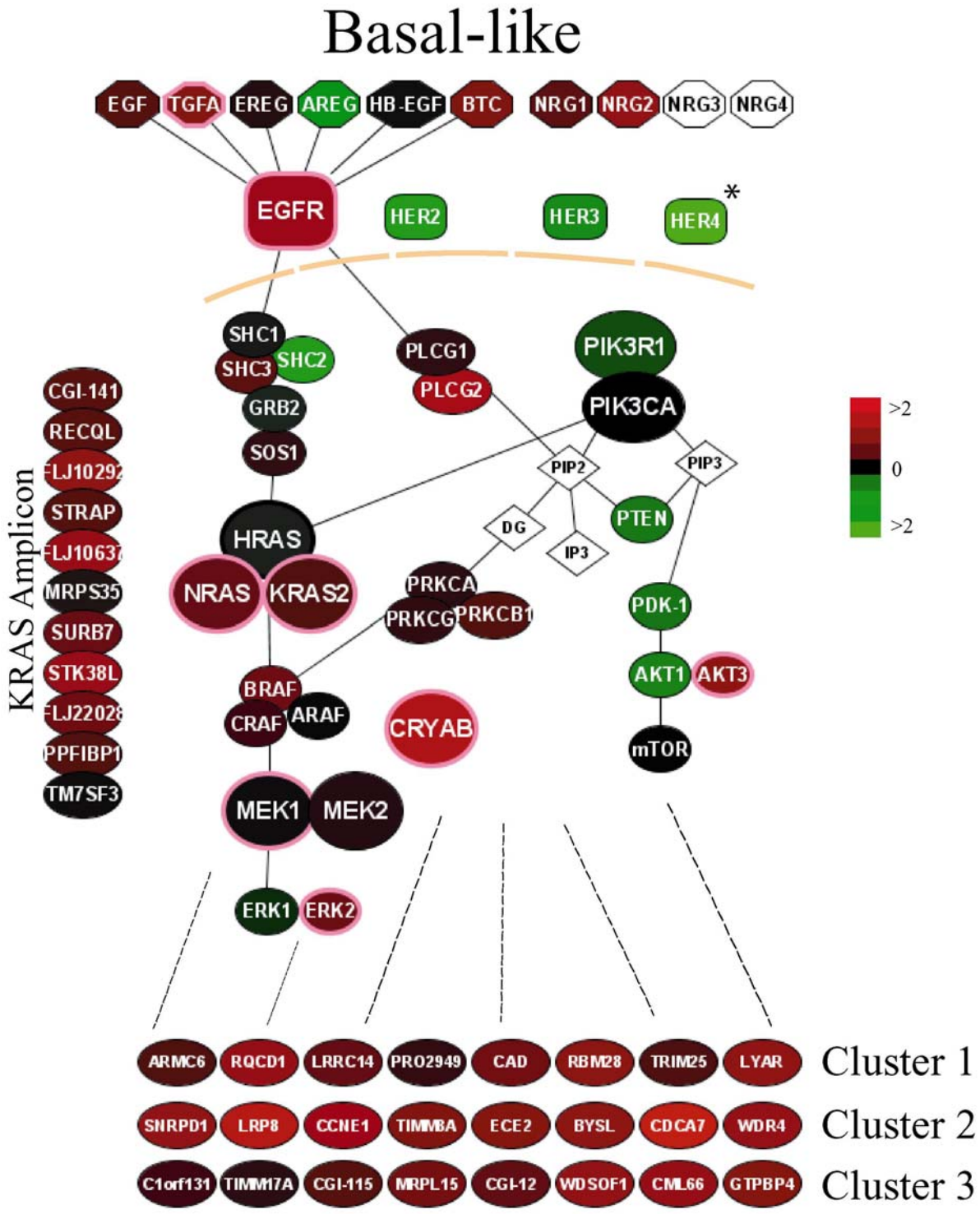
Basal-like
EGF
TGFA
EREG
AREG
HB-EGF
BTC
NRG1
NRG2
NRG3
NRG4
EGFR
HER2
HER3
HER4
*
SHC1
SHC3
SHC2
GRB2
SOS1
PLCG1
PLCG2
PIK3R1
PIK3CA
PIP2
PIP3
DG
IP3
PTEN
HRAS
NRAS
KRAS2
PRKCA
PRKCG
PRKCB1
PDK-1
AKT1
AKT3
mTOR
BRAF
CRAF
ARAF
CRYAB
MEK1
MEK2
ERK1
ERK2
KRAS Amplicon
CGI-141
RECQL
FLJ10292
STRAP
FLJ10637
MRPS35
SURB7
STK38L
FLJ22028
PPFIBP1
TM7SF3
>2
0
>2
ARMC6
RQCD1
LRRC14
PRO2949
CAD
RBM28
TRIM25
LYAR
Cluster 1
SNRPD1
LRP8
CCNE1
TIMMBA
ECE2
BYSL
CDCA7
WDR4
Cluster 2
C1orf131
TIMM17A
CGI-115
MRPL15
CGI-12
WDSOF1
CML66
GTPBP4
Cluster 3

Figure 4.5D

Figure 4.5. EGFR pathway diagram displayed for each breast tumor subtype. The average gene expression value for each gene within each subtype is displayed for the EGFR-pathway and for the three EGFR-activation profiles using the UNC tumor dataset. Eight genes from the middle of each of the three EGFR-activation clusters were used to view expression of the clusters in each of the subtypes. A pink node border identifies the genes that showed statistically significant associations with subtype. *Note: the NKI HER4 data spot was used since HER4 was not present in the UNC data set. *A*) Luminal A, *B*) Luminal B, *C*) HER2+/ER- and *D*) Basal-like.

subtype is characterized by the moderate to high expression of the EGFR-activation signatures, high *HRAS* expression and potentially high *MEK2* (Figure 5.4B). The EGFR/HER2 pathway has been often implicated as at least partially responsible for tamoxifen resistance in ER+ patients (Arpino et al., 2004; Gutierrez et al., 2005; Normanno et al., 2005; Britton et al., 2006; Dowsett et al., 2006; Ellis et al., 2006), and Clusters #2 and #3 were able to predict outcome differences in ER+ and tamoxifen-treated patients in both the UNC and NKI data sets (data not shown); however, the expression of Clusters #2 and #3 in ER+ patients closely parallels the genomic distinction of luminal A versus luminal B. Taken together, these results suggest that part of the luminal A versus luminal B distinction is due to the activation of the EGFR/HER2 pathway in luminal B tumors. Dowsett et al. (Dowsett et al., 2006) tested gefitinib in combination with anastrozole, an aromatase inhibitor, in ER+/PgR+ tumors and found that the combination gave no additional benefit over anastrozole alone in ability to suppress Ki67 expression. While our data are consistent with the non-genomic ER signaling, they suggest other downstream activations may be occurring such as HRAS, which may explain the lack of effect of gefitinib in this trial.

The HER2+/ER- tumors, as expected by definition, showed high expression of *HER2* and were also associated with high *HRAS* and *MEK1/MEK2* (Figure 4.5C). High *AKT1* levels were also associated with this tumor subtype, which is an association that has been identified before (Zhou et al., 2004; Tokunaga et al., 2006). Perhaps the most informative associations were from the basal-like subtype where numerous correlations were identified that included an association of basal-like tumors with high *EGFR, TGFA, MEK1, MEK2, AKT3, CRYAB, NRAS* and the *KRAS*-amplicon signature (Figure 4.5D). The basal-like tumors also had the highest expression of the three EGFR activation clusters suggesting a ubiquitous dependence

on this pathway; however, the mechanism of activation may vary across the basal-like tumors. As an example of an *in vitro* defined EGFR-RAS-MEK ligand independent activation mechanism, Moyano *et al*. showed that the ectopic expression of CRYAB in breast epithelial cells caused them to become transformed and EGF-independent (Moyano et al., 2006). This transformed phenotype resulted in the activation of the MEK-ERK pathway and was reverted by the addition of the MEK inhibitors PD98059 and U0126, while the PIK3CA inhibitor LY294002 had little effect. CRYAB may also confer resistance to EGFR inhibitors as well as chemotherapy by its anti-apoptotic mechanism, the inhibition of caspase-3 activation (Kamradt et al., 2001; Kamradt et al., 2005). A second example of a ligand independent activation event in this pathway is the high expression of a mutant RAS protein, where activating mutations have been shown to correlate with activated MEK/ERK (Martinez-Lacaci et al., 2000; Han et al., 2006). Interestingly, the expression of the *KRAS*-amplicon gene expression signature showed many more associations versus the *KRAS* gene alone, suggesting that either the amplicon signature is a better assay for detecting *KRAS* activity than the simple expression of *KRAS* alone, or more likely, that additional co-amplified and highly expressed genes synergize with *KRAS* to activate the RAS-MEK pathway. Only a small subset of basal-like tumors show the high expression of *EGFR* and one of its ligands (typically *TGFA*) and the low expression of *CRYAB* and *KRAS*; it is only this subset of basal-like tumors that might be responsive to EGFR inhibitors of any kind (i.e. cetuximab, erlotinib, or gefitinib) because it is only these tumors that activate this pathway using a ligand dependent activation of the EGFR-pathway. Examples of individual basal-like tumors that show each of these activation profiles as well as a tumor with high activation of all three are presented in Figure 4.6.

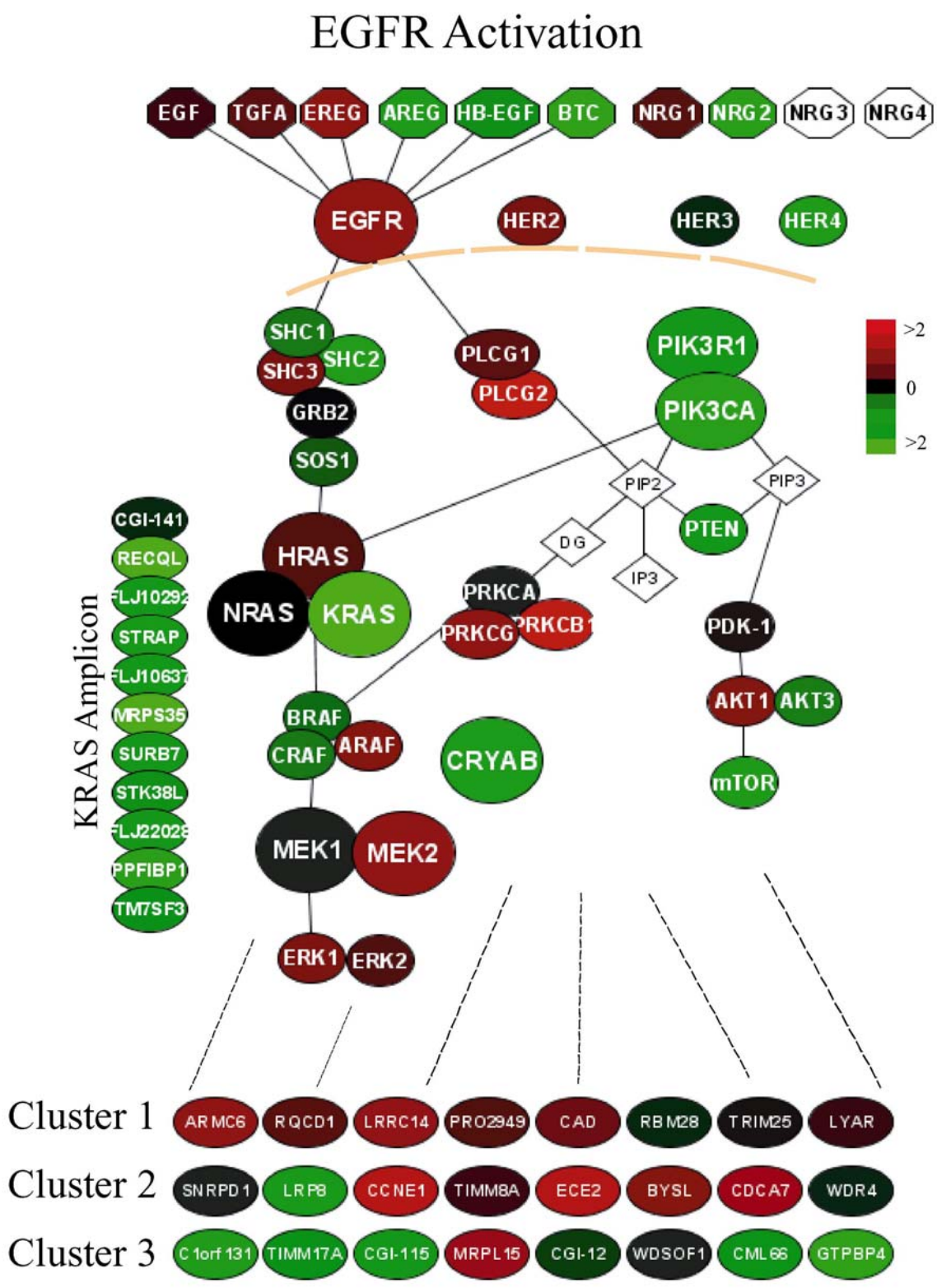
EGFR Activation
EGF
TGFA
EREG
AREG
HB-EGF
BTC
NRG1
NRG2
NRG3
NRG4
EGFR
HER2
HER3
HER4
SHC1
SHC2
SHC3
GRB2
SOS1
PLCG1
PLCG2
PIK3R1
PIK3CA
>2
0
>2
PIP2
PIP3
PTEN
DG
IP3
HRAS
NRAS
KRAS
PRKCA
PRKCG
PRKCB1
PDK-1
AKT1
AKT3
mTOR
BRAF
CRAF
ARAF
CRYAB
MEK1
MEK2
ERK1
ERK2
KRAS Amplicon
CGI-141
RECQL
FLJ10293
STRAP
FLJ10637
MRPS35
SURB7
STK38L
FLJ22028
PPFIBP1
TM7SF3
Cluster 1
ARMC6
RQCD1
LRRC14
PRO2949
CAD
RBM28
TRIM25
LYAR
Cluster 2
SNRPD1
LRP8
CCNE1
TIMM8A
ECE2
BYSL
CDCA7
WDR4
Cluster 3
C1orf131
TIMM17A
CGI-115
MRPL15
CGI-12
WDSOF1
CML66
GTPBP4

