5-2016

Expression of genes for peptide/protein hormones and their cognate receptors in breast carcinomas as biomarkers predicting risk of recurrence.

Michael Wesley Daniels

EXPRESSION OF GENES FOR PEPTIDE/PROTEIN HORMONES AND THEIR COGNATE RECEPTORS IN BREAST CARCINOMAS AS BIOMARKERS PREDICTING RISK OF RECURRENCE

By

Michael Wesley Daniels

B.S., University of Louisville, 2014

Thesis Submitted to the Faculty of the
School of Public Health & Information Sciences
in Partial Fulfillment of the Requirements
for the Degree of

Master of Science
in Biostatistics

Department of Bioinformatics and Biostatistics
University of Louisville
May 2016

EXPRESSION OF GENES FOR PEPTIDE/PROTEIN HORMONES AND THEIR COGNATE RECEPTORS IN BREAST CARCINOMAS AS BIOMARKERS PREDICTING RISK OF RECURRENCE

By

Michael Wesley Daniels
B.S., University of Louisville, 2014

A Thesis Approved on

April 1, 2016

by the following Thesis Committee:

_____
Thesis Director
Dr. James L. Wittliff

_____
Thesis Co-Director
Dr. Guy N. Brock

_____
Thesis Co-Director
Dr. Douglas J. Lorenz

_____
Dr. Richard A. Kerber

DEDICATION

This thesis is dedicated to my wife

Mrs. Katharine Marie Daniels

who has believed in me since the day we met.

ACKNOWLEDGMENTS

I would like to thank my co-mentors Dr. Guy Brock and Dr. Jim Wittliff. Dr. Brock's willingness to go above and beyond expectations by assisting so many students such as myself exemplifies not only the type of professor he is but the type of person he is. Dr. Wittliff has forever changed me. His insight and care has guided me towards a better version of myself. I would also like to thank the other committee members, Dr. Doug Lorenz and Dr. Rich Kerber. Many University of Louisville professors in the School of Public Health and the Department of Mathematics in the College of Arts and Sciences have been an integral part of my academic journey. Also, I cannot deny the drive my children, Maggie and Amelia, provide me when they express to me their feelings of pride for their father. Whether it is Maggie saying, "I am distinguished in Math, just like my Daddy" or Amelia dressing up as a biostatistician for career day at her elementary school, I know what I am doing has a positive influence in their lives. Just as any good parent wants their children to have a better life than they had. My parents, Gary Joseph Daniels, Sr. and Connie Lynn Daniels, are witnessing the joys of seeing their son's dream finally come true. With their guidance, I embraced hard work and developed the temperament to ride out so many storms. The seas are now calm and wind is behind our sails. So, I thank you for instilling in me the resolve to never give up.

ABSTRACT

EXPRESSION OF GENES FOR PEPTIDE/PROTEIN HORMONES AND THEIR COGNATE RECEPTORS IN BREAST CARCINOMAS AS BIOMARKERS PREDICTING RISK OF RECURRENCE

Michael W. Daniels

May 14, 2016

Certain hormones and/or receptors influencing normal cellular pathways were detected in breast cancers. The hypothesis is that gene subsets predict risk of breast carcinoma recurrence in patients with primary disease. Gene expression of 55 hormones and 73 receptors were determined by microarray with LCM-procured carcinoma cells of 247 de-identified biopsies. Univariate and multivariate Cox regressions were determined using expression levels of each hormone/receptor gene, individually or as a pair. Significant genes derived for each subset were analyzed to predict risk of cancer recurrence with 1000 LASSO training/test sets. A 14-gene molecular signature was identified for predicting clinical outcome without regard to estrogen or progestin receptor status of biopsies. A three-gene signature was derived for ER+ cancers while a 9-gene signature was deciphered for ER- cancers. Molecular signatures derived were compared with results in public databases. Collectively, results suggest gene subsets in primary breast cancer have been identified that predict recurrence.

TABLE OF CONTENTS

DEDICATION……..……. iii

ACKNOWLEDGMENTS……. iv

ABSTRACT…….…….. v

LIST OF TABLES…… vi

LIST OF FIGURES…….…… vii

CHAPTER I……..……. 1

CHAPTER II UNIVARIATE ANALYSES……... 8

Method and Materials…….…… 9

Microarray Database……..…… 9

Gene Expression Analyses…….. 11

Preliminary Gene Selection…….. 11

Univariate Cox Regression…….. 12

Influence of Estrogen Receptor and Progestin Receptor Status on Gene Expression Levels……..…… 14

Results…… 15

Univariate Cox Regression of PFS and OS without Regard to ER and PR Status……..…... 15

Univariate Cox Regression of PFS and OS According to ER Status……... 16

Univariate Cox Regression of PFS and OS According to PR Status……... 20
Univariate Cox Regression of PFS and OS According to ER/PR Status…..……..……... 23
Analyses of Clinical Relevance using Kaplan-Meier Plots…....…. 25
Evaluation of Influence of Steroid Hormone Receptor Status using Boxplots……..…… 27
Discussion……... 32
CHAPTER III MULTIVARIATE ANALYSES…….. 36
Method and Materials…….…….. 37
Multivariate Cox Regression……..……. 37
Least Absolute Shrinkage and Selection Operator…….. 38
Results…… 39
Multivariate Cox Regression with Interaction of PFS and OS without Regard to ER and PR Status of the Cancer Biopsy…… 39
Multivariate Cox Regression with Interaction of PFS and OS with Regard to ER and PR Status of the Cancer Biopsy……. 40
Multivariate Cox Regression for the Genes of Hormones of PFS and OS without Regard to ER and PR Status of the Cancer Biopsy…… 47
Multivariate Cox Regression for the Genes of

Receptors of PFS and OS without Regard to
ER and PR Status of the Cancer Biopsy……..….. 49
Kaplan-Meier Plots Discriminating Hormone-Receptor
Pairs Whose Expression Levels Predict Clinical Outcomes..... 53
Multivariable Gene Expression Model Fitted using LASSO…… 54
Discussion……... 57

CHAPTER IV META-ANALYSIS AND MOLECULAR SIGNATURES……. 60
Method and Materials……..……... 60
Public Databases……..……..……... 60
Molecular Signatures and Multivariable Cox Model……. 62
Results…… 64
Meta-analysis of HRL and Four Public Databases……..……. 65
Kaplan-Meier Plots Demonstrating Results from Meta-Analysis… 67
Gene Signatures Predicting Clinical Behavior of Breast
Carcinomas…… 67
Discussion……... 73
References…….. 76
Curriculum vitae…….…… 81

LIST OF TABLES

TABLE PAGE

1. Summary of Patient Characteristics in Microarray Database..... 10

2. Summary of Genes with Expression Levels Associated with PFS..... 18

3. Summary of Genes with Expression Levels Associated with OS..... 19

4. Summary of Genes with Expression Levels Correlating Either with PFS (A) or OS (B) for 146 ER+ Breast Carcinomas..... 21

5. Summary of Genes with Expression Levels Correlating either with PFS (A) or OS (B) for 101 ER- Breast Carcinomas..... 21

6. Summary of Genes with Expression Levels Associated either with PFS (A) or OS (B) for 151 PR+ Breast Carcinomas..... 22

7. Summary of Genes with Expression Levels Associated with PFS for 96 PR- Breast Carcinomas..... 23

8. Summary of Genes with Expression Levels Associated with OS for 118 ER+/PR+ Breast Carcinomas..... 28

9. Summary of Genes with Expression Levels Associated with PFS for 28 ER+/PR- Breast Carcinomas..... 28

10. Summary of Genes with Expression Levels Associated with PFS for 68 ER-/PR- Breast Carcinomas..... 28

11. Summary of the Interaction Term for Hormone and Receptor Gene Pairs with Expression Levels Associated with OS without Regard to Receptor Status……...…… 40
12. Summary of the Interaction Term for Hormone and Receptor Gene Pairs with Expression Levels Associated with OS for 146 ER+ Breast Carcinomas……...…. 41
13. Summary of the Interaction Term for Hormone and Receptor Gene Pairs with Expression Levels Associated with OS for 101 ER- Breast Carcinomas……...…… 42
14. Summary of the Interaction Term for Hormone and Receptor Gene Pairs with Expression Levels Associated with OS for 151 PR+ Breast Carcinomas…… 43
15. Summary of the Interaction Term for Hormone and Receptor Gene Pairs with Expression Levels Associated with OS for 96 PR- Breast Carcinomas…… 44
16. Summary of the Interaction Term for Hormone and Receptor Gene Pairs with Expression Levels Associated with PFS for 118 ER+/PR+ Breast Carcinomas…… 44
17. Summary of the Interaction Term for Hormone and Receptor Gene Pairs with Expression Levels Associated with OS for 28 ER+/PR- Breast Carcinomas…… 45
18. Summary of the Interaction Term for Hormone and Receptor Gene Pairs with Expression Levels Associated with either

PFS (A) or OS (B) for 33 ER-/PR+ Breast Carcinomas…….. 46

19. Summary of the Interaction Term for Hormone and Receptor Gene Pairs with Expression Levels Associated with OS for 68 ER-/PR- Breast Carcinomas…… 47

20. Summary of the Hormone Term from Two Variable Cox Models of Hormone and Receptor Gene Pairs with Expression Levels Associated with PFS……..…….. 48

21. Summary of the Hormone Term from Two Variable Cox Models of Hormone and Receptor Gene Pairs with Expression Levels Associated with OS……..…… 49

22. Representative Summary of the Receptor Term from Two Variable Cox Models of Hormone and Receptor Gene Pairs with Expression Levels Associated with PFS…….. 51

23. Representative Summary of the Receptor Term from Two Variable Cox Models of Hormone and Receptor Gene Pairs with Expression Levels Associated with OS……...… 52

24. Significant Genes and Corresponding Beta Coefficients for LASSO Performed on 142 Peptide/Protein Hormones and Their Cognate Receptors with and without Receptor Status of Breast Carcinomas…….....…… 57

25. Summary of the Genes from Meta-analysis without Regard to Receptor Status (n=1126) with Expression Levels Associated with PFS/DMFS…….....…….. 65