Figure 4.6A

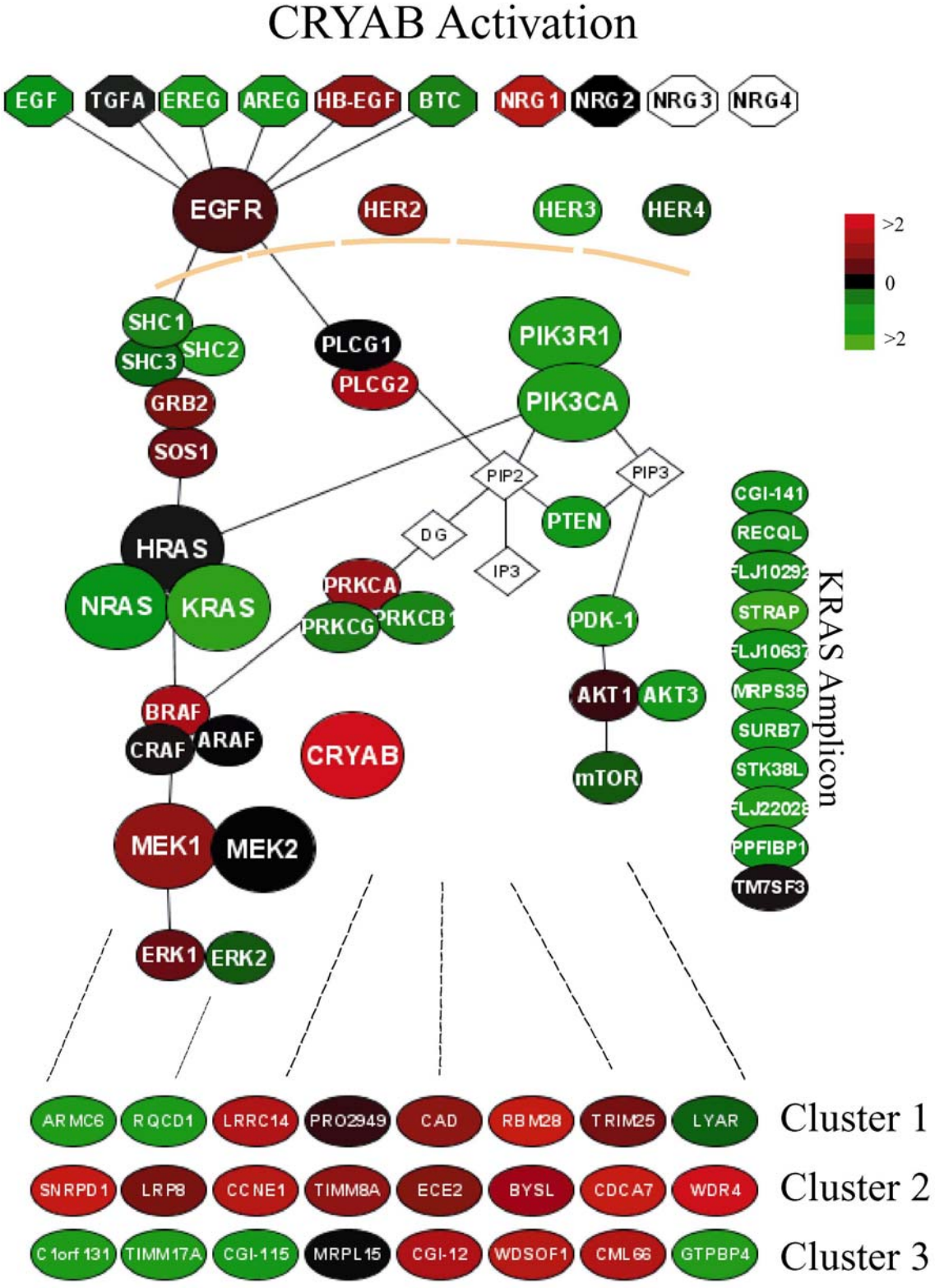
CRYAB Activation
EGF
TGFA
EREG
AREG
HB-EGF
BTC
NRG1
NRG2
NRG3
NRG4
EGFR
HER2
HER3
HER4
>2
0
>2
SHC1
SHC3
SHC2
GRB2
SOS1
PLCG1
PLCG2
PIK3R1
PIK3CA
PIP2
PIP3
DG
IP3
PTEN
HRAS
NRAS
KRAS
PRKCA
PRKCG
PRKCB1
PDK-1
AKT1
AKT3
mTOR
BRAF
CRAF
ARAF
CRYAB
MEK1
MEK2
ERK1
ERK2
CGI-141
RECQL
FLJ10292
STRAP
FLJ10637
MRPS35
SURB7
STK38L
FLJ22028
PPFIBP1
TM7SF3
KRAS Amplicon
ARMC6
RQCD1
LRRC14
PRO2949
CAD
RBM28
TRIM25
LYAR
Cluster 1
SNRPD1
LRP8
CCNE1
TIMM8A
ECE2
BYSL
CDCA7
WDR4
Cluster 2
C1orf131
TIMM17A
CGI-115
MRPL15
CGI-12
WDSOF1
CML66
GTPBP4
Cluster 3

Figure 4.6B

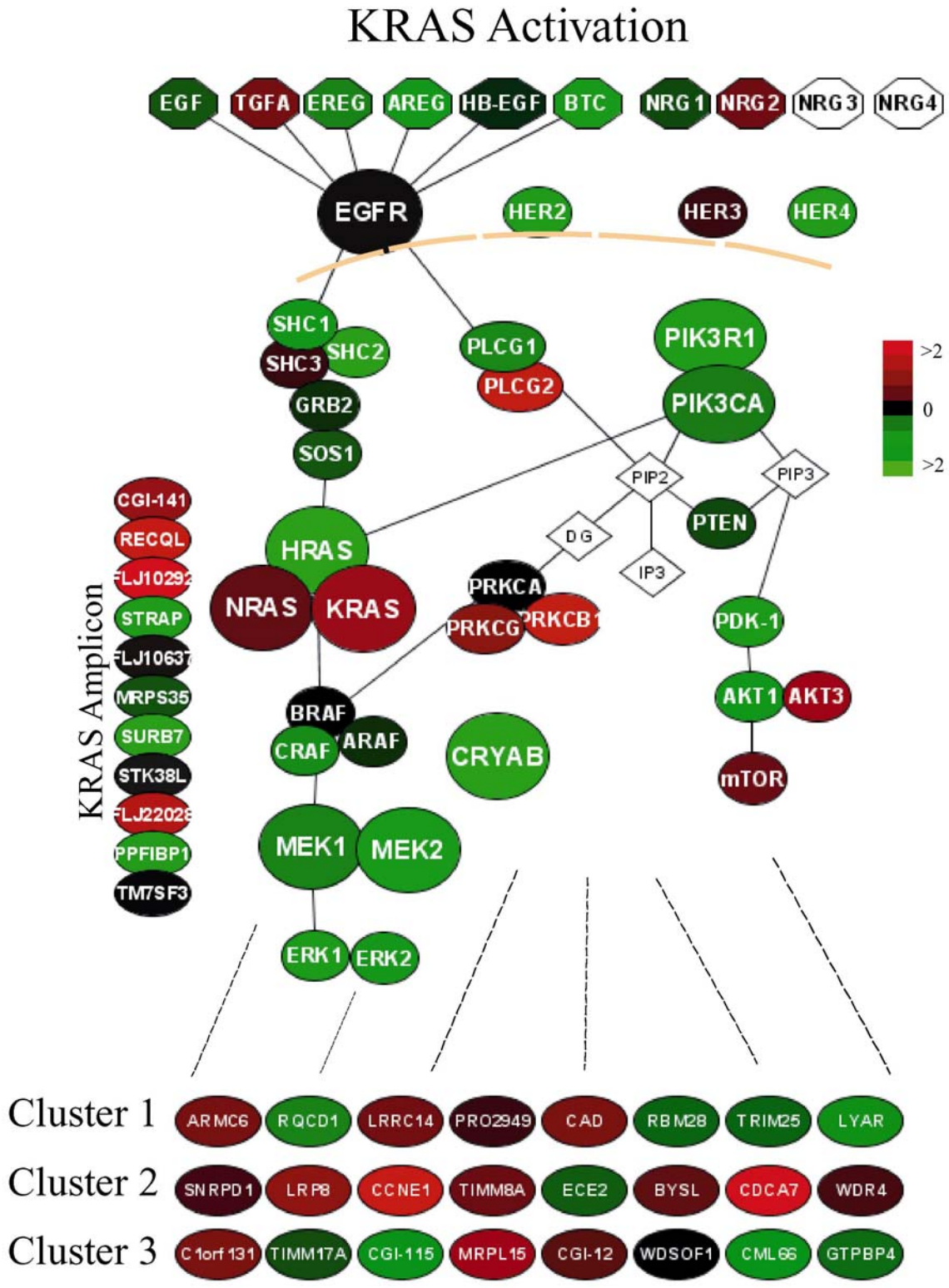
KRAS Activation
EGF
TGFA
EREG
AREG
HB-EGF
BTC
NRG1
NRG2
NRG3
NRG4
EGFR
HER2
HER3
HER4
SHC1
SHC2
SHC3
GRB2
SOS1
PLCG1
PLCG2
PIK3R1
PIK3CA
>2
0
>2
PIP2
PIP3
PTEN
DG
IP3
HRAS
NRAS
KRAS
PRKCA
PRKCG
PRKCB1
PDK-1
AKT1
AKT3
mTOR
BRAF
CRAF
ARAF
CRYAB
MEK1
MEK2
ERK1
ERK2
KRAS Amplicon
CGI-141
RECQL
FLJ10292
STRAP
FLJ10637
MRPS35
SURB7
STK38L
FLJ22028
PPFIBP1
TM7SF3
Cluster 1
ARMC6
RQCD1
LRRC14
PRO2949
CAD
RBM28
TRIM25
LYAR
Cluster 2
SNRPD1
LRP8
CCNE1
TIMM8A
ECE2
BYSL
CDCA7
WDR4
Cluster 3
C1orf131
TIMM17A
CGI-115
MRPL15
CGI-12
WDSOF1
CML66
GTPBP4

Figure 4.6C

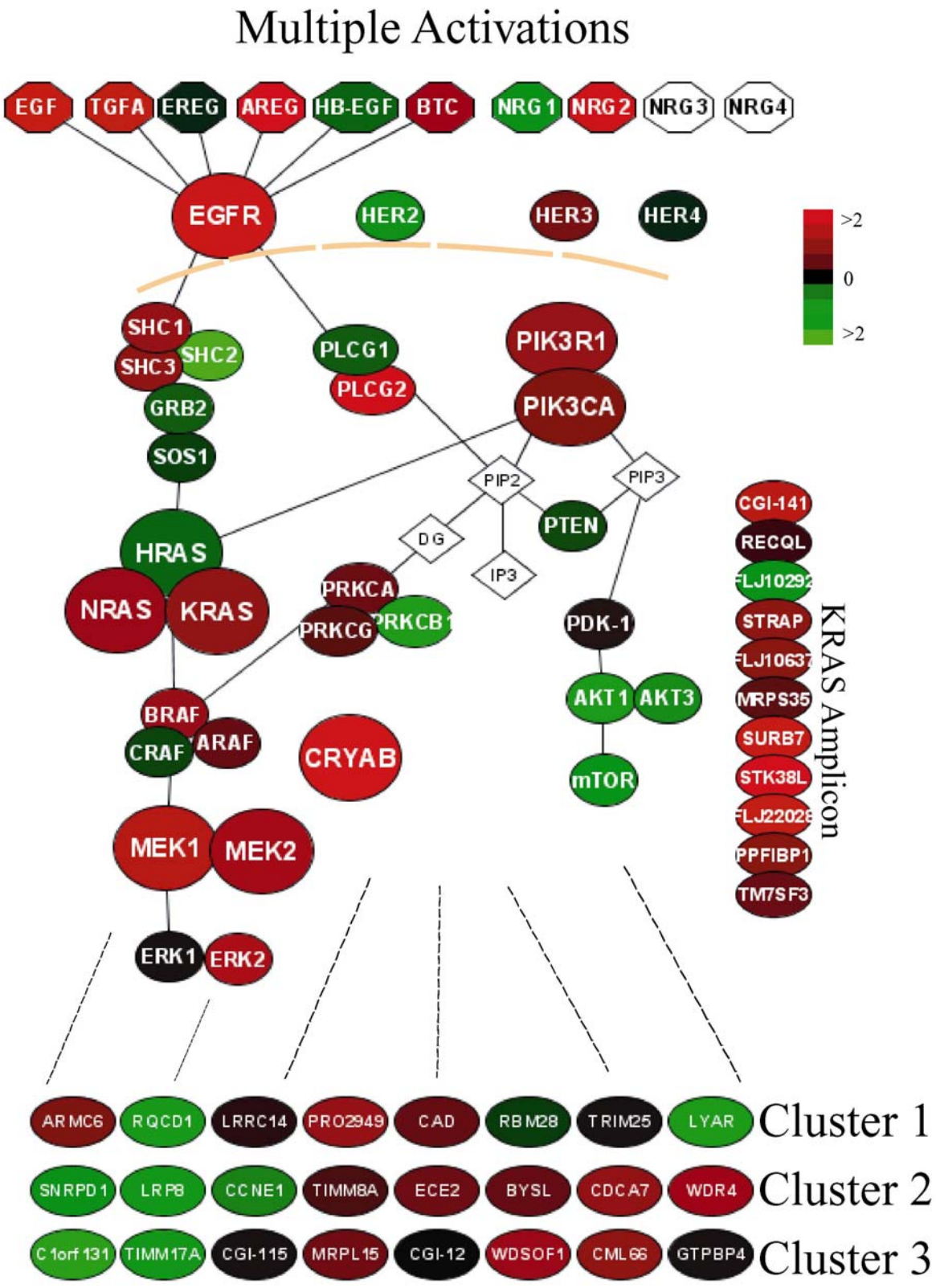
Multiple Activations
EGF
TGFA
EREG
AREG
HB-EGF
BTC
NRG1
NRG2
NRG3
NRG4
EGFR
HER2
HER3
HER4
>2
0
>2
SHC1
SHC3
SHC2
GRB2
SOS1
PLCG1
PLCG2
PIK3R1
PIK3CA
PIP2
PIP3
DG
IP3
PTEN
HRAS
NRAS
KRAS
PRKCA
PRKCG
PRKCB1
PDK-1
AKT1
AKT3
mTOR
BRAF
CRAF
ARAF
CRYAB
MEK1
MEK2
ERK1
ERK2
CGI-141
RECQL
FLJ10292
STRAP
FLJ10637
MRPS35
SURB7
STK38L
FLJ22028
PPFIBP1
TM7SF3
KRAS Amplicon
ARMC6
RQCD1
LRRC14
PRO2949
CAD
RBM28
TRIM25
LYAR
Cluster 1
SNRPD1
LRP8
CCNE1
TIMM8A
ECE2
BYSL
CDCA7
WDR4
Cluster 2
C1orf131
TIMM17A
CGI-115
MRPL15
CGI-12
WDSOF1
CML66
GTPBP4
Cluster 3

Figure 4.6D

Figure 4.6. EGFR pathway diagram displayed for each type of mechanism that could cause activation of the EGFR-RAS-MEK pathway in basal-like tumors. *A*) *EGFR*-ligand dependent activation profile, *B*) *CRYAB* activation profile, *C*) *KRAS*-amplicon activation profile, *D*) multiple simultaneous activation profiles.

These results paint a different portrait of the EGFR-pathway for each subtype that has important therapeutic implications. For the luminal A subtype, these data suggest that EGFR-RAS-MEK signaling is not a relevant therapeutic target and that if anything, *HER4* may be the critical HER family member for this subtype.

Conversely, for the Luminal B tumors, these data recapitulate the finding that the EGFR-RAS-MEK pathway appears to be an important target that may be activated by means other than EGFR ligand dependent activation. Our results in Luminal B tumors are also consistent with the hypothesis of the "non-genomic" effects of ER-pathway activation that occurs via HER2 predominantly, where membrane bound ER complexes with HER2 to cause activation of the RAS-MEK and p38 pathways (Arpino et al., 2004; Shou et al., 2004; Gutierrez et al., 2005). Ways to effectively target this pathway in luminal B tumors are not clear, but our data suggests that agents that target HER2, RAS or MEK might be worth investigating.

Targeting of the HER pathway in the HER2+/ER- tumors involves the administration of trastuzumab; however, given that the response rate to trastuzumab-containing therapies is approximately 50%, additional agents are needed. Candidates from this and other studies could include the direct or indirect targeting of *AKT1* and/or *MEK*1.

For basal-like patients, these data suggest that EGFR-RAS-MEK pathway activation is a requisite as almost every single basal-like tumor analyzed showed the high expression of one of the EGFR-activation signatures, and in fact, most showed the high expression of all three. The effective targeting of the EGFR-RAS-MEK pathway in basal-like tumors may require additional patient stratification based upon the mechanism of activation of the pathway. For example, those basal-like patients who show high *EGFR* and ligand and low

CRYAB and *KRAS*-amplicon could be candidates for EGFR inhibitors, while those who show high *CRYAB* and/or *KRAS*-amplicon would be candidates for MEK inhibitors, of which many are in clinical development. Alternatively, if MEK inhibitors show good efficacy and low side effects, all patients who show activation of the EGFR-RAS-MEK pathway as assessed by Clusters #2 and #3 (both of which were associated with high *MEK* expression-Table 4.4), then these may prove to be good drugs to administer to all basal-like patients, as well as some HER2+/ER- and some Luminal B patients.

ACKNOWLEDGEMENTS

We thank AstraZeneca for the gift of gefitinib. This work was supported by funds for CMP from the NCI Breast SPORE program to UNC-CH (P50-CA58223-09A1), by RO1-CA-101227-01, by M01RR00046, by the Breast Cancer Research Foundation, and National institute of Environmental Health Sciences Grants U19-ES11391. KH was supported by a Department of Defense Predoctoral Fellowship BC043145.

REFERENCES

Adams, G. P. and Weiner, L. M. (2005). Monoclonal antibody therapy of cancer. Nat Biotechnol 23(9): 1147-1157.

Agrawal, A., Gutteridge, E., Gee, J. M., Nicholson, R. I. and Robertson, J. F. (2005). Overview of tyrosine kinase inhibitors in clinical breast cancer. Endocr Relat Cancer 12 Suppl 1: S135-144.

Akita, R. W. and Sliwkowski, M. X. (2003). Preclinical studies with Erlotinib (Tarceva). Semin Oncol 30(3 Suppl 7): 15-24.

Anido, J., Matar, P., Albanell, J., Guzman, M., Rojo, F., Arribas, J., Averbuch, S. and Baselga, J. (2003). ZD1839, a specific epidermal growth factor receptor (EGFR) tyrosine kinase inhibitor, induces the formation of inactive EGFR/HER2 and EGFR/HER3 heterodimers and prevents heregulin signaling in HER2-overexpressing breast cancer cells. Clin Cancer Res 9(4): 1274-1283.

Arnes, J. B., Brunet, J. S., Stefansson, I., Begin, L. R., Wong, N., Chappuis, P. O., Akslen, L. A. and Foulkes, W. D. (2005). Placental cadherin and the basal epithelial phenotype of BRCA1-related breast cancer. Clin Cancer Res 11(11): 4003-4011.

Arpino, G., Green, S. J., Allred, D. C., Lew, D., Martino, S., Osborne, C. K. and Elledge, R. M. (2004). HER-2 amplification, HER-1 expression, and tamoxifen response in estrogen receptor-positive metastatic breast cancer: a southwest oncology group study. Clin Cancer Res 10(17): 5670-5676.

Baselga, J. (2002). Why the epidermal growth factor receptor? The rationale for cancer therapy. Oncologist 7 Suppl 4: 2-8.

Bertucci, F., Finetti, P., Rougemont, J., Charafe-Jauffret, E., Cervera, N., Tarpin, C., Nguyen, C., Xerri, L., Houlgatte, R., Jacquemier, J., Viens, P. and Birnbaum, D. (2005). Gene expression profiling identifies molecular subtypes of inflammatory breast cancer. Cancer Res 65(6): 2170-2178.

Bhattacharyya, A., Ear, U. S., Koller, B. H., Weichselbaum, R. R. and Bishop, D. K. (2000). The breast cancer susceptibility gene BRCA1 is required for subnuclear assembly of Rad51 and survival following treatment with the DNA cross-linking agent cisplatin. J Biol Chem 275(31): 23899-23903.

Bianco, R., Shin, I., Ritter, C. A., Yakes, F. M., Basso, A., Rosen, N., Tsurutani, J., Dennis, P. A., Mills, G. B. and Arteaga, C. L. (2003). Loss of PTEN/MMAC1/TEP in EGF receptor-expressing tumor cells counteracts the antitumor action of EGFR tyrosine kinase inhibitors. Oncogene 22(18): 2812-2822.

Britton, D. J., Hutcheson, I. R., Knowlden, J. M., Barrow, D., Giles, M., McClelland, R. A., Gee, J. M. and Nicholson, R. I. (2006). Bidirectional cross talk between ERalpha and EGFR signalling pathways regulates tamoxifen-resistant growth. Breast Cancer Res Treat 96(2): 131-146.

Chang, H. Y., Nuyten, D. S., Sneddon, J. B., Hastie, T., Tibshirani, R., Sorlie, T., Dai, H., He, Y. D., van't Veer, L. J., Bartelink, H., van de Rijn, M., Brown, P. O. and van de Vijver, M. J. (2005). Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. Proc Natl Acad Sci U S A 102(10): 3738-3743.

Chou, T. C. and Talalay, P. (1984). Quantitative analysis of dose-effect relationships: the combined effects of multiple drugs or enzyme inhibitors. Adv Enzyme Regul 22: 27-55.

Ciardiello, F., Caputo, R., Bianco, R., Damiano, V., Pomatico, G., De Placido, S., Bianco, A. R. and Tortora, G. (2000). Antitumor effect and potentiation of cytotoxic drugs activity in human cancer cells by ZD-1839 (Iressa), an epidermal growth factor receptor-selective tyrosine kinase inhibitor. Clin Cancer Res 6(5): 2053-2063.

Culy, C. R. and Faulds, D. (2002). Gefitinib. Drugs 62(15): 2237-2248; discussion 2249-2250.

Dowell, J., Minna, J. D. and Kirkpatrick, P. (2005). Erlotinib hydrochloride. Nat Rev Drug Discov 4(1): 13-14.

Dowsett, M., Houghton, J., Iden, C., Salter, J., Farndon, J., A'Hern, R., Sainsbury, R. and Baum, M. (2006). Benefit from adjuvant tamoxifen therapy in primary breast cancer patients according oestrogen receptor, progesterone receptor, EGF receptor and HER2 status. Ann Oncol 17(5): 818-826.

Eisen, M. B. and Brown, P. O. (1999). DNA arrays for analysis of gene expression. Methods Enzymol 303: 179-205.

Eisen, M. B., Spellman, P. T., Brown, P. O. and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 95(25): 14863-14868.

Ellis, M. J., Tao, Y., Young, O., White, S., Proia, A. D., Murray, J., Renshaw, L., Faratian, D., Thomas, J., Dowsett, M., Krause, A., Evans, D. B., Miller, W. R. and Dixon, J. M. (2006). Estrogen-independent proliferation is present in estrogen-receptor HER2-positive primary breast cancer after neoadjuvant letrozole. J Clin Oncol 24(19): 3019-3025.

Elstrodt, F., Hollestelle, A., Nagel, J. H., Gorin, M., Wasielewski, M., van den Ouweland, A., Merajver, S. D., Ethier, S. P. and Schutte, M. (2006). BRCA1 mutation analysis of 41

human breast cancer cell lines reveals three new deleterious mutants. Cancer Res 66(1): 41-45.

Fan, C., Oh, D. S., Wessels, L., Weigelt, B., Nuyten, D. S., Nobel, A. B., van't Veer, L. J. and Perou, C. M. (2006). Concordance among gene-expression-based predictors for breast cancer. N Engl J Med 355(6): 560-569.

Foulkes, W. D., Brunet, J. S., Stefansson, I. M., Straume, O., Chappuis, P. O., Begin, L. R., Hamel, N., Goffin, J. R., Wong, N., Trudel, M., Kapusta, L., Porter, P. and Akslen, L. A. (2004). The prognostic implication of the basal-like (cyclin E high/p27 low/p53+/glomeruloid-microvascular-proliferation+) phenotype of BRCA1-related breast cancer. Cancer Res 64(3): 830-835.