26. Summary of the Meta-analysis for (A) ER+ (n=737) and
(B) ER- (n=378) Breast Cancers with Expression Levels
Associated with PFS/DMFS…..... 66

LIST OF FIGURES

FIGURE PAGE

1. Flow Diagram of Approaches and Analyses used to Decipher Clinically Relevant Molecular Signatures…….……...…….. 7

2. Representative Kaplan-Meier Plots Comparing Above and Below the Median Relative Gene Expression (1)……..……...…. 29

3. Representative Kaplan-Meier Plots Comparing Above and Below the Median Relative Gene Expression (2)…….……. 30

4. Representative Boxplots Comparing Distributions of Relative Gene Expressions by Receptor Status……..……... 31

5. Representative Kaplan-Meier Plots Comparing Above and Below the Median Relative Gene Expression for Peptide/Protein Hormones and Their Cognate Receptors…….……...…. 56

6. Representative Kaplan-Meier Plots Comparing Above and Below the Median Relative Gene Expression of 142 Peptide/Protein Hormones or Their Cognate Receptors with and without Regard to ER Status for PFS from the Meta-analysis……..……. 70

7. Composition of Molecular Signatures Derived Showing Genes Common to Each…….…….. 71

8. Summary of 1000 C-indices of Each of the Three Molecular Signatures on Each of the Four Public Databases, (A) VDX,

(B) TRANSBIG, (C) UNT and (D) MAINZ……... 72

CHAPTER I

INTRODUCTION

The Surveillance, Epidemiology, and End Results (SEER) program estimated 231,840 new breast cancer cases and 40,290 deaths in the United States in 2015 ("Surveilance, Epidemiology, and End Results Program,"). Self-examination to clinical examination to mammogram is a typical route to detect breast cancer. When an area of the breast is identified as suspicious, a biopsy may be taken to help classify a tumor as malignant. If the tumor is diagnosed as carcinoma, then the pathologist will evaluate the lesion by stage and type. Stage is based on several factors such as size of the cancer, number of lymph nodes involved and signs of local invasion of the breast and metastasis to other organs (Greenspan, Gardner, & Shoback, 1997). Some types of breast cancer may be characterized as *in situ* versus invasive, ductal versus lobular and sex-hormone receptor status (Fleisher, Dnistrian, Sturgeon, Lamerz, & Wittliff, 2002; J. Wittliff, Pasic, & Bland, 1998). Classification of stage and grade as well as nodal status contributes significantly to determining the prognosis and treatment of the disease.

An important consideration for assessing breast cancer prognosis and treatment was provided by the discovery that estrogen (ER) and progestin

receptor (PR) proteins were clinically useful biomarkers (Fisher et al., 1983; Fisher et al., 1981; Fleisher et al., 2002; Hammond et al., 2010; James, 1984; J. Wittliff et al., 1998). Briefly, the presence of significant levels of ER and PR in a breast cancer biopsy was correlated with better prognosis than patients with lesion lacking ER and PR proteins. In addition, the presence of ER and PR in a breast cancer tissue biopsy was strongly associated with the patient's response to Tamoxifen, an antiestrogen-like drug that binds to ER (Fisher et al., 1983; Fisher et al., 1981). Receptor status is now used as a combination of ER and PR with the addition of epidermal growth factor receptor-2 (HER-2/neu) protein, which is involved in growth regulation of cancer cells given their presence in the cell (Fleisher et al., 2002; Hammond et al., 2010).

Common treatments for breast cancers with elevated levels of ER and PR are the antiestrogen-like drugs (e.g., Tamoxifen, Evista/Raloxifene and Fareston/Toremifene), which are termed SERMs (Selective Estrogen Receptor Modulators) ("Raloxifene Hydrochloride," ; "Tamoxifen Citrate," ; "Toremifene,"). Aromatase Inhibitors such as Arimidex (Anastrozole) and Femara (Letrozole), which block the production of estrogens from androgenic precursors produced by the adrenal glands, are used as hormone therapies for post-menopausal patients (Hong & Chen, 2011). HER/2 oncoprotein serves as a biomarker for treating a patient with the drug Trastuzumab (Herceptin) which attaches itself to the HER-2/neu protein that is present in the surface membranes of certain breast carcinoma cells. For decades the Hormone Receptor Laboratory, which holds both CLIA and Commonwealth of Kentucky licenses as a Clinical Laboratory, has determined the

levels of these biomarkers in thousands of breast cancer tissue biopsies for management of patients.

One area of research that is likely to improve the survival rates of women diagnosed with breast cancer is personalized or precision medicine. Due to the clinical and molecular heterogeneity of breast cancer, identifying genes and gene products involved in driving the progression of the disease may provide opportunities to design and synthesize new drugs for these new molecular targets. In addition, knowledge of their expression in relationship to risk of breast cancer recurrence provides additional information regarding a patient's prognosis. Thus research in these areas of genomics and proteomics hold promise for developing a larger arsenal of personalized treatment options for breast cancer patients. Expression patterns of sets of genes that accurately predict the clinical behavior of cancers are called molecular signatures. Many genomics based signatures have been shown effective in predicting clinical outcomes such as progression free survival, which measures the length of time a patient lives without the appearance of a metastasis of the disease. (Gingras, Desmedt, Ignatiadis, & Sotiriou, 2015; J. Wittliff et al., 2002). A number of gene expression profiles (i.e., molecular signatures) have been developed and commercialized.

It is widely documented that numerous peptide/protein hormones such as insulin act as growth factors impacting carcinoma cell growth and may play a role in carbohydrate metabolism during differentiation and growth of the lesion (Chen et al., 2002; Falzon & Du, 2000). Each of the hydrophilic peptide/protein hormones circulate freely in the blood and bring about their particular physiological actions in

normal target cells by associating with high affinity with their cognate receptor proteins which are located on cell surfaces (Norman & Litwack, 1997; Pierce, 1982). Thus this diverse family of hormones which influences a wide variety of normal cellular pathways in the many organs composing the endocrine system when bound to their cognate receptor proteins provides a fertile and unchartered area to explore in breast cancer. The overarching goal of this study is to determine the relationships of the expression of the genes for each of the peptide/protein hormones and that of their cognate receptors with clinical outcomes of breast cancer patients. Our hypothesis is that expression profiles of subsets of these genes may be used to predict risk of breast carcinoma recurrence in patients with primary disease.

To the best of our knowledge, no study has analyzed collectively the gene expression patterns of all of the peptide/protein hormones and their cognate receptors in relationship to their association with breast cancer behavior. The association of a peptide/protein hormone with its cognate receptor results in the release of signaling molecules (second messengers) inside the cell to trigger a variety of physiological changes ("Journal of receptor and signal transduction research," 1995).

Peptide/protein hormones are produced and secreted by organs of the endocrine system throughout our bodies affecting adjacent cells (paracrine action) and cells located in distant organs (endocrine action). For example, insulin and glucagon are produced by the pancreas and secreted into the bloodstream where they influence a variety of other organs to control blood sugar levels. Surprisingly,

breast carcinoma cells have been found to overexpress some of these hormone and receptor genes (Chen et al., 2002; Falzon & Du, 2000). From a variety of previous reports addressing relationships of hormones and cancer, and my co-mentor's years of experience investigating endocrine mechanisms of breast cancer, the following questions were developed. Do certain breast carcinomas express elevated levels of mRNA for genes of peptide/protein hormones and their cognate receptors? Are the gene expression levels related to clinical outcomes of the patients? Can one discern gene expression profiles (i.e., molecular signatures) that may be useful clinically in predicting risk of breast cancer recurrence?

To begin to answer these questions we took a step-wise global approach using the gene expression levels of 22,000 genes that had been determined by microarray of RNA isolated from LCM-procured breast carcinoma cells (Figure 1 Flow Diagram). Determination of which peptide/protein hormones and their cognate receptors may be playing a role in the clinical behavior of breast cancers required surveying research literature describing the diverse family of endocrine regulators as will be described in Chapter II. Due to the complicated nature of analyzing multiple genes and various breast cancer subtypes, the statistical analyses were performed in a stepwise manner as outlined in the flow diagram in Figure 1. Each of the following Chapters (II, III and IV) describes the manner in which the investigations were conducted, the results in the comprehensive databases that were employed and the statistical methods used for the analyses to identify interrelationships. These relationships are the essence of our hypothesis that expression profiles of subsets of genes for peptide/protein hormones and their

cognate receptors may be used to predict risk of breast carcinoma recurrence in patients with primary disease.

In Chapter II, we evaluate univariate association of each of the identified hormone and receptors with breast cancer progression and overall survival. The univariate analyses in Chapter II will be extended to associations by ER/PR status. In a similar fashion ER/PR will be evaluated in Chapters III and IV. Multivariate analyses performed on hormone and receptor gene pairs in Chapter III explore an exhaustive evaluation of the relationships between hormones and their cognate receptors. Also in Chapter III, regularization techniques are employed to analyses all 142 genes at one time. In Chapter IV, a meta-analysis was performed with public databases to compare our results with a few highly referenced databases. Molecular signatures close out our analyses in Chapter IV as described in the Flow Diagram (Figure 1). The Flow Diagram is a recipe of approaches to develop molecular signatures of clinical relevance in the future for the Hormone Receptor Laboratory.

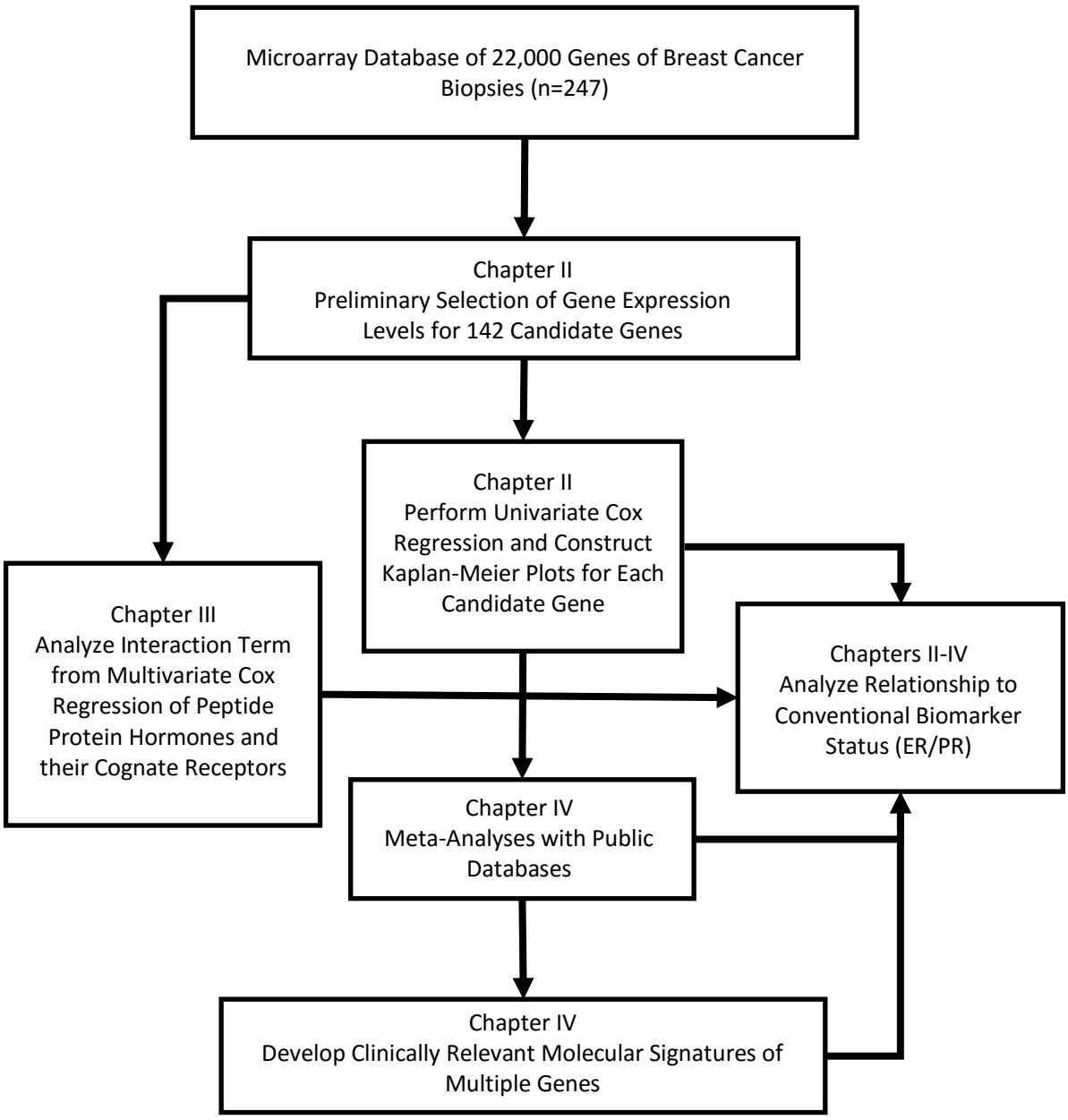

Figure 1. Flow Diagram of Approaches and Analyses used to Decipher Clinically Relevant Molecular Signatures

CHAPTER II

UNIVARIATE ANALYSES

The microarray, protein biomarker and clinical follow-up databases used for analyses described in these investigations were established by my Co-mentor, Dr. Wittliff and collaborators almost two decades ago. (Sarah A Andres et al., 2015; Andres, Brock, & Wittliff, 2013; Andres, Edwards, & Wittliff, 2012; Andres & Wittliff, 2011, 2012; Kerr II & Wittliff, 2011; Kidd et al., 2010; Kruer, Cummins, Powell, & Wittliff, 2013; Ma et al., 2003; Tecimer et al., 2000; J. Wittliff et al., 1998; J. L. Wittliff, 2010) Briefly, intact frozen tissue sections of de-identified human breast carcinoma tissue biopsies as well as those that were processed for a sophisticated technique called Laser Capture Microdissection (LCM) for the microarray analyses allowed collection of specific cell types in a non-destructive manner. (Andres & Wittliff, 2011, 2012) The results collected from microarray of LCM-procured cells are truly unique in that only the expressions of carcinoma cell genes were determined. This database was complemented by results in other comprehensive databases that contained quantitative results of protein biomarker levels (ER, PR and HER-2) that are used routinely in clinical management of breast cancer, qPCR validated expression levels of almost 100 genes as well as extensive results on the features of the carcinoma biopsy and numerous patient parameters including clinical outcomes.

Methods and Materials

Microarray Database

Using an IRB-approved biorepository and associated databases composed of tissue specimens previously collected by Dr. Wittliff's laboratory (CLIA and Commonwealth of Kentucky licensed) at the University of Louisville for clinical assays of estrogen (ER) and progestin receptors (PR), de-identified tissue specimens of primary breast cancers obtained from 1988 - 1996 were examined using REMARK criteria (McShane et al., 2006) as described in previous studies. (Sarah A Andres et al., 2015; Andres et al., 2013; Andres & Wittliff, 2011; Kerr II & Wittliff, 2011; Kruer et al., 2013) Patients were treated with the standard of care at the time of diagnosis. Tissue-based properties (e.g., pathology, grade, size, and tumor marker expression) and patient-related characteristics (e.g., age, race, smoking status, menopausal status, stage, and nodal status) were utilized to determine relationships between gene expression and clinical parameters. Microarray analyses were performed on LCM procured carcinoma cells from 247 breast cancer tissue biopsies as described. (Ma et al., 2003; J. L. Wittliff, 2010) Figure 2 describes the characteristics of the patient analyzed in the microarray database.

Patient Characteristics	n
Median Age (range)	
59 years (21-89.5)	247
Median Observation Time (range)	
65 months (3-155)	247
Race	
White	211
Black	34
Other	2
Histology	
Invasive ductal carcinoma	193
Lobular carcinoma	15
Medullary carcinoma	8
Other/Unknown	31
Median Tumor Size (range)	
27 mm (3-100)	218
Stage	
0	3
1	60
2A	85
2B	55
3A	22
3B	13
4	4
Unknown	5
Grade	
1	14
2	70
3	94
4	1
Unknown	68
Estrogen Receptor Status	
Negative	98
Positive	151
Lymph Node Status	
Negative	125
Positive	102
Unknown	20
Recurrence Status	
Yes	98
No	146
Never disease-free	3

Table 1. Summary of Patient Characteristics in Microarray Database

Gene Expression Analyses

Levels of mRNA expression were analyzed by microarray and qPCR according to the protocols described in publications of previous investigators with Dr. Wittliff after isolation and the quality of RNA was evaluated with Agilent RNA 6000 Nano Kits and the Bioanalyzer™ Instrument (Agilent Technologies, Palo Alto, CA) (Sarah A Andres et al., 2015; Andres et al., 2013; Andres & Wittliff, 2011, 2012; Kerr II & Wittliff, 2011). Relative gene expression levels were determined from qPCR with the ΔΔCt method using ACTB for normalization and Universal Human Reference RNA (Stratagene, La Jolla, CA) as the calibrator.

Preliminary Gene Selection

An exhaustive inquiry of prevalent literature revealed 63 peptide/protein hormones and 82 cognate receptor proteins as shown in the second box of Figure 1 flow diagram. (Greenspan et al., 1997; Norman & Litwack, 1997) There are a number of peptide/protein hormones that have been reported to associate with multiple receptor proteins. For example, somatostatin (SST) is known to bind with five receptor isoforms – somatostatin receptor 1 (SSTR1), somatostatin receptor 2 (SSTR2), somatostatin receptor 3 (SSTR3), somatostatin receptor 4 (SSTR4) and somatostatin receptor 5 (SSTR5). (Hoyer et al., 1995) Similarly, receptors may pair with more than a single peptide/protein hormone. For instance, Spier and de Lecea (2000) demonstrated that another hormone, cortistatin (CORT), was also recognized by each of the five somatostatin receptors described above. Among the extensive repository of expression levels for ~22,000 genes in our microarray results of LCM-procured carcinoma cells, those genes for 61 hormones and 81

receptors were identified as candidates for analyses. The 247 tissue biopsies from patients with primary invasive ductal carcinoma served as the principal study population.

Univariate Cox Regression

Univariate Cox regression was performed on each of the 142 gene candidates using their relative expression levels as described in the third step of the flow diagram (Figure 1). Both Progression Free Survival (PFS) and Over-All Survival (OS) were analyzed as clinical outcomes for each gene. Univariate Cox models employed relative gene expression values as a single covariate (i.e., SST) and investigated the extent to which expression levels of a single gene in the cohort predicted the risk of recurrence of breast cancer (PFS) or succumbing to that disease (OS).

The general formula for the hazard function of a Cox proportional hazard model is as follows: $h(t) = h_0(t)\exp(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p)$. (Bradburn, Clark, Love, & Altman, 2003) The hazard function ($h(t)$) is time dependent, which changes over time along with the baseline hazard ($h_0(t)$). The baseline hazard measures the risk when all covariates ($x_1, x_2 \ldots x_p$) are equal to 0. Although the baseline hazard may change over time, one of the assumptions for Cox models is that the hazard for different subsets of the population will be proportional to the baseline hazard at all times. The beta coefficients ($\beta_1, \beta_2 \ldots \beta_p$), which link the covariates to the hazard at time t, are determined by maximizing the partial likelihood function associated with the model. Covariates are variables such as

relative gene expression values for each of the hormones or their receptors. Univariate Cox regressions have a single beta coefficient calculated for a single covariate as shown here: $h(t) = h_0(t)\exp(\beta_1 x_1)$.

Hazard ratios (HR) were derived from univariate Cox regression models by the exponentiation of β_1 and calculated for each of the 142 candidate genes. A hazard ratio for these models compares hazards (or risks for the clinical outcome being measured) at two different levels of gene expression. (Klein & Moeschberger, 2003) Since relative gene expression values are continuous covariates, the HR compares the hazard at any relative gene expression value to the hazard at a one unit level increase in relative gene expression.

$$HR = \frac{h_0(t)\exp(\beta_1(x_1 + 1))}{h_0(t)\exp(\beta_1 x_1)} = \exp(\beta_1)$$

An HR of greater than one represents an increase in risk, whereas a value of less than one represents a decrease in risk. An HR equal to one represents no difference in risk.