Foulkes, W. D., Stefansson, I. M., Chappuis, P. O., Begin, L. R., Goffin, J. R., Wong, N., Trudel, M. and Akslen, L. A. (2003). Germline BRCA1 mutations and a basal epithelial phenotype in breast cancer. J Natl Cancer Inst 95(19): 1482-1485.

Fountzilas, G., Pectasides, D., Kalogera-Fountzila, A., Skarlos, D., Kalofonos, H. P., Papadimitriou, C., Bafaloukos, D., Lambropoulos, S., Papadopoulos, S., Kourea, H., Markopoulos, C., Linardou, H., Mavroudis, D., Briasoulis, E., Pavlidis, N., Razis, E., Kosmidis, P. and Gogas, H. (2005). Paclitaxel and carboplatin as first-line chemotherapy combined with gefitinib (IRESSA) in patients with advanced breast cancer: a phase I/II study conducted by the Hellenic Cooperative Oncology Group. Breast Cancer Res Treat 92(1): 1-9.

Gasparini, G., Sarmiento, R., Amici, S., Longo, R., Gattuso, D., Zancan, M. and Gion, M. (2005). Gefitinib (ZD1839) combined with weekly epirubicin in patients with metastatic breast cancer: a phase I study with biological correlate. Ann Oncol 16(12): 1867-1873.

Graham, J., Muhsin, M. and Kirkpatrick, P. (2004). Cetuximab. Nat Rev Drug Discov 3(7): 549-550.

Gutierrez, M. C., Detre, S., Johnston, S., Mohsin, S. K., Shou, J., Allred, D. C., Schiff, R., Osborne, C. K. and Dowsett, M. (2005). Molecular changes in tamoxifen-resistant breast cancer: relationship between estrogen receptor, HER-2, and p38 mitogen-activated protein kinase. J Clin Oncol 23(11): 2469-2476.

Hambek, M., Baghi, M., Strebhardt, K., Baumann, H., Gstottner, W. and Knecht, R. (2005). Reduction of cisplatin dosage by ZD 1839. Anticancer Res 25(6B): 3985-3988.

Han, S. W., Kim, T. Y., Jeon, Y. K., Hwang, P. G., Im, S. A., Lee, K. H., Kim, J. H., Kim, D. W., Heo, D. S., Kim, N. K., Chung, D. H. and Bang, Y. J. (2006). Optimization of patient selection for gefitinib in non-small cell lung cancer by combined analysis of epidermal growth factor receptor mutation, K-ras mutation, and Akt phosphorylation. Clin Cancer Res 12(8): 2538-2544.

Herschkowitz, J. I., Simin, K., Weigman, V. J., Mikaelian, I., Hu, Z., Rasmussen, K. E., Chandrasekharan, S., Backlund, M. G., Yin, Y., Glazer, R. I., Brown, P. H., Green, J. E., Kopelovich, K., Churchill, G. A., Van Dyke, T. and Perou, C. M. Identification of conserved gene expression features across human and murine mammary tumors. Submitted.

Hosack, D. A., Dennis, G., Jr., Sherman, B. T., Lane, H. C. and Lempicki, R. A. (2003). Identifying biological themes within lists of genes with EASE. Genome Biol 4(10): R70.

Hu, Z., Fan, C., Oh, D. S., Marron, J. S., He, X., Qaqish, B. F., Livasy, C., Carey, L. A., Reynolds, E., Dressler, L., Nobel, A., Parker, J., Ewend, M. G., Sawyer, L. R., Wu, J., Liu, Y., Nanda, R., Tretiakova, M., Ruiz Orrico, A., Dreher, D., Palazzo, J. P., Perreard, L., Nelson, E., Mone, M., Hansen, H., Mullins, M., Quackenbush, J. F., Ellis, M. J., Olopade, O. I., Bernard, P. S. and Perou, C. M. (2006). The molecular portraits of breast tumors are conserved across microarray platforms. BMC Genomics 7: 96.

Hu, Z., Troester, M. and Perou, C. M. (2005). High reproducibility using sodium hydroxide-stripped long oligonucleotide DNA microarrays. Biotechniques 38(1): 121-124.

Husain, A., He, G., Venkatraman, E. S. and Spriggs, D. R. (1998). BRCA1 up-regulation is associated with repair-mediated resistance to cis-diamminedichloroplatinum(II). Cancer Res 58(6): 1120-1123.

Janmaat, M. L., Rodriguez, J. A., Gallegos-Ruiz, M., Kruyt, F. A. and Giaccone, G. (2006). Enhanced cytotoxicity induced by gefitinib and specific inhibitors of the Ras or phosphatidyl inositol-3 kinase pathways in non-small cell lung cancer cells. Int J Cancer 118(1): 209-214.

Johnston, S. R. and Leary, A. (2006). Lapatinib: a novel EGFR/HER2 tyrosine kinase inhibitor for cancer. Drugs Today (Barc) 42(7): 441-453.

Kamradt, M. C., Chen, F. and Cryns, V. L. (2001). The small heat shock protein alpha B-crystallin negatively regulates cytochrome c- and caspase-8-dependent activation of caspase-3 by inhibiting its autoproteolytic maturation. J Biol Chem 276(19): 16059-16063.

Kamradt, M. C., Lu, M., Werner, M. E., Kwan, T., Chen, F., Strohecker, A., Oshita, S., Wilkinson, J. C., Yu, C., Oliver, P. G., Duckett, C. S., Buchsbaum, D. J., LoBuglio, A. F., Jordan, V. C. and Cryns, V. L. (2005). The small heat shock protein alpha B-crystallin is a novel inhibitor of TRAIL-induced apoptosis that suppresses the activation of caspase-3. J Biol Chem 280(12): 11059-11066.

Kennedy, R. D., Quinn, J. E., Mullan, P. B., Johnston, P. G. and Harkin, D. P. (2004). The role of BRCA1 in the cellular response to chemotherapy. J Natl Cancer Inst 96(22): 1659-1668.

Lev, D. C., Kim, L. S., Melnikova, V., Ruiz, M., Ananthaswamy, H. N. and Price, J. E. (2004). Dual blockade of EGFR and ERK1/2 phosphorylation potentiates growth inhibition of breast cancer cells. Br J Cancer 91(4): 795-802.

Lu, C., Speers, C., Zhang, Y., Xu, X., Hill, J., Steinbis, E., Celestino, J., Shen, Q., Kim, H., Hilsenbeck, S., Mohsin, S. K., Wakeling, A., Osborne, C. K. and Brown, P. H. (2003). Effect of epidermal growth factor receptor inhibitor on development of estrogen receptor-negative mammary tumors. J Natl Cancer Inst 95(24): 1825-1833.

Lynch, T. J., Bell, D. W., Sordella, R., Gurubhagavatula, S., Okimoto, R. A., Brannigan, B. W., Harris, P. L., Haserlat, S. M., Supko, J. G., Haluska, F. G., Louis, D. N., Christiani, D. C., Settleman, J. and Haber, D. A. (2004). Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. N Engl J Med 350(21): 2129-2139.

Martinez-Lacaci, I., Kannan, S., De Santis, M., Bianco, C., Kim, N., Wallace-Jones, B., Ebert, A. D., Wechselberger, C. and Salomon, D. S. (2000). RAS transformation causes sustained activation of epidermal growth factor receptor and elevation of mitogen-activated protein kinase in human mammary epithelial cells. Int J Cancer 88(1): 44-52.

Mass, R. D. (2004). The HER receptor family: a rich target for therapeutic development. Int J Radiat Oncol Biol Phys 58(3): 932-940.

Moasser, M. M., Basso, A., Averbuch, S. D. and Rosen, N. (2001). The tyrosine kinase inhibitor ZD1839 ("Iressa") inhibits HER2-driven signaling and suppresses the growth of HER2-overexpressing tumor cells. Cancer Res 61(19): 7184-7188.

Morelli, M. P., Cascone, T., Troiani, T., De Vita, F., Orditura, M., Laus, G., Eckhardt, S. G., Pepe, S., Tortora, G. and Ciardiello, F. (2005). Sequence-dependent antiproliferative effects of cytotoxic drugs and epidermal growth factor receptor inhibitors. Ann Oncol 16 Suppl 4: iv61-iv68.

Moyano, J. V., Evans, J. R., Chen, F., Lu, M., Werner, M. E., Yehiely, F., Diaz, L. K., Turbin, D., Karaca, G., Wiley, E., Nielsen, T. O., Perou, C. M. and Cryns, V. L. (2006). AlphaB-crystallin is a novel oncoprotein that predicts poor clinical outcome in breast cancer. J Clin Invest 116(1): 261-270.

Nielsen, T. O., Hsu, F. D., Jensen, K., Cheang, M., Karaca, G., Hu, Z., Hernandez-Boussard, T., Livasy, C., Cowan, D., Dressler, L., Akslen, L. A., Ragaz, J., Gown, A. M., Gilks, C. B., van De Rijn, M. and Perou, C. M. (2004). Immunohistochemical and clinical

characterization of the basal-like subtype of invasive breast carcinoma. Clin Cancer Res 10(16): 5367-3574.

Normanno, N., De Luca, A., Maiello, M. R., Campiglio, M., Napolitano, M., Mancino, M., Carotenuto, A., Viglietto, G. and Menard, S. (2006). The MEK/MAPK pathway is involved in the resistance of breast cancer cells to the EGFR tyrosine kinase inhibitor gefitinib. J Cell Physiol 207(2): 420-427.

Normanno, N., De Luca, A., Maiello, M. R., Mancino, M., D'Antonio, A., Macaluso, M., Caponigro, F. and Giordano, A. (2005). Epidermal growth factor receptor (EGFR) tyrosine kinase inhibitors in breast cancer: current status and future development. Front Biosci 10: 2611-2617.

Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H. and Kanehisa, M. (1999). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 27(1): 29-34.

Oh, D. S., Troester, M. A., Usary, J., Hu, Z., He, X., Fan, C., Wu, J., Carey, L. A. and Perou, C. M. (2006). Estrogen-regulated genes predict survival in hormone receptor-positive breast cancers. J Clin Oncol 24(11): 1656-1664.

Paez, J. G., Janne, P. A., Lee, J. C., Tracy, S., Greulich, H., Gabriel, S., Herman, P., Kaye, F. J., Lindeman, N., Boggon, T. J., Naoki, K., Sasaki, H., Fujii, Y., Eck, M. J., Sellers, W. R., Johnson, B. E. and Meyerson, M. (2004). EGFR Mutations in Lung Cancer: Correlation with Clinical Response to Gefitinib Therapy. Science 304(5676): 1497-1500.

Perou, C. M., Sørlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., Fluge, O., Pergamenschikov, A., Williams, C., Zhu, S. X., Lonning, P. E., Borresen-Dale, A. L., Brown, P. O. and Botstein, D. (2000). Molecular portraits of human breast tumours. Nature 406(6797): 747-752.

Perreard, L., Fan, C., Quackenbush, J. F., Mullins, M., Gauthier, N. P., Nelson, E., Mone, M., Hansen, H., Buys, S. S., Rasmussen, K., Orrico, A. R., Dreher, D., Walters, R., Parker, J., Hu, Z., He, X., Palazzo, J. P., Olopade, O. I., Szabo, A., Perou, C. M. and Bernard, P. S. (2006). Classification and risk stratification of invasive breast carcinomas using a real-time quantitative RT-PCR assay. Breast Cancer Res 8(2): R23.

Quinn, J. E., Kennedy, R. D., Mullan, P. B., Gilmore, P. M., Carty, M., Johnston, P. G. and Harkin, D. P. (2003). BRCA1 functions as a differential modulator of chemotherapy-induced apoptosis. Cancer Res 63(19): 6221-6228.

Salomon, D. S., Brandt, R., Ciardiello, F. and Normanno, N. (1995). Epidermal growth factor-related peptides and their receptors in human malignancies. Crit Rev Oncol Hematol 19(3): 183-232.

Sartor, C. I., Dziubinski, M. L., Yu, C. L., Jove, R. and Ethier, S. P. (1997). Role of epidermal growth factor receptor and STAT-3 activation in autonomous proliferation of SUM-102PT human breast cancer cells. Cancer Res 57(5): 978-987.

Schraml, P., Bucher, C., Bissig, H., Nocito, A., Haas, P., Wilber, K., Seelig, S., Kononen, J., Mihatsch, M. J., Dirnhofer, S. and Sauter, G. (2003). Cyclin E overexpression and amplification in human tumours. J Pathol 200(3): 375-382.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Res. 13(11): 2498-2504.

She, Q. B., Solit, D., Basso, A. and Moasser, M. M. (2003). Resistance to gefitinib in PTEN-null HER-overexpressing tumor cells can be overcome through restoration of PTEN function or pharmacologic modulation of constitutive phosphatidylinositol 3'-kinase/Akt pathway signaling. Clin Cancer Res 9(12): 4340-4346.