Each of the statistical computations was performed with R version 3.2.3. The commands *coxph* and *cox.zph* in the R package *survival* (Therneau, 2013) were used to calculate all univariate and multivariate Cox regressions and to validate the assumption of proportionality in significant genes. Cox regression p-values were adjusted for multiple comparisons using the Benjamini & Hochberg method (Benjamini & Hochberg, 1995) with 0.30 “discovery” cutoff. Relative gene expression levels determined with LCM- procured carcinoma cells from 247 patients were divided into groups expressing values that were above and below

the median for each gene candidate. These two groups of results were analyzed in Kaplan-Meier (KM) plots using the commands *plot* and *survfit*. The *survdiff* command performed a log-rank test comparing survival times between the two groups.

Influence of Estrogen Receptor (ER) and Progestin Receptor (PR) Status on Gene Expression Levels

Measurements of the protein biomarkers, ER and PR, have significant importance in predicting clinical outcomes of breast cancer patients such as risk of recurrence and over-all survival. (Fisher et al., 1987; Fisher et al., 1983; Fisher et al., 1981; Fleisher et al., 2002; Hammond et al., 2010; J. Wittliff et al., 1998) For example, patients with breast cancers exhibiting both ER and PR are reported to have a better prognosis and are candidates for anti-hormone therapy compared to those with ER and PR negative breast cancers. (Fisher et al., 1987; Fisher et al., 2001; Fisher et al., 1981) This is thought to be due in part, to the observation that ER, when complexed with its native ligand, estradiol-17 B or an estrogen mimic (e.g., Tamoxifen), stimulates the production of PR. (Cormier, Wolf, & Jordan, 1989) Comprehensive clinical trials of breast cancer patients treated with Tamoxifen, such as those of the NSABP support this conclusion. (Fisher et al., 1987; Fisher et al., 1983; Fisher et al., 1981)

Patients were categorized into the breast cancer subsets based on the ER and PR protein status of the tissue biopsy. Patients were also stratified according to carcinoma cells exhibiting one of the four combinations of the two protein biomarkers, either ER+/PR+, ER+/PR-, ER-/PR+ or ER-/PR-. Univariate Cox

regression was first performed for patients with cancers stratified by each protein status, ER+, ER-, PR+ and PR-. Then patients were evaluated according to one of the four possible combinations of ER and PR (e.g., ER+/PR+, ER+/PR-, ER-/PR+ or ER-/PR-) exhibited by their carcinoma cells with each of the 142 gene candidates. Kaplan-Meier plots were constructed for each gene using the various subsets of ER and PR. Box plots were constructed according to receptor status for each of the 142 candidate genes using the R command *boxplot*. The boxplots displayed the relative gene expression in the LCM-procured carcinoma cells for patients bifurcated by ER and PR status. The null hypotheses that the two groups come from identical populations was tested using the R command *wilcox.test*. The R command employed an unpaired independent two-sample Mann-Whitley-Wilcoxon test by comparing ER+ to ER- groups and PR+ to PR- groups of breast cancers.

Results

Univariate Cox Regression of PFS and OS without Regard to ER and PR Status

From univariate Cox regression for PFS, expression levels of fifteen genes for peptide/protein hormones and nineteen receptors showed significance at the adjusted p-value of <0.30 (Table 2). Noteworthy among the findings, the expression of POMC, whose mRNA transcript is quite large containing the sequences for a number of hormones, exhibited the highest statistical significance. (Lee et al., 2006) POMC had a HR of 1.72, which can be interpreted as patients with breast carcinomas expressing this gene at one unit higher have a 72% increase in risk of recurrence of metastatic breast cancer. The results obtained for

SST and SSTR1 are of particular interest because they are a hormone-cognate receptor pair, which appeared among the top five genes for univariate Cox regression for PFS.

From univariate Cox regression for OS, expression levels of seven genes for peptide/protein hormones and ten receptor genes showed significance at the adjusted p-value of <0.30 (Table 3). Interestingly, expression of POMC gene related to breast cancer OS mirrored the highest statistical significance observed when PFS was calculated. All seventeen significant genes related to OS exhibited significant relationships with PFS. Also, the hormone-receptor pair POMC and MC5R exhibited statistical significance individually for univariate Cox regression for OS.

Univariate Cox Regression of PFS and OS According to ER Status

Using gene expression levels from 146 ER+ breast cancer carcinomas, univariate Cox regression determined that two peptide/protein hormones, SST and renin (REN), exhibited significant mRNA expression levels related to PFS and one receptor (SSTR2) was related to OS (Table 4). SST reappeared as significant for predicting PFS as it did for predicting PFS without regards to hormone receptor status. SSTR2, a cognate receptor for SST, exhibited significance for OS and will reappear in our molecular signature for ER+ breast cancers. Also showing significance for PFS is REN, which is normally produced in the kidneys when intrarenal blood pressure drops. The inhibition of REN has been demonstrated to mitigate angiogenesis, the production of new blood vessels. (Rodrigues-Ferreira & Nahmias, 2015). Cancer cells need more blood supply to provide them with the

nutrients to grow faster. The HR of REN is 2.93 meaning that cancers with higher levels of expression of REN are more likely to have a shorter time to progression.

Gene Symbol	β	HR	95% CI (HR)	adj. p
POMC	0.54	1.72	(1.37,2.15)	<0.001
CALCR	0.44	1.56	(1.21,2.01)	0.03
SST	0.55	1.73	(1.27,2.35)	0.03
IGF1R.clone	-0.17	0.84	(0.76,0.94)	0.05
SSTR1	0.65	1.91	(1.28,2.84)	0.05
TMSB15A	0.14	1.16	(1.05,1.27)	0.07
PPY	0.93	2.55	(1.32,4.92)	0.10
CRH	0.55	1.73	(1.18,2.54)	0.10
SSTR3	0.97	2.64	(1.28,5.46)	0.13
REN	0.61	1.85	(1.16,2.93)	0.13
TMSB10	0.4	1.49	(1.1,2.02)	0.13
ACVR2B	0.35	1.42	(1.08,1.86)	0.14
ACVR2A	0.29	1.34	(1.06,1.7)	0.16
PRL	0.32	1.38	(1.06,1.79)	0.16
VIPR1	-0.26	0.77	(0.62,0.96)	0.18
NPY1R	-0.16	0.85	(0.74,0.98)	0.21
SCT	0.58	1.79	(1.08,2.97)	0.21
NPY6R	-0.99	0.37	(0.15,0.9)	0.22
INSR.AL365454	-0.25	0.78	(0.61,0.99)	0.23
ACVR1	-0.27	0.76	(0.58,1)	0.23
ACVR1B	0.25	1.29	(1.01,1.65)	0.23
RLN1	-0.21	0.81	(0.66,0.99)	0.23
RXFP3	1.2	3.32	(1.07,10.27)	0.23
MC5R	0.6	1.83	(1.02,3.29)	0.23
GHR	0.4	1.5	(1.02,2.2)	0.23
AVPR2	0.75	2.11	(1.04,4.28)	0.23
PTH	-1.28	0.28	(0.08,0.98)	0.23
EDN1	0.37	1.45	(1.03,2.05)	0.23
GHSR	-1.07	0.34	(0.13,0.94)	0.23
RLN2	-0.13	0.88	(0.77,1)	0.23
AGTRAP	-0.31	0.73	(0.53,1.01)	0.27
VIPR2	-0.1	0.91	(0.82,1)	0.27
AVPR1A	0.88	2.41	(0.95,6.11)	0.28
GAL	0.11	1.12	(0.99,1.26)	0.30

Table 2. Summary of Genes with Expression Levels Associated with PFS. As described in Methods and Materials, expression levels of the 142 candidate genes were determined using LCM-procured cells from 247 breast cancer biopsies.

Gene Symbol	β	HR	95% CI (HR)	adj. p
POMC	0.52	1.68	(1.34,2.1)	<0.001
CALCR	0.45	1.57	(1.19,2.08)	0.08
TMSB15A	0.18	1.19	(1.07,1.33)	0.08
IGF1R.clone	-0.17	0.85	(0.75,0.96)	0.16
SSTR3	1.1	3.01	(1.34,6.76)	0.16
SCT	0.79	2.21	(1.27,3.83)	0.16
NPY6R	-1.36	0.26	(0.09,0.71)	0.16
TMSB10	0.45	1.57	(1.12,2.2)	0.16
ACVR2A	0.35	1.42	(1.08,1.86)	0.19
PPY	0.87	2.38	(1.15,4.93)	0.26
PTH	-1.77	0.17	(0.04,0.77)	0.26
NPY1R	-0.2	0.82	(0.69,0.97)	0.26
MC5R	0.72	2.06	(1.07,3.97)	0.28
VIPR1	-0.27	0.76	(0.59,0.98)	0.28
VIPR2	-0.13	0.88	(0.78,0.99)	0.28
EDN1	0.44	1.56	(1.05,2.3)	0.28
GHSR	-1.34	0.26	(0.08,0.87)	0.28

Table 3. Summary of Genes with Expression Levels Associated with OS. As described in Methods and Materials, expression levels of the 142 candidate genes were determined using LCM-procured cells from 247 breast cancer biopsies.

Using gene expression levels from 101 ER- breast cancer carcinomas, univariate Cox regression determined that mRNA levels of four genes for peptide/protein hormones and two receptor genes were correlated with a significant for predicting PFS and four genes for peptide/protein hormones predicted OS (Table 5). The gene expression levels of POMC were significant for predicting PFS and OS for ER- cancers similarly to our observation for predicting PFS and OS without regard to receptor status. CALCR, POMC, GH1 and PRL were significant for PFS and will reappear in our molecular signature for ER- cancers. Growth hormone (GH1) is typically produced in the pituitary gland

regulated by GHRH and SST from the hypothalamus. Models with SST and GHRH could be performed in future studies to see if the cancer cells may be controlling levels of GH1 by autocrine means.

Univariate Cox Regression of PFS and OS according to PR status

Using gene expression levels from 151 PR+ breast cancer carcinomas, univariate Cox regression determined that one peptide/protein hormone, thymosin β10 (TMSB10), exhibited significance for predicting PFS and three peptide/protein hormones, CALCA, TMSB10 and POMC, were significant for predicting OS (Table 6). TMSB10 was first discovered in the thymus but is made throughout the body. TMSB10 has been shown to be overexpressed in lung and pancreatic cancer and even targeted for therapy. (Alldinger et al., 2005; Langevin, Kratzke, & Kelsey, 2015) Are the mechanisms that cause TMSB10 to be overexpressed in lung and pancreatic cancers somehow related to breast cancers that exhibit high levels of PR but not necessarily ER? Interestingly, POMC is significant for OS but not PFS in a sex hormone positive (PR+) subtype. Could early deaths from patients who are ER- but PR+ be driving the significance for POMC?

A. PFS ER+

Gene Symbol	β	HR	95% CI (HR)	adj. p
SST	0.73	2.08	(1.28,3.39)	0.23
REN	1.08	2.93	(1.47,5.85)	0.23

B. OS ER+

Gene Symbol	β	HR	95% CI (HR)	adj. p
SSTR2	0.28	1.32	(1.11,1.57)	0.27

Table 4. Summary of Genes with Expression Levels Correlating either with PFS (A) or OS (B) for 146 ER+ Breast Carcinomas. As described in Methods and Materials, expression levels of 142 candidate genes were determined using LCM-procured cells from 247 breast cancer biopsies.

A. PFS ER-

Gene Symbol	β	HR	95% CI (HR)	adj. p
CALCR	0.43	1.53	(1.15,2.04)	0.16
POMC	0.4	1.49	(1.17,1.9)	0.16
GH1	-0.76	0.47	(0.28,0.79)	0.16
PRL	0.58	1.79	(1.18,2.71)	0.16
AVPR1A	1.83	6.23	(1.81,21.4)	0.16
SCT	0.8	2.24	(1.19,4.2)	0.29

B. OS ER-

Gene Symbol	β	HR	95% CI (HR)	adj. p
POMC	0.37	1.45	(1.13,1.87)	0.28
GH1	-0.84	0.43	(0.24,0.79)	0.28
PRL	0.54	1.72	(1.15,2.56)	0.28
SCT	0.94	2.55	(1.33,4.88)	0.28

Table 5. Summary of Genes with Expression Levels Correlating either with PFS (A) or OS (B) for 101 ER- Breast Carcinomas. As described in Methods and Materials, expression levels of 142 candidate genes were determined using LCM-procured cells from 247 breast cancer biopsies.

A. PFS PR+

Gene Symbol	β	HR	95% CI (HR)	adj. p
TMSB10	0.97	2.63	(1.56,4.44)	0.04

B. OS PR+

Gene Symbol	β	HR	95% CI (HR)	adj. p
CALCA	0.38	1.46	(1.14,1.85)	0.23
TMSB10	0.87	2.39	(1.34,4.27)	0.23
POMC	0.86	2.36	(1.29,4.34)	0.27

Table 6. Summary of Genes with Expression Levels Associated either with PFS (A) or OS (B) for 151 PR+ Breast Carcinomas. As described in Methods and Materials, expression levels of 142 candidate genes were determined using LCM-procured cells from 247 breast cancer biopsies.

Using gene expression levels from 96 PR- breast cancer carcinomas, univariate Cox regression determined that five genes for peptide/protein hormones and five receptor genes were significant for predicting PFS. No genes were found to be significant after adjusting p-values for multiple comparisons (Table 7). CALCR, POMC and PRL are shown to have an association with PFS for PR- breast cancer carcinomas and were also associated with ER- breast cancer carcinomas. These three genes will be common to two molecular signatures. Since 1995, researchers have known about the ability of breast cancer cells to secrete active forms prolactin (PRL) possibly through autocrine mechanisms, although paracrine mechanisms could not be ruled out. (Ginsburg & Vonderhaar, 1995) We will look at the relationship between PRL and its cognate receptor PRLR in Chapter Three but further study could be performed on the statistical interactions of PRL with other hormones and receptors.

PFS PR-

Gene Symbol	β	HR	95% CI (HR)	adj. p
CALCR	0.56	1.75	(1.3,2.37)	0.04
POMC	0.43	1.54	(1.18,2.01)	0.11
SST	0.71	2.03	(1.26,3.26)	0.14
SSTR1	0.67	1.96	(1.23,3.13)	0.14
CRH	0.62	1.86	(1.21,2.87)	0.14
AVPR2	1.18	3.25	(1.39,7.57)	0.15
AVPR1A	1.74	5.71	(1.51,21.57)	0.20
ACVR1	-0.4	0.67	(0.48,0.93)	0.30
PRL	0.42	1.52	(1.07,2.16)	0.30
PTH	-2.31	0.1	(0.01,0.71)	0.30

Table 7. Summary of Genes with Expression Levels Associated with PFS for 96 PR- Breast Carcinomas. As described in Methods and Materials, expression levels of 142 candidate genes were determined using LCM-procured cells from 247 breast cancer biopsies.

Univariate Cox Regression of PFS and OS according to ER/PR status

Since both the status of ER and PR are considered in the clinical setting when assessing a patient's risk of recurrence and selection of a therapeutic agent, these studies utilized the results of clinical determinations of the steroid hormone receptors. (Fisher et al., 1987; Fisher et al., 1983; Fisher et al., 1981; Fleisher et al., 2002; Hammond et al., 2010; J. Wittliff et al., 1998) The analyses described include the four combinations ER+/PR+, ER+/PR-, ER-/PR+ and ER-/PR- status of each of the breast cancer surgical biopsies submitted for ER and PR protein analyses. In general, a patient with a breast carcinoma determined to be ER+/PR+ exhibits the best prognosis compared to the patients with cancers exhibiting the other combinations. Best prognosis implies that the patient is expected to have a lower risk of breast cancer recurrence compared with patients with cancer biopsies exhibiting the other three combinations of ER and PR. A patient with ER-/PR-

breast carcinoma exhibits the worst prognosis compared to that the other patients. In addition, patients with ER-/PR- tumors are not candidates for hormone-receptor protein based therapies such as tamoxifen. (Fisher et al., 1987; Hammond et al., 2010; Kerr II & Wittliff, 2011)

Since our goal is to decipher molecular signature for predicting breast cancer outcomes, we undertook the following analyses. Using gene expression levels from 118 ER+/PR+ breast cancer carcinomas, univariate Cox regression determined that the receptor SSTR2 exhibited significance for predicting OS (Table 8). Using gene expression levels determined from LCM-procured of 28 ER+/PR- breast cancer carcinomas, univariate Cox regression determined that one peptide/protein hormone, EPO, and two genes of receptors, CALCR and ACVR1B, had significance for predicting PFS (Table 9). However since this sample group only contained 28 patients, caution must be taken in the interpretation of these analyses. We noted that erythropoietin, EPO had not appeared in any of the gene subsets detected thus far. Furthermore CALCR has been significant for predicting PFS of ER+/PR-, PR-, ER- breast cancer patients and for predicting PFS and OS without regard to receptor status.

The expression of ER-/PR+ protein biomarkers in a breast cancer is infrequently observed and since the action of ER is known to provoke the production of PR, its appearance is the focus of research. (Andres & Wittliff, 2012; Fleisher et al., 2002; J. Wittliff et al., 1998) Unfortunately, using gene expression levels from 33 ER-/PR+ breast cancer carcinomas, univariate Cox regression determined no genes of significance after adjusting p-values for multiple

comparisons. This may be due to a small sample size. A few significant findings from unadjusted p-values could have been novel in their implications.

Breast carcinoma biopsies that lack both ER and PR protein (ER-/PR-) are widely known to correlate with poor prognosis of breast cancer patients and their lack of response to anti-hormone therapies. (Andres & Wittliff, 2012; Fleisher et al., 2002; J. Wittliff et al., 1998) Using gene expression levels determined by microarray from LCM-procured carcinoma cells of 68 ER-/PR- breast cancer carcinomas, univariate Cox regression determined mRNA expression levels of two peptide/protein hormones, POMC and PRL, significant for predicting PFS (Table 10). The genes for PRL and POMC appear later in our molecular signature for ER- breast carcinomas.

Analyses of Clinical Relevance using Kaplan-Meier Plots

Due to the numerous genes exhibiting significance for predicting PFS and OS of breast cancer for each category of receptor status, a representative sample of Kaplan-Meier plots has been presented. (Rich et al., 2010) Without regards to ER/PR status, IGF1R.clone and GAL genes showed significant differentiation for predicting PFS and OS between cancers that expressed the gene above the median versus compared to those that expressed the gene below the median. The median is taken as a first discriminator in this thesis, however later analyses for the development of manuscripts will also derive outcomes as a function of quartiles. TMSB10 expression was significant for predicting PFS while that of activin A receptor type IIA, ACVR2A, was significant for predicting OS. The risk of recurrence and survivorship was worse for patients whose tumors expressed

IGF1R.clone mRNA below the median and GAL, TMSB10 and ACVR2A mRNA expression levels that were above the median.

For ER+ cancers, SST gene expression showed a significant differentiation for predicting PFS and a poor prognosis for patients whose lesions expressed the gene above the median. SSTR2 was significant for predicting OS with a poor prognosis for patients whose tumors expressed the gene above the median. For ER- carcinomas, expression of growth hormone, GH1, predicted both PFS and OS that was highly statistically significant with a poor prognosis for those whose breast cancers expressed GH1 below the median.