Shou, J., Massarweh, S., Osborne, C. K., Wakeling, A. E., Ali, S., Weiss, H. and Schiff, R. (2004). Mechanisms of tamoxifen resistance: increased estrogen receptor-HER2/neu cross-talk in ER/HER2-positive breast cancer. J Natl Cancer Inst 96(12): 926-935.

Sieuwerts, A. M., Look, M. P., Meijer-van Gelder, M. E., Timmermans, M., Trapman, A. M., Garcia, R. R., Arnold, M., Goedheer, A. J., de Weerd, V., Portengen, H., Klijn, J. G. and Foekens, J. A. (2006). Which cyclin E prevails as prognostic marker for breast cancer? Results from a retrospective study involving 635 lymph node-negative breast cancer patients. Clin Cancer Res 12(11 Pt 1): 3319-3328.

Singh, A. B. and Harris, R. C. (2005). Autocrine, paracrine and juxtacrine signaling by EGFR ligands. Cell Signal 17(10): 1183-1193.

Slamon, D. J., Leyland-Jones, B., Shak, S., Fuchs, H., Paton, V., Bajamonde, A., Fleming, T., Eiermann, W., Wolter, J., Pegram, M., Baselga, J. and Norton, L. (2001). Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. N Engl J Med 344(11): 783-792.

Sørlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Thorsen, T., Quist, H., Matese, J. C., Brown, P. O., Botstein, D., Eystein Lonning, P. and Borresen-Dale, A. L. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 98(19): 10869-10874.

Sørlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J. S., Nobel, A., Deng, S., Johnsen, H., Pesich, R., Geisler, S., Demeter, J., Perou, C. M., Lonning, P. E., Brown, P. O., Borresen-Dale, A. L. and Botstein, D. (2003). Repeated observation of breast tumor

subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A 100(14): 8418-8423.

Tamura, K. and Fukuoka, M. (2003). Molecular target-based cancer therapy: tyrosine kinase inhibitors. Int J Clin Oncol 8(4): 207-211.

Tassone, P., Tagliaferri, P., Perricelli, A., Blotta, S., Quaresima, B., Martelli, M. L., Goel, A., Barbieri, V., Costanzo, F., Boland, C. R. and Venuta, S. (2003). BRCA1 expression modulates chemosensitivity of BRCA1-defective HCC1937 human breast cancer cells. Br J Cancer 88(8): 1285-1291.

Tokunaga, E., Kimura, Y., Oki, E., Ueda, N., Futatsugi, M., Mashino, K., Yamamoto, M., Ikebe, M., Kakeji, Y., Baba, H. and Maehara, Y. (2006). Akt is frequently activated in HER2/neu-positive breast cancers and associated with poor prognosis among hormone-treated patients. Int J Cancer 118(2): 284-289.

Troester, M. A., Hoadley, K. A., Parker, J. S. and Perou, C. M. (2004). Prediction of toxicant-specific gene expression signatures after chemotherapeutic treatment of breast cell lines. Environ Health Perspect 112(16): 1607-1613.

Troester, M. A., Hoadley, K. A., Sorlie, T., Herbert, B.-S., Borresen-Dale, A.-L., Lonning, P. E., Shay, J. W., Kaufmann, W. K. and Perou, C. M. (2004). Cell-Type-Specific Responses to Chemotherapeutics in Breast Cancer. Cancer Res 64(12): 4218-4226.

Tusher, V., Tibshirani, R. and Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A 98(9): 5116-5121.

van de Vijver, M. J., He, Y. D., van't Veer, L. J., Dai, H., Hart, A. A., Voskuil, D. W., Schreiber, G. J., Peterse, J. L., Roberts, C., Marton, M. J., Parrish, M., Atsma, D., Witteveen, A., Glas, A., Delahaye, L., van der Velde, T., Bartelink, H., Rodenhuis, S., Rutgers, E. T., Friend, S. H. and Bernards, R. (2002). A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 347(25): 1999-2009.

von Minckwitz, G., Jonat, W., Fasching, P., du Bois, A., Kleeberg, U., Luck, H. J., Kettner, E., Hilfrich, J., Eiermann, W., Torode, J. and Schneeweiss, A. (2005). A multicentre phase II study on gefitinib in taxane- and anthracycline-pretreated metastatic breast cancer. Breast Cancer Res Treat 89(2): 165-172.

Weigelt, B., Hu, Z., He, X., Livasy, C., Carey, L. A., Ewend, M. G., Glas, A. M., Perou, C. M. and Van't Veer, L. J. (2005). Molecular portraits and 70-gene prognosis signature are preserved throughout the metastatic process of breast cancer. Cancer Res 65(20): 9155-9158.

Whitfield, M. L., George, L. K., Grant, G. D. and Perou, C. M. (2006). Common markers of proliferation. Nat Rev Cancer 6(2): 99-106.

Whitfield, M. L., Sherlock, G., Saldanha, A. J., Murray, J. I., Ball, C. A., Alexander, K. E., Matese, J. C., Perou, C. M., Hurt, M. M., Brown, P. O. and Botstein, D. (2002). Identification of genes periodically expressed in the human cell cycle and their expression in tumors. Mol Biol Cell 13(6): 1977-2000.

Yarden, Y. and Sliwkowski, M. X. (2001). Untangling the ErbB signalling network. Nat Rev Mol Cell Biol 2(2): 127-137.

Zhou, C., Huang, P. and Liu, J. (2005). The carboxyl-terminal of BRCA1 is required for subnuclear assembly of RAD51 after treatment with cisplatin but not ionizing radiation in human breast and ovarian cancer cells. Biochem Biophys Res Commun 336(3): 952-960.

Zhou, X., Tan, M., Stone Hawthorne, V., Klos, K. S., Lan, K. H., Yang, Y., Yang, W., Smith, T. L., Shi, D. and Yu, D. (2004). Activation of the Akt/mammalian target of rapamycin/4E-BP1 pathway by ErbB2 overexpression predicts tumor progression in breast cancers. Clin Cancer Res 10(20): 6779-6788.

CHAPTER V

CONCLUSION

Many breast cancer gene expression studies have shown that regardless of tumor tissue collection method, RNA extraction method, or microarray platform, the identification of the intrinsic subtypes is robust (Perou et al., 2000; Sørlie et al., 2001; Sørlie et al., 2003; Sotiriou et al., 2003; Rouzier et al., 2005; Hu et al., 2006). The classification based on the inherent molecular gene expression differences is also predictive of outcome (Sørlie et al., 2001; Sørlie et al., 2003; Fan et al., 2006). Two of these subtypes are characterized by expression of markers, ER+ (luminal tumors) and HER2+, which have been used to guide treatment. In addition, there are at least two types of epithelial cells that can be identified by cytokeratin staining and can distinguish the luminal and basal-like tumor subtypes (Perou et al., 2000). This heterogeneity implies that treatments should vary according to subtype, and that responses to the same drug may also vary according to subtype.

My research focused on the basal-like subtype, its comparison with the more prevalent luminal subtype in response to chemotherapy, and on the identification of genes or pathways that could be targeted for therapeutic intervention. Since the basal-like subtype lacks ER expression and HER2 amplification, it cannot be treated with the widely used targeted biologics of tamoxifen (or aromatase inhibitors) and trastuzumab, and hence, cytotoxic chemotherapy regimens are the only options. At the beginning of my research most preclinical and clinical data had not specifically evaluated whether any differences in

response to chemotherapy occurred among the tumor subtypes. Therefore, we addressed this issue by using cell line models of basal-like and luminal tumors to examine their transcriptional responses to commonly prescribed chemotherapeutics. When any cell line was treated with a chemotherapeutic agent, the predominant response in both luminal and basal-like lines was a general stress response (Troester et al., 2004). However, the general stress response varied quite dramatically between the two subtypes. The luminal subtype induced a greater number of genes and more dynamic response to chemotherapy with down-regulation of proliferation genes and the strong induction of DNA damage-response genes, including $p21^{waf1}$ and *GADD45*. The basal-like subtype had a much reduced response without the observed reduction in proliferation genes, which suggests that the basal-like cell type might have a blunted G1-S checkpoint response. The basal-like cell lines also had a much smaller induction of DNA damage-response genes, which showed fold-changes much less than those observed in the luminal cell lines. This observation was also visible in the *in vivo* tumor data obtained by studying the expression response in tumors assayed by microarray before and after neoadjuvant treatment with either 5FU or DOX.

The dominant transcriptional response to chemotherapeutics was a general stress response because it was similarly induced when these two distinct chemotherapeutics were used (5FU and DOX); however, a much smaller drug-specific response was identified that correlated with DOX or 5FU treatment and which showed many more similarities between the two subtypes (Troester et al., 2004). Some of the genes that made up the drug-specific patterns correlated with data observed from other researchers and some were new with no current correlation with the known mechanism of action. We evaluated whether these drug-specific signatures could be used for the classification of a new drug that showed a similar

mechanism of action. Using several prediction classifiers we could accurately predict etoposide, a topoisomerase II inhibitor, as being highly similar to doxorubicin, which is also a topoisomerase II inhibitor. We were also able to classify and predict four classes (subtype and treatment) in our training and test sets. This hints at mechanistic differences between subtypes and within individual cell lines, which again suggests that the basal-like and luminal cell types respond differently and have different chemosensitivities.

While our gene expression analysis was not set up to dictate treatment, the expanded chemosensitivity data with additional cell lines and chemotherapeutics gave some suggestions for subtype specific therapies. There were many individual variations across the cell lines concerning chemosensitivity; however, the tumor-derived basal-like cell lines were more sensitive to carboplatin than the normal HMECs and the luminal cell lines. As discussed in Chapter 4, platinum derivatives may represent a good chemotherapeutic based on the biology of the basal-like subtype and their link with germline BRCA1 mutations. The BRCA1 protein is responsible for platinum-based DNA damage repair (Bhattacharyya et al., 2000; Zhou et al., 2005) and there is increasing evidence that breast tumors with BRCA1 mutations or reduced expression are almost always of basal-like subtype (Foulkes et al., 2003; Sørlie et al., 2003; Arnes et al., 2005; Turner et al., 2006). Therefore, treatment with a platinum-based therapy could be targeted to basal-like breast tumors. This is precisely the chemotherapeutic that is being used in the UNC initiated "triple-negative" patient trial where metastatic basal-like patients will be treated with cetuximab alone, or cetuximab plus carboplatin [http://www.clinicaltrials.gov/ct/show/NCT00232505].

Since the basal-like and luminal cell types varied in their chemosensitivities, both in cell lines and tumors, and had different general stress responses, it is important to analyze

multiple clinical endpoints when assessing whether a given regimen has been "successful". For example, recently several papers have examined clinical response in relation to subtype. Rouzier et al. (Rouzier et al., 2005) used fine needle aspirates of 82 stage I-III breast tumors before treatment with neoadjuvant paclitaxel followed by 5FU, DOX, and cyclophosphamide and determined subtype by hierarchical clustering with a breast tumor intrinsic gene list. Pathological clinical response (pCR) strongly correlated with subtype with 45% of both HER2+/ER- and basal-like tumors showing a complete pCR, while only 6% of the luminal and none of the normal-like exhibited pCR. If the basal-like and HER2+/ER- subtypes have previously been shown to have the poorest outcomes, as seen in both treated and untreated datasets (Sørlie et al., 2001; Sørlie et al., 2003; Fan et al., 2006; Hu et al., 2006), why then would they have the best response to chemotherapy? Carey et al. (Carey et al., Submitted) further confirmed that the higher clinical responses and pCR to neoadjuvant DOX and cyclophosphamide were in the HER2+/ER- and basal-like subtype; however, they went on to record relapse free and overall survival in these patients and found that the basal-like and HER2+/ER- patients that did not have a pCR still had higher rates of relapse (while relapse of patients with pCR was rare). Therefore, they showed that while the basal-like and HER2+/ER- have higher pCR rates, those that do not achieve a complete response have a much higher chance of recurring relative to the luminal tumors, and therefore still have a poor outcome. These analyses suggest that while pCR is an important endpoint, relapse and overall survival information is still extremely valuable data.

Currently, cytotoxic therapy is the only treatment option for basal-like patients, and those that do not achieve a complete response to chemotherapy have few treatment options. Therefore, the next step is to determine if any biologically targetable proteins exist within the

basal-like tumors that could be used to improve therapeutic options. EGFR, a receptor tyrosine kinase, is highly expressed in about 50% of basal-like tumors and represents a good candidate for targeted therapy because many drugs already exist within the oncology clinic that target this protein (Baselga, 2002; Nielsen et al., 2004). Observations of our cell line and tumor data also indicate high EGFR mRNA and protein expression in most of the basal-like lines, whereas no expression was observed in the luminal lines. We determined that the basal-like cell lines were more sensitive to several EGFR inhibitors relative to the luminal lines, ranging from 2- to 100-fold less sensitive. These analyses further suggest the EGFR pathway may represent a biologically target therapy for basal-like patients.