For PR+ cancers, glycoprotein hormone alpha polypeptide, CGA, showed a significant differentiation for predicting PFS while that of endothelin, EDN1, showed significance for predicting OS using Kaplan-Meier plots. Patients expressing CGA and EDN1 above the median were correlated with a worse prognosis compared to patients expressing these genes at levels below the median. For PR- cancers, arginine vasopressin receptor 2, AVPR2, significantly differentiated patients above and below the median for PFS and GH1 significantly differentiated patients above and below the median for OS. Patients with breast cancers expressing CGA above the median and patients with breast cancers expressing GH1 below the median had a worse prognosis.

Examination of ER+/PR+ carcinomas revealed that CALCA significantly differentiated patient outcome for PFS while SSTR2 showed significance for predicting OS. Patients expressing CALCA and SSTR2 above the median were correlated with a worse prognosis than those with gene expression levels below

the median. The most significant findings from the Kaplan-Meier plots were revealed in carcinoma cells classified as ER-/PR-, the biomarker status that is related to the worse prognosis among the four combinations. When ER-/PR- cancers were examined, GH1 expression was highly significant for predicting PFS and OS with those whose lesions expressed GH1 below the median having the highest risk of recurrence and survivorship.

Evaluation of Influence of Steroid Hormone Receptor Status using Boxplots

One of the goals of this study was to determine if there were candidate genes whose expression was related to the expression levels of either ER or PR protein. Such relationships would suggest that steroid hormone action may be related to the expression of certain peptide/protein hormone genes and the genes for their receptors. Due to the magnitude of the analyses and results for the 142 candidate genes only representative boxplots that showed the greatest statistical significance are displayed. Considering influence of ER, the relative gene expression of receptor activity modifying protein 2 (RAMP2), IGF1R, IGF1R.clone, angiotensin II receptor, type 1 (AGTR1) and thyrotropin-releasing hormone (TRH) were significantly higher levels in ER+ carcinoma cells compared to their ER- counterparts. The relative gene expression of ACVR2A and GAL in ER- cancers was significantly higher than observed in ER+ cancer cells. The relative gene expression of AGTR1 and IGF1R.clone in PR+ breast cancers was expressed at a significantly higher level than their expression levels in PR- carcinoma cells. A boxplot (results not shown) revealed that GH1 expression was significant for both

ER and PR. GH1 was overexpressed in ER- and PR- subgroups compared to ER+ and PR+ carcinomas.

OS ER+/PR+

Gene Symbol	β	HR	95% CI (HR)	adj. p
SSTR2	0.34	1.4	(1.14,1.72)	0.16

Table 8. Summary of Genes with Expression Levels Associated with OS for 118 ER+/PR+ Breast Carcinomas. As described in Methods and Materials, expression levels of 142 candidate genes were determined using LCM-procured cells from 247 breast cancer biopsies.

PFS ER+/PR-

Gene Symbol	β	HR	95% CI (HR)	adj. p
CALCR	1.11	3.03	(1.46,6.29)	0.15
ACVR1B	1.16	3.2	(1.48,6.95)	0.15
EPO	0.61	1.85	(1.25,2.74)	0.15

Table 9. Summary of Genes with Expression Levels Associated with PFS for 28 ER+/PR- Breast Carcinomas. As described in Methods and Materials, expression levels of 142 candidate genes were determined using LCM-procured cells from 247 breast cancer biopsies.

PFS ER-/PR-

Gene Symbol	β	HR	95% CI (HR)	adj. p
POMC	0.4	1.49	(1.14,1.94)	0.25
PRL	0.63	1.88	(1.23,2.86)	0.25

Table 10. Summary of Genes with Expression Levels Associated with PFS for 68 ER-/PR- Breast Carcinomas. As described in Methods and Materials, expression levels of 142 candidate genes were determined using LCM-procured cells from 247 breast cancer biopsies.

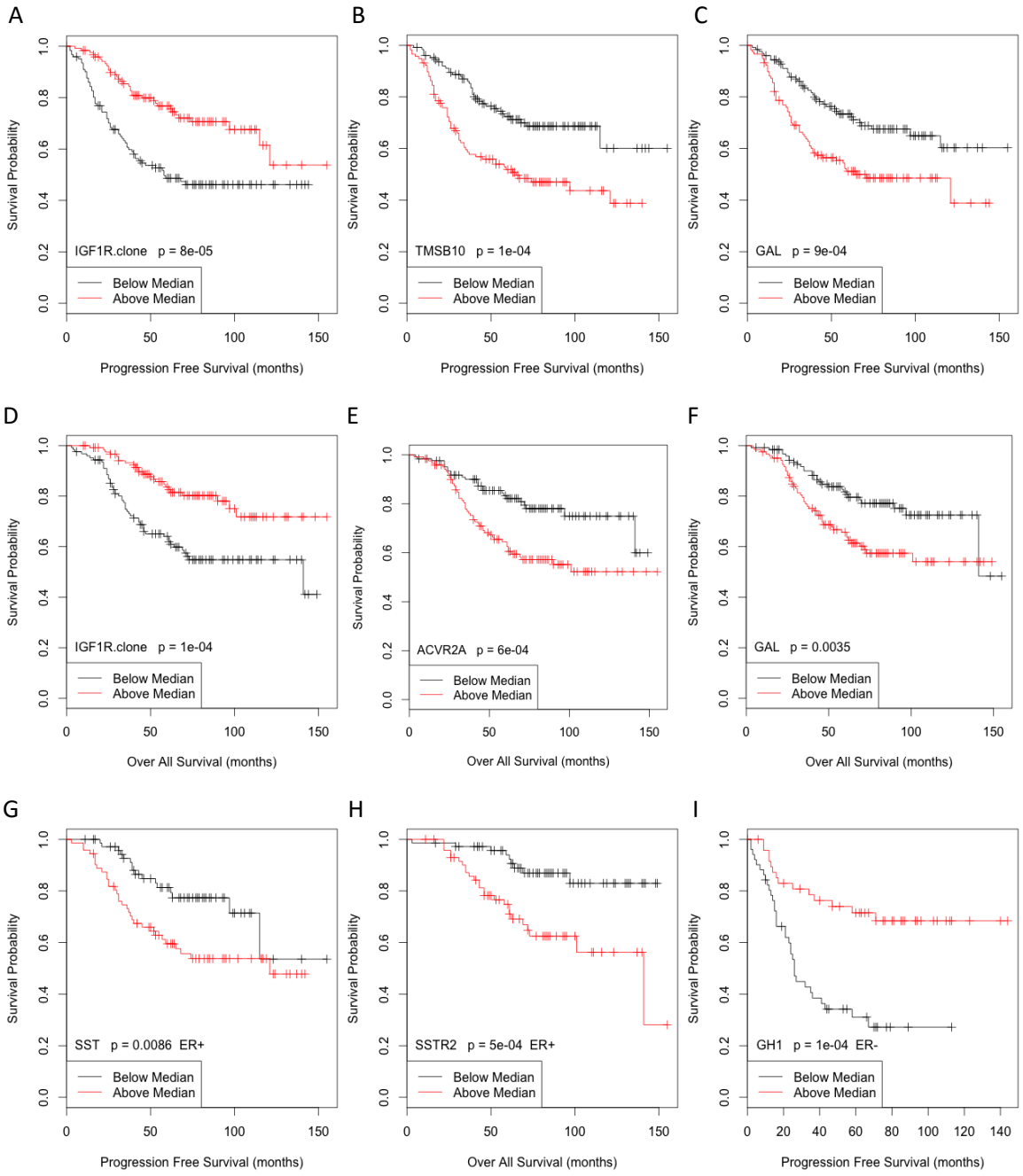

Figure 2 Representative Kaplan-Meier Plots Comparing Above and Below the Median Relative Gene Expression (1). Without regards toward receptor status, IGF1R.clone (A), TMSB10 (B) and GAL (C) displayed significance difference between groups for PFS, while IGF1R.clone (D), ACVR2A (E) and GAL (F) were significant for OS. For ER+ tumors, SST (G) displayed significance difference between groups for PFS, while SSTR2 (H) was significant for OS. For ER- tumors, GH1 (I) displayed significance difference between groups for PFS.