To further evaluate the EGFR pathway in breast cancers, we used a cell line model of the basal-like subtype that was sensitive to EGFR, MEK, and PI3K inhibitors and developed a gene expression signature for EGFR pathway activation. We analyzed this homogenous cell line pattern across a large panel of primary breast tumors and noted that the activation signature was not a single and homogeneous pattern across the tumor samples. Three different signatures were identified that were highest in three different subsets of tumors, all of which showed that the high expression of each gene set significantly predicted a poor outcome. While only 50% of the basal-like tumors had high protein expression of EGFR, greater than 95% of this subtype had high expression for at least one of the three EGFR-activation signatures. This is evidence that the EGFR signaling pathway may represent a major and consistent signaling pathway for growth in the basal-like subtype akin to ER for luminal tumors and HER2 for HER2+/ER- tumors.

These EGFR pathway analyses also identified genes whose high expression has been implicated in ligand independent activation of the EGFR-RAS-MEK pathway and that were

correlated with the basal-like subtype and the EGFR-activation signatures. These genes included high expression of an amplicon on chromosome 12p that contains KRAS and at least eleven other genes, MEK, ERK, and CRYAB as identified by (Herschkowitz et al.). Overexpression of CRYAB has been found to constitutively activate the MEK/ERK pathway and confer EGF-independent growth on an EGF-dependent cell line (Moyano et al., 2006). Only a few basal-like tumors expressed EGFR and one of its ligands without downstream activation of the MEK/ERK pathway via high CRYAB or high KRAS-amplicon expression signature. Most likely these are the only basal-like tumors that would respond to direct EGFR inhibition, which must be tested using EGFR inhibitor-treated breast cancer patients.

In addition to the basal-like tumors showing high expression of the EGFR-activation signatures, half of the luminal B and HER2+/ER- tumors also showed high expression of these signatures. Since HER2 is an EGFR family member, and HER2 and EGFR often heterodimerize to send the activation signal, it is logical that there are shared EGFR-activation signatures across subtypes. Luminal B tumors, which by gene expression have higher levels of proliferation genes, may be using "nongenomic" mechanisms of ER for enhanced growth through the HER pathway (Schiff et al., 2005). The observations of poor outcomes in luminal B patients when given endocrine therapy could be the result of activation of the EGFR-HER2 pathway.

The EGFR-activation signatures identify tumors with a presumed activation of the EGFR-RAS-MEK pathway and predict a poor outcome. We hypothesize that these signatures might also be important in predicting response to EGFR inhibitors since EGFR expression itself does not predict response (Fountzilas et al., 2005; Gasparini et al., 2005; von Minckwitz et al., 2005). We have been limited in our ability to correlate our EGFR

activation signatures with inhibitor-treated *in vivo* datasets because of the lack of such data sets. Currently, there is only one publicly available microarray data set of breast tumors treated with an EGFR inhibitor. However, that study used an unselected population that contained only 10 tumors of which only one was EGFR positive by IHC, and no clinical responses were observed (Yang et al., 2005). As mentioned in Chapter 4, there is a phase II metastatic study of breast cancer patients who are being randomized to cetuximab alone versus cetuximab plus carboplatin [http://www.clinicaltrials.gov/ct/show/NCT00232505]. This study is selecting for basal-like tumors; and therefore, we will be able to assess our EGFR activation signature and its ability to predict response on the subset (approximately 20-30%) of patients from whom we are able to obtain a tumor biopsy. Our lab will be receiving tumor samples at several stages through the treatment process and will be performing gene expression analysis and comparing these data with the clinical response data. We would also like to assess the expression of the genes we identified as potential EGFR-independent activators of the EGFR pathway (i.e. CRYAB and KRAS-amplicon) and determine if they correlate with a lack of response. We predict that only 10% of all basal-like tumors will actually respond to direct inhibition of EGFR. Indeed, if any show a response to cetuximab, we can evaluate our EGFR signature; however, the lack of response in the presence of the EGFR-activation signatures will demonstrate that this pathway is activated downstream of EGFR.

If basal-like patients fail to respond to direct EGFR inhibition as we predict, then therapeutic intervention downstream of EGFR may be a better option. While gene expression activation signatures of the MEK/ERK pathway and the PI3K/AKT pathway did not appear to be that dissimilar, high expression of many genes known to activate the

MEK/ERK pathway were more significantly correlated with the activation signatures. Studies with CRYAB showed that only MEK inhibitors, and not PIK3CA inhibitors, could overcome the reversal of the transforming effects of high CRYAB expression (Moyano et al., 2006). Many second-generation MEK inhibitors are currently in phase I and II clinical trials for a variety of tumors types, including breast (NCT00147550 and NCT00174369) (Sebolt-Leopold and Herrera, 2004). Studies with mTOR inhibitors, a downstream target in the PI3K/AKT, are also actively recruiting patients (NCT00360542). These studies are currently set up as single agent, non-randomized, and unselected patient based studies. It will be interesting to see how patients respond in these clinical trials, and how these responses might correlate with the EGFR-activation signatures. However, since the EGFR signaling pathway is intricate with much cross talk, it will most likely require inhibition at several steps, in combination with chemotherapy, to adequately treat basal-like breast cancers.

In summary, my work has shown that basal-like and luminal breast epithelial cells, both *in vitro* and *in vivo*, respond differently to the same chemotherapeutics, and thus these two distinct type of breast cancer should be considered separately and unique therapies developed for each. To address the need of developing therapies specific for basal-like tumors, a subtype realized only six years ago, we showed that they have a unique chemosensitivity profile (i.e. sensitive to carboplatin and more resistant to 5FU). In addition, the basal-like tumors all showed an EGFR-RAS-MEK pathway activation signature, and thus agents targeting this pathway may represent an effective biologically-based therapy. These studies have provided important pre-clinical data that has lead to the initiation of the first clinical trial focused on basal-like patients (LCCC0403) [http://www.clinicaltrials.gov/ct/show/NCT00232505] and a second trial in the late planning

stage where basal-like patients will receive paclitaxel or paclitaxel plus carboplatin (CALGB 40603). We hope to obtain improved response rates and prolonged survival times in these trials, but at least we have begun the characterization of the biology of the basal-like subtype and have initiated clinical trials aimed at this aggressive tumor subtype.

REFERENCES

Arnes, J. B., Brunet, J. S., Stefansson, I., Begin, L. R., Wong, N., Chappuis, P. O., Akslen, L. A. and Foulkes, W. D. (2005). Placental cadherin and the basal epithelial phenotype of BRCA1-related breast cancer. Clin Cancer Res 11(11): 4003-4011.

Baselga, J. (2002). Why the epidermal growth factor receptor? The rationale for cancer therapy. Oncologist 7 Suppl 4: 2-8.

Bhattacharyya, A., Ear, U. S., Koller, B. H., Weichselbaum, R. R. and Bishop, D. K. (2000). The breast cancer susceptibility gene BRCA1 is required for subnuclear assembly of Rad51 and survival following treatment with the DNA cross-linking agent cisplatin. J Biol Chem 275(31): 23899-23903.

Carey, L., Dees, E. C., Sawyer, L. R., Gatti, L., Moore, D. T., Collicho, F., Ollila, D. W. and Perou, C. M. (Submitted). The triple negative paradox: Primary tumor chemosensitivity of breast cancer subtypes. Clin Cancer Res.

Fan, C., Oh, D. S., Wessels, L., Weigelt, B., Nuyten, D. S., Nobel, A. B., van't Veer, L. J. and Perou, C. M. (2006). Concordance among gene-expression-based predictors for breast cancer. N Engl J Med 355(6): 560-569.

Foulkes, W. D., Stefansson, I. M., Chappuis, P. O., Begin, L. R., Goffin, J. R., Wong, N., Trudel, M. and Akslen, L. A. (2003). Germline BRCA1 mutations and a basal epithelial phenotype in breast cancer. J Natl Cancer Inst 95(19): 1482-1485.

Fountzilas, G., Pectasides, D., Kalogera-Fountzila, A., Skarlos, D., Kalofonos, H. P., Papadimitriou, C., Bafaloukos, D., Lambropoulos, S., Papadopoulos, S., Kourea, H., Markopoulos, C., Linardou, H., Mavroudis, D., Briasoulis, E., Pavlidis, N., Razis, E., Kosmidis, P. and Gogas, H. (2005). Paclitaxel and carboplatin as first-line chemotherapy combined with gefitinib (IRESSA) in patients with advanced breast cancer: a phase I/II study conducted by the Hellenic Cooperative Oncology Group. Breast Cancer Res Treat 92(1): 1-9.

Gasparini, G., Sarmiento, R., Amici, S., Longo, R., Gattuso, D., Zancan, M. and Gion, M. (2005). Gefitinib (ZD1839) combined with weekly epirubicin in patients with metastatic breast cancer: a phase I study with biological correlate. Ann Oncol 16(12): 1867-1873.

Herschkowitz, J. I., Simin, K., Weigman, V. J., Mikaelian, I., Hu, Z., Rasmussen, K. E., Chandrasekharan, S., Backlund, M. G., Yin, Y., Glazer, R. I., Brown, P. H., Green, J. E., Kopelovich, K., Churchill, G. A., Van Dyke, T. and Perou, C. M. Identification of conserved gene expression features across human and murine mammary tumors. Submitted.

Hu, Z., Fan, C., Oh, D. S., Marron, J. S., He, X., Qaqish, B. F., Livasy, C., Carey, L. A., Reynolds, E., Dressler, L., Nobel, A., Parker, J., Ewend, M. G., Sawyer, L. R., Wu, J., Liu, Y., Nanda, R., Tretiakova, M., Ruiz Orrico, A., Dreher, D., Palazzo, J. P., Perreard, L., Nelson, E., Mone, M., Hansen, H., Mullins, M., Quackenbush, J. F., Ellis, M. J., Olopade, O. I., Bernard, P. S. and Perou, C. M. (2006). The molecular portraits of breast tumors are conserved across microarray platforms. BMC Genomics 7: 96.

Moyano, J. V., Evans, J. R., Chen, F., Lu, M., Werner, M. E., Yehiely, F., Diaz, L. K., Turbin, D., Karaca, G., Wiley, E., Nielsen, T. O., Perou, C. M. and Cryns, V. L. (2006). AlphaB-crystallin is a novel oncoprotein that predicts poor clinical outcome in breast cancer. J Clin Invest 116(1): 261-270.

Nielsen, T. O., Hsu, F. D., Jensen, K., Cheang, M., Karaca, G., Hu, Z., Hernandez-Boussard, T., Livasy, C., Cowan, D., Dressler, L., Akslen, L. A., Ragaz, J., Gown, A. M., Gilks, C. B., van De Rijn, M. and Perou, C. M. (2004). Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. Clin Cancer Res 10(16): 5367-3574.

Perou, C. M., Brown, P. O. and Botstein, D. (2000). Tumor classification using gene expression patterns from DNA microarrays. New Technologies for life sciences: A Trends Guide: 67-76.

Rouzier, R., Perou, C. M., Symmans, W. F., Ibrahim, N., Cristofanilli, M., Anderson, K., Hess, K. R., Stec, J., Ayers, M., Wagner, P., Morandi, P., Fan, C., Rabiul, I., Ross, J. S., Hortobagyi, G. N. and Pusztai, L. (2005). Breast cancer molecular subtypes respond differently to preoperative chemotherapy. Clin Cancer Res 11(16): 5678-5685.

Schiff, R., Massarweh, S. A., Shou, J., Bharwani, L., Arpino, G., Rimawi, M. and Osborne, C. K. (2005). Advanced concepts in estrogen receptor biology and breast cancer endocrine resistance: implicated role of growth factor signaling and estrogen receptor coregulators. Cancer Chemother Pharmacol 56 Suppl 1: 10-20.

Sebolt-Leopold, J. S. and Herrera, R. (2004). Targeting the mitogen-activated protein kinase cascade to treat cancer. Nat Rev Cancer 4(12): 937-947.

Sørlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Thorsen, T., Quist, H., Matese, J. C., Brown, P. O., Botstein, D., Eystein Lonning, P. and Borresen-Dale, A. L. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 98(19): 10869-10874.

Sørlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J. S., Nobel, A., Deng, S., Johnsen, H., Pesich, R., Geisler, S., Demeter, J., Perou, C. M., Lonning, P. E., Brown, P. O., Borresen-Dale, A. L. and Botstein, D. (2003). Repeated observation of breast tumor

subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A 100(14): 8418-8423.

Sotiriou, C., Neo, S. Y., McShane, L. M., Korn, E. L., Long, P. M., Jazaeri, A., Martiat, P., Fox, S. B., Harris, A. L. and Liu, E. T. (2003). Breast cancer classification and prognosis based on gene expression profiles from a population-based study. Proc Natl Acad Sci U S A 100(18): 10393-10398.

Troester, M. A., Hoadley, K. A., Parker, J. S. and Perou, C. M. (2004). Prediction of toxicant-specific gene expression signatures after chemotherapeutic treatment of breast cell lines. Environ Health Perspect 112(16): 1607-1613.

Troester, M. A., Hoadley, K. A., Sorlie, T., Herbert, B.-S., Borresen-Dale, A.-L., Lonning, P. E., Shay, J. W., Kaufmann, W. K. and Perou, C. M. (2004). Cell-Type-Specific Responses to Chemotherapeutics in Breast Cancer. Cancer Res 64(12): 4218-4226.

Turner, N. C., Reis-Filho, J. S., Russell, A. M., Springall, R. J., Ryder, K., Steele, D., Savage, K., Gillett, C. E., Schmitt, F. C., Ashworth, A. and Tutt, A. N. (2006). BRCA1 dysfunction in sporadic basal-like breast cancer. Oncogene.

von Minckwitz, G., Jonat, W., Fasching, P., du Bois, A., Kleeberg, U., Luck, H. J., Kettner, E., Hilfrich, J., Eiermann, W., Torode, J. and Schneeweiss, A. (2005). A multicentre phase II study on gefitinib in taxane- and anthracycline-pretreated metastatic breast cancer. Breast Cancer Res Treat 89(2): 165-172.

Yang, S. X., Simon, R. M., Tan, A. R., Nguyen, D. and Swain, S. M. (2005). Gene expression patterns and profile changes pre- and post-erlotinib treatment in patients with metastatic breast cancer. Clin Cancer Res 11(17): 6226-6232.

Zhou, C., Huang, P. and Liu, J. (2005). The carboxyl-terminal of BRCA1 is required for subnuclear assembly of RAD51 after treatment with cisplatin but not ionizing radiation in human breast and ovarian cancer cells. Biochem Biophys Res Commun 336(3): 952-960.

APPENDIX IIA

Genes significantly altered by treatment across all tumor subtypes as determined by Significance Analysis of Microarrays[a]

Gene	Accession Number	Average Ratio Before	Average Ratio After
activating transcription factor 3	H21041	0.10	0.93
carboxypeptidase X	AA598945	0.13	0.87
connective tissue growth factor	AA598794	-0.78	0.39
connective tissue growth factor	AA044993	-0.93	0.35
core promoter element binding protein	AA013481	0.44	1.06
core promoter element binding protein	AA055584	0.35	0.71
corticotropin releasing hormone binding protein	AA286752	0.16	0.91
cyclin-dependent kinase 5, regulatory subunit 1 (p35)	AA442853	0.73	1.45
cyclin-dependent kinase inhibitor 1A, $p21^{waf1}$	N23941	-0.02	0.96
dihydropyrimidinase-like 3	AI831083	0.20	1.13
**dopachrome tautomerase	AA478553	-0.51	0.75
early growth response 1	AA486533	-0.30	1.27
early growth response 2	AA446027	0.63	1.15
early response protein NAK1	N94487	0.26	1.14
elastin	AA459308	1.19	1.99
elongin A	AA128607	-0.82	0.38
fibulin 1	AA134757	1.28	2.07
FOS	N36944	0.07	0.87
FOS	AA485377	-0.20	0.69
FOS	R12840	-0.30	1.22
Homo sapiens mRNA	AA135912	0.88	1.47
JUN	W96134	0.91	1.62
JUN	AA293362	0.82	1.37
GTP-binding protein overexpressed in skeletal muscle	AA418077	0.76	1.38
matrix metalloproteinase 9	T64837	2.28	3.14
RAB21	AA076645	-0.69	0.36
regulator of G-protein signaling 1	AA017417	2.53	3.72
regulator of G-protein signaling 16	AA453774	0.97	1.56

[a]Data has not been median centered.

**A potentially chimeric cDNA clone that maps to two different Unigene entries.

APPENDIX IIB

Genes significantly altered by therapy in luminal tumors as determined by Significance Analysis of Microarrays[a]

Gene Name	Accession Number	Average Ratio Before	Average Ratio After
connective tissue growth factor	AA598794	-0.92	0.20
connective tissue growth factor	AA044993	-1.08	0.12
corticotropin releasing hormone binding protein	AA286752	0.09	0.75
cyclin-dependent kinase 5, regulatory subunit 1 (p35)	AA442853	0.57	1.36
cyclin-dependent kinase inhibitor 1A, $p21^{wafl}$	N23941	0.05	1.13
dihydropyrimidinase-like 3	AI831083	0.006	0.96
**dopachrome tautomerase	AA478553	-0.72	0.44
early growth response 1	AA486533	-0.38	1.10
early response protein NAK1	N94487	0.21	1.09
elongin A	AA128607	-0.99	0.17
FOS	N36944	0.08	0.82
Kinase-inducible Ras-like protein	AA418077	0.57	1.31
**prostate differentiation factor	N26311	-1.63	-0.46
spondin 1	H09099	-2.54	-1.55
thrombospondin 1	AA464532	0.09	0.79

[a]Data has not been median-centered.
**A potentially chimeric cDNA clone that maps to two different Unigene entries.

APPENDIX IIC

Genes altered by therapy in basal tumors as determined by Significance Analysis of Microarrays[a]

Gene Name	Accession Number	Average Ratio Before	Average Ratio After
connective tissue growth factor	AA598794	-0.79	0.57
connective tissue growth factor	AA044993	-0.97	0.81
core promoter element binding protein	AA013481	0.97	1.81
dermatan sulfate proteoglycan 3	AA131238	0.03	1.11
early growth response 1	AA486533	-0.69	1.01
early response protein NAK1	N94487	0.14	1.41
elongin A	AA128607	-0.77	0.88
FOS	R12840	-0.63	1.65
Homo sapiens mRNA	AA135912	1.19	2.16
RAB21	AA076645	-1.20	0.79

[a]Data has not been median-centered.
**A potentially chimeric cDNA clone that maps to two different Unigene entries.

APPENDIX IIIA

Results of Cross-Validation (CV) Analyses using Two-Class PAM Method

Delta Value	Average No. of Genes	Average CV Accuracy (%)
4	1123.1	64
3.5	1928.3	76
3	2323.6	78
2.75	2460.4	80
2.5	2584.6	76
2.25	2706.8	76
2	2829.4	76
1.75	2948	76
1.5	3075	76
1.25	3213.9	76
1	3371.1	74
0.75	3543	70
0.5	3716.2	70
0.25	3902.5	70
0	4077	70

APPENDIX IIIB

Cross-Validation (CV) Accuracy (%) Using Two-Class KNN Method

No. of Genes	*k*=	1	3	5	7	9	11
10		58	72	72	66	66	64
30		84	92	90	92	94	94
50		74	80	90	96	94	92
70		80	86	92	92	90	92
100		80	82	94	94	96	98
200		74	76	86	88	88	90
4077		66	64	70	68	68	76

APPENDIX IIIC

Results of Cross-Validation (CV) Analyses Using Four-Class PAM Method

Delta Value	Average No. of Genes	Average CV Accuracy (%)
6	17.4	66
5	45.6	68
4	136.9	76
3.5	652	76
3	1830.7	50
2.75	2211.2	46
2.5	2510.9	42
2.25	2763.2	48
2	2975.5	48
1.75	3183.5	48
1.5	3379.4	48
1.25	3549.1	46
1	3693.8	46
0.75	3826.3	46
0.5	3944.6	46
0.25	4037.7	42
0	4077	38

APPENDIX IIID

Cross-Validation (CV) Accuracy (%) Using Four-Class KNN Method

No. of Genes	*k*=	1	3	5	7	9	11
10		40	50	48	54	56	56
30		60	62	76	76	72	76
50		66	78	74	70	70	74
70		62	74	76	76	78	68
100		58	66	70	70	80	78
200		54	58	64	62	76	74
4077		48	48	48	46	48	52

APPENDIX IVA

Genes from clusters #1-3 identified from the 500 SUM102 genes clustered on the UNC tumor data set.

Cluster #1

Gene Symbol	Gene Name	Accession #	ClusterID
MGC11256	Hypothetical protein MGC11256	NM_024324	Hs.211282
FLJ20397	Hypothetical protein FLJ20397	NM_017802	Hs.521328
GAL	Galanin	BC030241	Hs.278959
		AC010889	
ALDH1B1	Aldehyde dehydrogenase 1 family, member B1	NM_000692	Hs.436219
NKX2-5	NK2 transcription factor related, locus 5 (Drosophila)	BC025711	Hs.54473
SERPINB8	Serine (or cysteine) proteinase inhibitor, clade B (ovalbumin), member 8	NM_198833	Hs.368077
MTP18	Mitochondrial protein 18 kDa	NM_001003704	Hs.25199
POLR3B	Polymerase (RNA) III (DNA directed) polypeptide B	NM_018082	Hs.62696
FLJ14800	Hypothetical protein FLJ14800	NM_032840	Hs.343334
FARSLA	Phenylalanine-tRNA synthetase-like, alpha subunit	NM_004461	Hs.23111
NOL5A	Nucleolar protein 5A (56kDa with KKE/D repeat)	NM_006392	Hs.376064
NCLN	Nicalin homolog (zebrefish)	NM_020170	Hs.501420
SGTA	Small glutamine-rich tetratricopeptide repeat (TPR)-containing, alpha	NM_003021	Hs.203910
PPM1G	Protein phosphatase 1G (formerly 2C), magnesium-dependent, gamma isoform	NM_002707	Hs.17883
PPAN	Peter pan homolog (Drosophila)	NM_020230	Hs.14468
ZMYND11	Zinc finger, MYND domain containing 11	NM_212479	Hs.292265
MTA2	Metastasis associated gene family, member 2	NM_004739	Hs.173043
BRI3BP	BRI3 binding protein	NM_080626	Hs.507227
GRPEL1	GrpE-like 1, mitochondrial (E. coli)	NM_025196	Hs.443723
STIP1	Stress-induced-phosphoprotein 1 (Hsp70/Hsp90-organizing protein)	NM_006819	Hs.337295
LYAR	Hypothetical protein FLJ20425	NM_017816	Hs.425427
TRIM25	Tripartite motif-containing 25	NM_005082	Hs.528952
RBM28	RNA binding motif protein 28	NM_018077	Hs.274263
CAD	Carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase	NM_004341	Hs.377010
PRO2949	Hypothetical protein PRO2949	AF119907	Hs.391480
LRRC14	Leucine rich repeat containing 14	NM_014665	Hs.459391
RQCD1	RCD1 required for cell differentiation1 homolog (S. pombe)	AC012510	
ARMC6	Armadillo repeat containing 6	NM_033415	Hs.77876
ATAD3A	ATPase family, AAA domain containing 3A	NM_018188	Hs.227067
ATAD3B	ATPase family, AAA domain containing 3B	NM_031921	Hs.23413
B4GALT2	UDP-Gal:betaGlcNAc beta 1,4- galactosyltransferase, polypeptide 2	NM_003780	Hs.474083
TOMM40	Translocase of outer mitochondrial membrane 40 homolog (yeast)	NM_006114	Hs.310542
UPP1	Uridine phosphorylase 1	BC047030	Hs.488240
FLJ12438	Hypothetical protein FLJ12438	AK095928	Hs.8595

Cluster #2

Gene Symbol	Gene Name	Accession #	ClusterID
NLN	Neurolysin (metallopeptidase M3 family)	NM_020726	Hs.247460
WDR12	WD repeat domain 12	NM_018256	Hs.73291
MRPS17	Mitochondrial ribosomal protein S17	NM_015969	Hs.44298
SLC25A19	Solute carrier family 25 (mitochondrial deoxynucleotide carrier), member 19	BC001075	Hs.514470
C6orf66	Chromosome 6 open reading frame 66	NM_014165	Hs.512144
WDR4	WD repeat domain 4	NM_033661	Hs.248815
CDCA7	Cell division cycle associated 7	NM_031942	Hs.470654
BYSL	Bystin-like	NM_004053	Hs.106880
ECE2	Endothelin converting enzyme 2	NM_032331	Hs.146161
TIMM8A	Translocase of inner mitochondrial membrane 8 homolog A (yeast)	NM_004085	Hs.447877
CCNE1	Cyclin E1	NM_057182	Hs.244723
LRP8	Low density lipoprotein receptor-related protein 8, apolipoprotein e receptor	NM_017522	Hs.444637
SNRPD1	Small nuclear ribonucleoprotein D1 polypeptide 16kDa	NM_006938	Hs.464734
KLHL7	Kelch-like 7 (Drosophila)	BC009555	Hs.385861
DC13	DC13 protein	NM_020188	Hs.388255
NOC4	Neighbor of COX4	NM_006067	Hs.173162
FREQ	Frequenin homolog (Drosophila)	NM_014286	Hs.301760
FTHFSDC1	Formyltetrahydrofolate synthetase domain containing 1	AL117452	Hs.268698
PTDSS1	Phosphatidylserine synthase 1	NM_014754	Hs.292579
MGC2574	Hypothetical protein MGC2574	NM_024098	Hs.4253
MGC5352	Hypothetical protein MGC5352	AK097688	Hs.102558
FLJ20989	Hypothetical protein FLJ20989	NM_023080	Hs.169615
LOC51236	Brain protein 16	NM_016458	Hs.300224
BOP1	Block of proliferation 1	NM_015201	Hs.535901
GPR172A	G protein-coupled receptor 172A	NM_024531	Hs.6459
FBXL6	F-box and leucine-rich repeat protein 6	NM_024555	Hs.12271
SIAHBP1	Fuse-binding protein-interacting repressor	NM_078480	Hs.521924