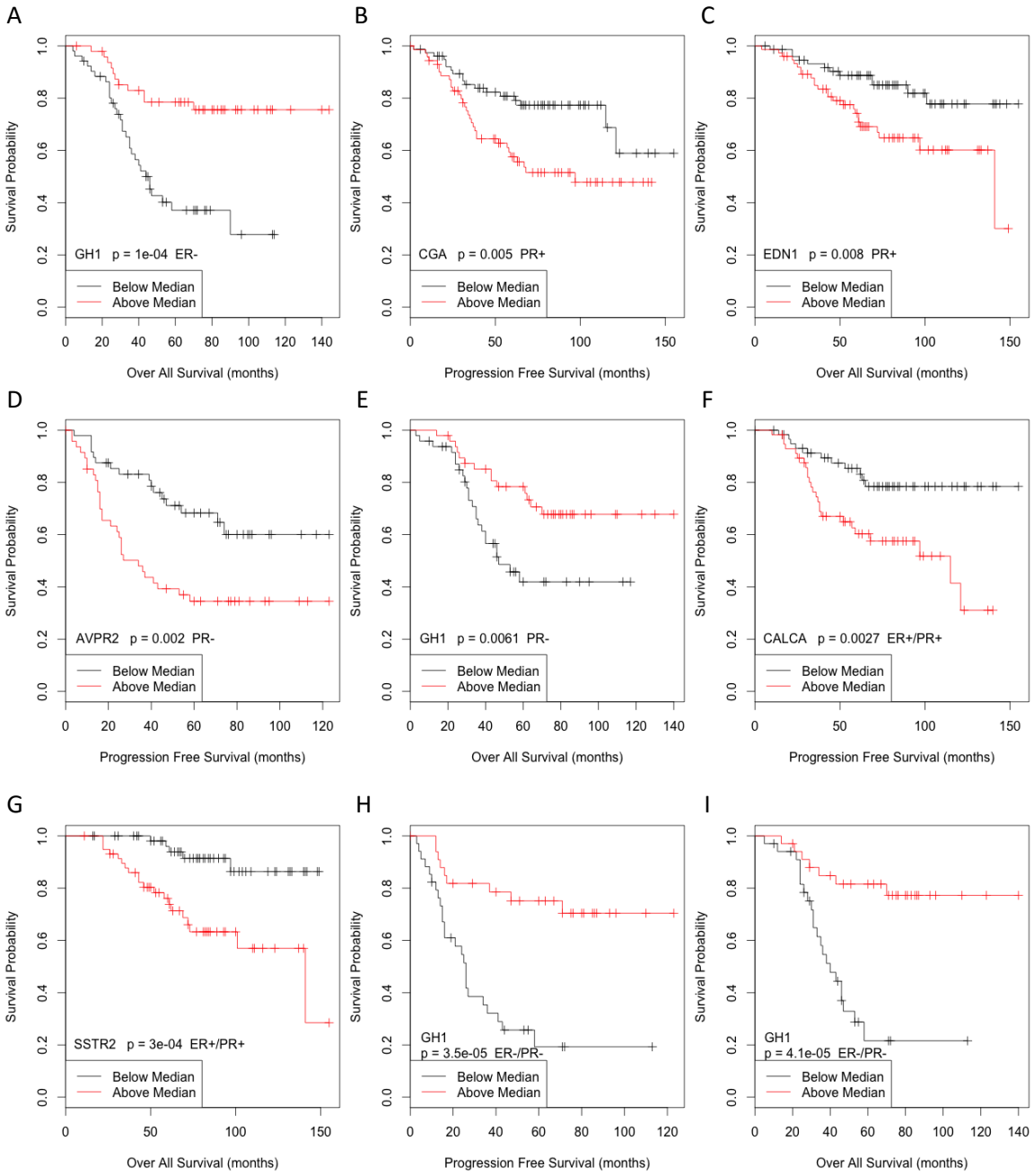

Figure 3. Representative Kaplan-Meier Plots Comparing Above and Below the Median Relative Gene Expression (2). For ER- tumors, GH1 (A) displayed significance difference between groups for OS. For PR+ tumors, CGA (B) displayed significance difference between groups for PFS, while EDN1 (C) was significant for OS. For PR- tumors, AVPR2 (D) displayed significance difference between groups for PFS, while GH1 (E) was significant for OS. For ER+/PR+ tumors, CALCA (F) displayed significance difference between groups for PFS, while SSTR2 (G) was significant for OS. For ER-/PR- tumors, GH1 (H) displayed significance difference between groups for both PFS and OS.

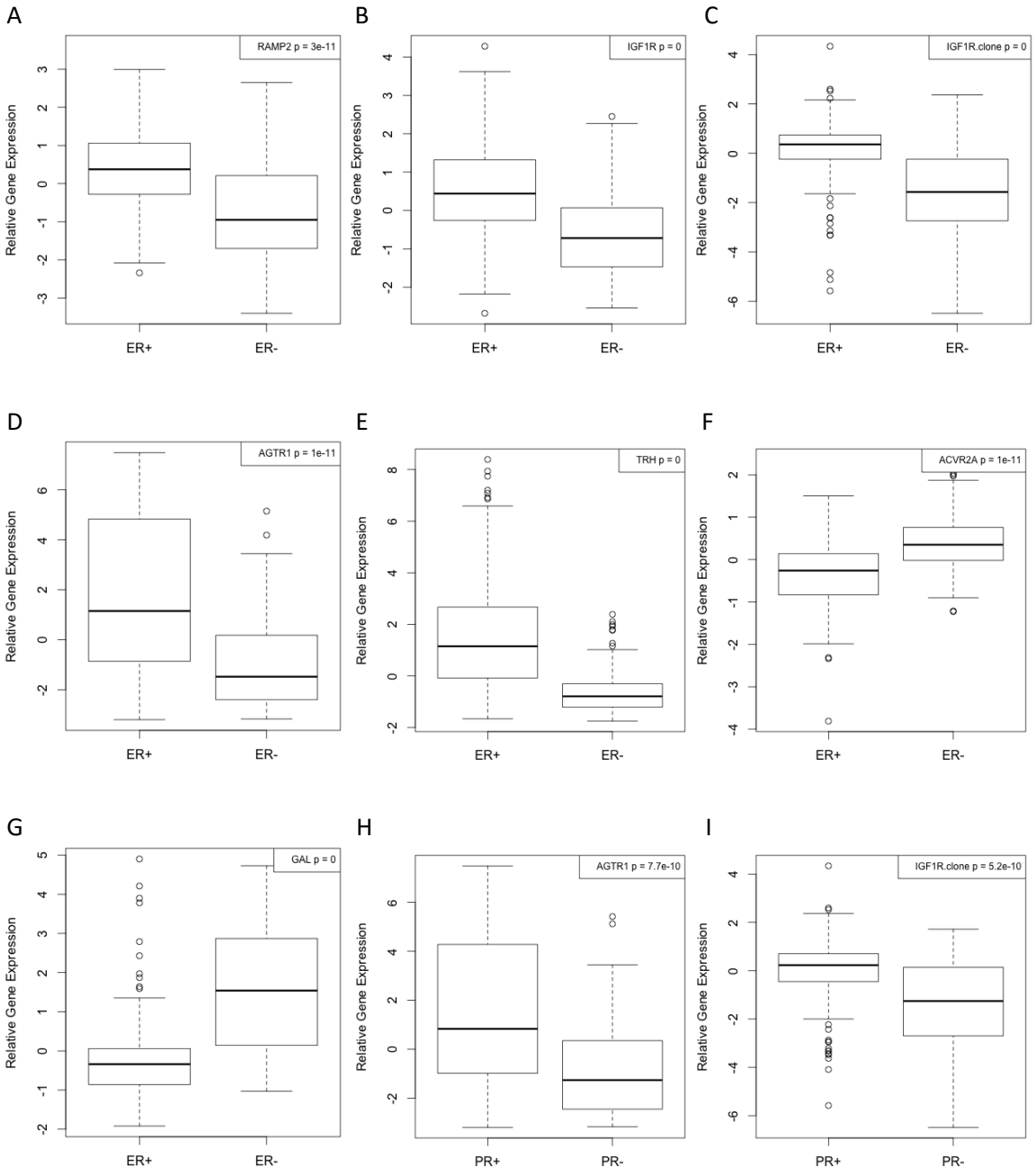

Figure 4. Representative Boxplots Comparing Distributions of Relative Gene Expressions by Receptor Status. The gene expression of RAMP2 (A), IGF1R (B), IGF1R.clone (C), AGTR1 (D) and TRH (E) in ER+ tumors was expressed at a significantly higher level than ER- tumors. The gene expression of ACVR2A (F) and GAL (G) in ER- tumors was expressed at a significantly higher level than ER+ tumors. The gene expression of AGTR1 (H) and IGF1R.clone (I) in PR+ tumors was expressed at a significantly higher level than PR- tumors.

Discussion

The use of these comprehensive, de-identified databases containing a variety of molecular biological results complemented by characteristics of the breast carcinoma and the clinical outcome of the patient have allowed the exploration of original questions and hypotheses related to biomarkers of clinical relevance in breast carcinoma management. These matchless resources combined with a range of statistical tests, bioinformatics tools and novel, sophisticated software have given me a unique opportunity to learn new approaches and integrate results of complex molecular assays with clinical outcomes. As a result, we have uncovered a gold mine of potential interrelationships between gene expression/ protein biomarker levels and breast carcinoma behavior.

In support of our approach outlined in Figure 1 (Flow Diagram), the integrity of the databases mined in these investigations is based on the following facts. Firstly, each of the tissue biopsies of breast cancer was collected in Pathology Departments of hospitals utilizing standardized protocols for specimen handling developed by Dr. Wittliff and then processed and stored deep frozen under stringent conditions established in the Hormone Receptor Laboratory. The HRL holds both CLIA and Commonwealth of Kentucky licenses as a clinical laboratory.

Secondly, each of the specimens in the Biorepository and results accumulated in the comprehensive databases were de-identified under approval of the Institutional Review Board (IRB) of the University of Louisville. Next, using these frozen biopsy specimens, tissue sections were uniformly processed and only

breast carcinoma cells were Laser Capture Microdissected in a non-destructive fashion using protocols established by Dr. Wittliff and collaborators. (J. L. Wittliff, 2010) The extraction and purification of RNA from LCM-procured cells as well as RNA amplification and microarray analyses of gene expression were performed in well controlled, highly reproducible assays. This is documented by acceptance of various publications (Andres et al., 2013; Metzler et al., 2015; J. L. Wittliff, 2010) as well as acceptance of patent applications filed describing molecular profiling of breast cancers (JLW & Arcturus patent applications).

Lastly, microarray results of expression of certain genes were validated by qPCR (Andres & Wittliff, 2011, 2012; Kerr II & Wittliff, 2011) and determinations of ER and PR protein levels in tissue biopsies were performed with FDA-approved kits that gave quantitative results. My Co-mentor, Dr. Wittliff developed the radio ligand-binding assays which not only measured the levels of the active ER and PR proteins but gave affinities of the association of the sex hormone receptors for their ligands which were radioactively labeled. Thus the databases and statistical approaches that I have used to investigate expression of genes for 142 peptide/protein hormone and their receptor proteins in breast carcinoma cells to derive clinically useful molecular signatures have great reliability.

POMC was the most significant gene for univariate Cox regression without regard to receptor status. POMC is a 241 amino acid long polypeptide typically produced in the pituitary gland and is cleaved at various sites to make a number of other proteins including adrenocorticotropic hormone (ACTH) and β-lipotropin. (Kaushal & Sinha, 2015) POMC is secreted in response to corticotropin-releasing

hormone (CRH), a gene also significant for univariate Cox regression. ACTH is necessary for steroidogenesis, which converts cholesterol into steroids such as progesterone, estrogen and cortisol (an immunosuppressant). (Hanukoglu, 1992) The gene pairs POMC & MC5R, SST & SSTR1, SST & SSTR3, RLN1 & RXFP3 and RLN2 & RXFP3 had both genes individually among the univariate significant genes.