Cluster #3

Gene Symbol	Gene Name	Accession #	ClusterID
LCMT2	Leucine carboxyl methyltransferase 2	NM_014793	Hs.200596
ANXA7	Annexin A7	NM_004034	Hs.386434
CDR2	Cerebellar degeneration-related protein 2, 62kDa	NM_001802	Hs.513430
SOCS6	Suppressor of cytokine signaling 6	NM_004232	Hs.44439
BID	BH3 interacting domain death agonist	NM_197966	Hs.474150
XPOT	Exportin, tRNA (nuclear export receptor for tRNAs)	NM_007235	Hs.85951
DDX20	DEAD (Asp-Glu-Ala-Asp) box polypeptide 20	NM_007204	Hs.485810
HRMT1L6	HMT1 hnRNP methyltransferase-like 6 (S. cerevisiae)	NM_018137	Hs.26006
HMGN4	High mobility group nucleosomal binding domain 4	NM_006353	Hs.236774
SLC30A7	Solute carrier family 30 (zinc transporter), member 7	AI740796	Hs.533903
ACTL6A	Actin-like 6A	NM_004301	Hs.435326
SRPRB	Signal recognition particle receptor, B subunit	NM_021203	Hs.12152
YWHAH	Chromosome 22 open reading frame 24	NM_003405	Hs.226755
C7orf30	Chromosome 7 open reading frame 30	NM_138446	Hs.87385
THUMPD3	THUMP domain containing 3	NM_015453	Hs.443081
HSPC128	HSPC128 protein	NM_014167	Hs.90527
SPATA5L1	Spermatogenesis associated 5-like 1	NM_024063	Hs.369657

Gene Symbol	Gene Name	Accession #	ClusterID
GNB4	Guanine nucleotide binding protein (G protein), beta polypeptide 4	NM_021629	Hs.270543
		NM_001018159	
DSCR2	Down syndrome critical region gene 2	NM_203433	Hs.473838
CCT8	Chaperonin containing TCP1, subunit 8 (theta)	NM_006585	Hs.125113
RBM8A	RNA binding motif protein 8A	NM_005105	Hs.515755
KIAA0179	KIAA0179	D80001	Hs.129621
UCK2	Uridine-cytidine kinase 2	NM_012474	Hs.458360
CTPS	CTP synthase	NM_001905	Hs.473087
GART	Phosphoribosylglycinamide formyltransferase, phosphoribosylglycinamide synthetase, phosphoribosylaminoimidazole synthetase	NM_000819	Hs.473648
HDAC2	Histone deacetylase 2	NM_001527	Hs.3352
ILF2	Interleukin enhancer binding factor 2, 45kDa	NM_004515	Hs.75117
TEX10	Testis expressed sequence 10	NM_017746	Hs.494648
EXOSC3	Exosome component 3	NM_016042	Hs.493887
SERF1A	Small EDRK-rich factor 1A (telomeric)	NM_022978	Hs.32567
POLR3F	Polymerase (RNA) III (DNA directed) polypeptide F, 39 kDa	NM_006466	Hs.472227
CGI-09	CGI-09 protein	NM_015939	Hs.128791
HSA9761	Putative dimethyladenosine transferase	NM_014473	Hs.533222
		NM_001024227	
NKIRAS1	NFKB inhibitor interacting Ras-like 1	NM_020345	Hs.173202
TSN	Translin	NM_004622	Hs.75066
DDX18	DEAD (Asp-Glu-Ala-Asp) box polypeptide 18	NM_006773	Hs.363492
RNASEH1	Ribonuclease H1	NM_002936	Hs.502765
SEC61B	Sec61 beta subunit	NM_006808	Hs.191887
UBE2J1	Ubiquitin-conjugating enzyme E2, J1 (UBC6 homolog, yeast)	NM_016021	Hs.163776
ZCSL2	Zinc finger, CSL domain containing 2	NM_206831	Hs.388087
TPRT	Trans-prenyltransferase	NM_014317	Hs.546357
SUV39H2	Suppressor of variegation 3-9 homolog 2	NM_024670	Hs.85567
RAN	RAN, member RAS oncogene family	NM_006325	Hs.10842
EIF4E	Eukaryotic translation initiation factor 4E	NM_001968	Hs.249718
		NM_001015891	
SFPQ	Splicing factor proline/glutamine rich (polypyrimidine tract binding protein associated)	NM_005066	Hs.355934
PNPT1	Polyribonucleotide nucleotidyltransferase 1	NM_033109	Hs.388733
DNAJA1	DnaJ (Hsp40) homolog, subfamily A, member 1	NM_001539	Hs.445203
FLJ10874	Hypothetical protein FLJ10874	NM_018252	Hs.445386
EIF2S1	Eukaryotic translation initiation factor 2, subunit 1 alpha, 35kDa	NM_004094	Hs.151777
BRIX	BRIX	NM_018321	Hs.38114
METTL2	Methyltransferase like 2	NM_018396	Hs.433213
PSMD12	Proteasome (prosome, macropain) 26S subunit, non-ATPase, 12	NM_174871	Hs.4295
DKFZP586L0724	DKFZP586L0724 protein	NM_015462	Hs.463936
		NM_001009182	
UBE2D1	Ubiquitin-conjugating enzyme E2D 1 (UBC4/5 homolog, yeast)	NM_003338	Hs.129683
PAQR3	Progestin and adipoQ receptor family member III	NM_177453	Hs.368305
RNF138	Ring finger protein 138	NM_198128	Hs.302408
FLJ38973	Hypothetical protein FLJ38973	NM_153689	Hs.471040

Gene Symbol	Gene Name	Accession #	ClusterID
KIF2	Kinesin heavy chain member 2	NM_001011663 NM_004520	Hs.113319
FLJ21908	Hypothetical protein FLJ21908	NM_024604	Hs.437855
C13orf6	Chromosome 13 open reading frame 6	NM_032859	Hs.183528
LOC134218	Hypothetical protein LOC134218	NM_194283	Hs.131887
RG9MTD1	RNA (guanine-9-) methyltransferase domain containing 1	NM_017819	Hs.57898
MRPL50	Mitochondrial ribosomal protein L50	NM_019051	Hs.288224
GFM1	G elongation factor, mitochondrial 1	NM_024996	Hs.518355
MASA	E-1 enzyme	NM_021204	Hs.18442
SYNCRIP	Synaptotagmin binding, cytoplasmic RNA interacting protein	NM_006372	Hs.485877
GTPBP4	GTP binding protein 4	NM_012341	Hs.215766
CML66	Chronic myelogenous leukemia tumor antigen 66	NM_032869	Hs.195870
DKFZP564O0463	DKFZP564O0463 protein	NM_015420	Hs.532265
CGI-12	CGI-12 protein	NM_015942	Hs.308613
MRPL15	Mitochondrial ribosomal protein L15	NM_014175	Hs.18349
CGI-115	CGI-115 protein	NM_016052	Hs.408101
TIMM17A	Translocase of inner mitochondrial membrane 17 homolog A (yeast)	NM_006335	Hs.20716
DKFZp547B1713	Hypothetical protein DKFZp547B1713	NM_152379	Hs.434945
FLJ20533	Hypothetical protein FLJ20533	NM_017866	Hs.106650
SLBP	Stem-loop (histone) binding protein	NM_006527	Hs.298345
UBA2	SUMO-1 activating enzyme subunit 2	XR_000199 NM_005499	Hs.511739
SFRS2	Splicing factor, arginine/serine-rich 2	NM_003016	Hs.73965
HNRPDL	Heterogeneous nuclear ribonucleoprotein D-like	NM_005463	Hs.527105
SFRS10	Splicing factor, arginine/serine-rich 10 (transformer 2 homolog, Drosophila)	NM_001031684 NM_004593	Hs.533122
TFAM	Transcription factor A, mitochondrial	NM_003201	Hs.75133
FLJ14753	Hypothetical protein FLJ14753	NM_032558	Hs.13453
Rif1	Telomere-associated protein RIF1 homolog	NM_018151	Hs.536537
ABCB10	ATP-binding cassette, sub-family B (MDR/TAP), member 10	NM_012089	Hs.17614
HSPD1	Heat shock 60kDa protein 1 (chaperonin)	XM_370704 NM_002156	Hs.113684
UBQLN1	Ubiquilin 1	NM_013438	Hs.9589
MSH6	MutS homolog 6 (E. coli)	NM_000179	Hs.445052
FBXO28	F-box protein 28	NM_015176	Hs.64691
PTS	6-pyruvoyltetrahydropterin synthase	NM_000317	Hs.503860
C10orf119	Chromosome 10 open reading frame 119	NM_024834	Hs.124246
IARS	Isoleucine-tRNA synthetase	NM_013417	Hs.445403
SEH1L	SEH1-like (S. cerevisiae)	NM_031216	Hs.301048
EIF2S2	Eukaryotic translation initiation factor 2, subunit 2 beta, 38kDa	NM_003908	Hs.429180
FAM3C	Family with sequence similarity 3, member C	NM_014888	Hs.434053
HPRT1	Hypoxanthine phosphoribosyltransferase 1 (Lesch-Nyhan syndrome)	NM_000194	Hs.412707
DRG1	Developmentally regulated GTP binding protein 1	NM_004147	Hs.115242
AHSA1	AHA1, activator of heat shock 90kDa protein ATPase homolog 1 (yeast)	NM_012111	Hs.204041
CBFB	Core-binding factor, beta subunit	NM_022845	Hs.460988

Gene Symbol	Gene Name	Accession #	ClusterID
UMPS	Uridine monophosphate synthetase (orotate phosphoribosyl transferase and orotidine-5'-decarboxylase)	NM_000373	Hs.2057
C12orf5	Chromosome 12 open reading frame 5	NM_020375	Hs.504545
FLJ40432	Hypothetical protein FLJ40432	NM_152523	Hs.471234
DEGS	Degenerative spermatocyte homolog, lipid desaturase (Drosophila)	NM_003676	Hs.299878
C6orf93	Chromosome 6 open reading frame 93	NM_032860	Hs.185675
EEF1E1	Eukaryotic translation elongation factor 1 epsilon 1	NM_004280	Hs.88977
ATR	Ataxia telangiectasia and Rad3 related	NM_001184	Hs.271791
DHX15	DEAH (Asp-Glu-Ala-His) box polypeptide 15	NM_001358	Hs.5683
C14orf138	Chromosome 14 open reading frame 138	NM_024558	Hs.546431
CGI-48	CGI-48 protein	NM_016001	Hs.463465
SMYD2	SET and MYND domain containing 2	NM_020197	Hs.66170
CCT2	Chaperonin containing TCP1, subunit 2 (beta)	NM_006431	Hs.189772
FLJ12806	Hypothetical protein FLJ12806	NM_022831	Hs.534965
MAPRE1	Microtubule-associated protein, RP/EB family, member 1	NM_012325	Hs.472437
NOLC1	Nucleolar and coiled-body phosphoprotein 1	D21262	Hs.523238
		AL500527	
C10orf117	Chromosome 10 open reading frame 117	NM_022451	Hs.74899
PPP1R8	Protein phosphatase 1, regulatory (inhibitor) subunit 8	NM_138558	Hs.533474
		AK223118	
HCCS	Holocytochrome c synthase (cytochrome c heme-lyase)	NM_005333	Hs.211571
MGC2714	Hypothetical protein MGC2714	NM_032299	Hs.503716
DKFZP566E144	Small fragment nuclease	NM_015523	Hs.7527
UCHL3	Ubiquitin carboxyl-terminal esterase L3 (ubiquitin thiolesterase)	NM_006002	Hs.162241
HCNGP	Transcriptional regulator protein	NM_013260	Hs.546381
		NM_001008892	
UAP1	UDP-N-acteylglucosamine pyrophosphorylase 1	NM_003115	Hs.492859
C13orf7	Chromosome 13 open reading frame 7	NM_024546	Hs.93956
STRAP	Serine/threonine kinase receptor associated protein	NM_007178	Hs.504895
NCBP1	Nuclear cap binding protein subunit 1, 80kDa	NM_002486	Hs.522309
LSM6	LSM6 homolog, U6 small nuclear RNA associated (S. cerevisiae)	NM_007080	Hs.190520
KPNA1	Karyopherin alpha 1 (importin alpha 5)	AF035311	Hs.161008
ABCE1	ATP-binding cassette, sub-family E (OABP), member 1	NM_002940	Hs.12013
TIMM23	Translocase of inner mitochondrial membrane 23 homolog (yeast)	NM_006327	Hs.524308