Twelve of the 34 genes significant for univariate Cox regression for PFS without regard to receptor status were also significant for the log-rank test derived for Kaplan-Meier plots. Of the twelve genes showing significance for univariate Cox regression and Kaplan-Meier plots, POMC, VIPR1, RLN1, ACVR2A, SST and CALCR will be in the 14-gene signature for predicting PFS as described in the flow diagram. (Figure 1)

Nearly half the genes (67 of 142) had significantly different relative gene expressions for either ER+ versus ER- or PR+ versus PR- as shown by a representative sample of boxplots. An important question for these 67 genes is whether estrogen or progesterone is regulating them. The Kaplan-Meier plots without regards to receptor status provide some evidence for this question. Among the 22 significant genes for PFS or OS for Kaplan-Meier plots, thirteen are common to those significant for differences between ER and PR status expressed in the boxplots. The common genes are IGF1R.clone, AGTRAP, ACVR2A, RLN1, TRH, POMC, GH1, VIPR1, EDN1, GAL, GHSR, TMSB10 and TMSB15A. These genes are associated with the signaling pathways for estrogen or progesterone and require further investigation to discern the association.

Growth hormone (GH1), POMC and CALCR were highly significant for univariate Cox regression and Kaplan-Meier plots for ER- breast carcinomas and will be in the 9-gene signature for ER- cancers as described in the flow diagram. (Figure 1) Interestingly, the only gene that showed a protective effect as relative gene expression increases (negative beta coefficient) for PFS was GH1. GH1 is secreted in the pituitary gland and up regulated by growth hormone releasing hormone (GHRH) and down regulated by somatostatin in the hypothalamus. (Wagner et al., 2006) GH1 promotes insulin-like growth factor (IGF1) in the liver. IGF1 has been targeted in a pathway to regulate the metastasis of breast cancer. (Sachdev, 2008; Yang & Yee, 2012) There was no surprise to find IGF1 eventually show up in our 9-gene model for ER- breast cancers. Why does the increase in the expression of a gene, which promotes a gene in a known breast cancer metastasis pathway, has a protective effect against the progression of breast cancer?

Clinically, cancers identified as ER-/PR- have the worst prognosis. Prolactin (PRL) was the most significant gene for PFS and OS for these cancers followed closely by GH1, POMC and CALCR, mentioned above for ER- cancers. Elevated circulating PRL has been shown to increase the risk of breast cancer *in situ*. (Tikk et al., 2015) PRL and prolactin receptor will appear in our 9-gene molecular signature for ER- breast cancers. Before modeling any molecular signatures, we need to evaluate how these hormones and receptors may be affecting one another statistically. Collectively, these results lead us to look at various models and statistical techniques for multivariate analyses.

CHAPTER III

MULTIVARIATE ANALYSES

The analyses described in Chapter One using gene expression levels from LCM-procure breast carcinoma cells were expanded with various techniques for multivariate analyses as described in Figure 1 flow diagram. Cox regressions were computed for two variable models and models with an interaction term with gene expression results from the candidate peptide/protein hormones and their receptors. Least absolute shrinkage and selection operator (LASSO) was used to fit a model with each of the 142 candidate genes at one time to determine the relationship with prediction of risk of breast cancer recurrence.

In Chapter Two, univariate Cox regression evaluated the expression of each gene's influence on a breast cancer patient's clinical outcome. A unique aspect to this study is the ability to examine a peptide/protein hormone and their cognate receptor/s as a combination and the manner in which these pairs of gene expressions may play a role in predicting DFS and OS. A novelty of our study is that we have discovered many primary breast carcinomas exhibit elevated levels of expression of numerous genes for peptide/protein hormones and their receptors which are known to regulate physiologic pathways. There is no biochemical interaction between mRNAs but the proteins they eventually encode definitely interact. The problem with measuring mRNA is that we cannot say with certainty

that they will be coded into proteins. The correlation is there but the cell may have mutations that fail to translate the mRNA into the protein. If we wanted to ensure the levels of protein match the levels of mRNA then we would need to do qPCR or other techniques that measure proteins in the cell. (Wu & Singh, 2012) Although there is no biological interaction/association between each of the mRNA molecules of a hormone and that of its receptor, a statistical interaction may be computed.

Methods and Materials

Multivariate Cox Regression

Two types of Cox models were employed to evaluate the relationships between gene expression results for peptide/protein hormones and their cognate receptors with clinical outcomes. The first computes a two variable model (hormone + receptor) and the second model adds a third interaction term to the model. The two variable model has the form $h(t) = h_0(t)\ \exp\ (\beta_1 x_1 + \beta_2 x_2)$ and the interaction model has the form $h(t) = h_0(t)\exp(\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2)$. (Klein & Moeschberger, 1997) The expression of the hormone and receptor is represented by x_1 and x_2, respectively. The baseline hazard at time *t* is represented by $h_0(t)$ while the hazard at time *t* is represented by $h(t)$. In a two-variable Cox model, the beta coefficient for the hormone reflects the weight for which the expression for the hormone's gene has on the hazard function while controlling for the expression of the receptor gene. Likewise, the beta coefficient for the receptor gene expression reflects its weight on the hazard function while controlling for the gene expression of the hormone. These models provide the HR for cancers expressing a gene while accounting for the levels of the expression of its partner

gene. In the interaction model, if the beta coefficient for the interaction term is significant, then the gene expression results for the hormone and its receptor have a strong statistical dependence on each other. Hazard ratios for the hormone and receptor gene expression levels are conditioned on the level of its associated partner. The exponentiation of β_3 represents the excess hazard from the statistical interaction of the gene expressions of the hormone and its cognate receptors.

Multivariate Cox regression analyses were extended to all of the subsets of receptor status as discussed in Chapter Two Methods and Materials. Kaplan-Meier (KM) plots were constructed along with each pair with and without regards to ER/PR status to assess association with prediction of breast cancer recurrence. KM plots were constructed with four categories: above the median gene expression for both the hormone and receptor, below the median for both the hormone and receptor and the two discordant combinations.

Least Absolute Shrinkage and Selection Operator

With such a large number of genes (e.g., 142), stepwise selection methods to determine significant genes in a model are not feasible (Austin & Tu, 2004). An increase in the standard error of coefficients due to the multicollinearity of gene expression values may lead to type II errors – failure to reject a false hypothesis. An alternative method, Least Absolute Shrinkage and Selection Operator (LASSO), was used to evaluate each of the 142 candidate genes in primary breast carcinoma cells of 247 patients in models for DFS and OS. LASSO penalizes the size of $\vec{\beta}$ and removes genes whose coefficients are close to zero. The maximum

likelihood estimates $\hat{\beta}$ are derived by maximizing the penalized Cox log partial likelihood with the form $l(\beta) - \sum_{j=1}^{p} \lambda|\beta_j|$, where $l\,(\beta)$ represents the standard log Cox log partial likelihood and λ is the shrinkage parameter. Using 10-fold cross-validation to minimizing the mean square error determined the optimal value for λ (Andres et al., 2013). A larger λ corresponds to a larger penalty on the Cox log partial likelihood and thus removes more variables from the model alleviated the more predictive ability of overfitting a model. Further examination of the two clinical outcomes (PFS & OS) with combinations of biomarker profiles were assessed as well. The commands *penalized* and *optL1* in the R package *penalized* were used to perform an optimized L1 penalty, "LASSO", for a Cox model. (J. Goeman, Meijer, & Chaturvedi, 2014; J. J. Goeman, 2010)

Results

Multivariate Cox Regression with Interaction of PFS and OS without Regard to ER and PR Status of the Cancer Biopsy

For interaction models, the sign of the beta coefficient for the interaction term and its corresponding p-value are the most informative outputs from these types of Cox regressions. A negative beta coefficient reflects an antagonistic effect on the hazard, while a positive beta coefficient has a synergistic effect. For instance in Table 11, HCRT (orexin) and HCRTR2 (hypocretin receptor 2) exhibited a significant interaction term (adjusted p=0.23) for a false discovery rate cutoff of p<0.3 for OS with a positive beta coefficient (β=1.53) that indicates there is a significant increase in risk when expression levels of the two genes are

considered together. HRs for the main effects (i.e., hormone or receptor) of these Cox models are conditional HRs that depend upon the expression levels of the other gene in the model and are not shown for the sake of brevity. Table 1 shows the HR and 95% confidence interval for the HR of the interaction term and represents the excess risk from a synergistic effect or reductive risk from an antagonistic effect.

OS

Gene Symbol	β	HR	95% CI (HR)	adj. p
HCRT & HCRTR2	1.53	4.63	(1.75,12.23)	0.23

Table 11. Summary of the Interaction Term for Hormone and Receptor Gene Pairs with Expression Levels Associated with OS without Regard to Receptor Status. As described in Methods and Materials, expression levels of 142 candidate genes were determined using LCM-procured cells from 247 breast cancer biopsies.

Multivariate Cox Regression with Interaction of PFS and OS with Regard to ER and PR Status of the Cancer Biopsy

For ER+ cancers, two hormone-receptor pairs, INHA-ACVR2B and GNRH2-GNRHR, exhibited significance when associated with OS (Table 12). Interpretation of these interactions should be done with care. For example, inhibin (INHA) and activin A receptor, type IIB (ACVR2B) has an antagonistic effect on the overall hazard since the interaction beta coefficient (β_3) is negative. Yet, the beta coefficients for the main effects (i.e., INHA (β_1) and ACVR2B (β_2)) are positive (β_1=0.52, β_2=1.00), which can be interpreted as an increase in the expression level of INHA or ACVR2B increases the risk of death (OS) at time *t*. In other words, the

overall hazard is a convolution of the increased risk from the levels of INHA and ACVR2B and the decreased risk from the statistical interaction of the two at these levels. INHA-ACVR2B and GNRH2-GNRHR would be candidates for further study to test signaling pathways for endocrine autonomy.

OS Interaction ER+

Gene Symbol	β	HR	95% CI (HR)	adj. p
INHA & ACVR2B	-1.62	0.2	(0.08,0.51)	0.09
GNRH2 & GNRHR	9.99	21907.54	(26.77,17930991)	0.20

Table 12. Summary of the Interaction Term for Hormone and Receptor Gene Pairs with Expression Levels Associated with OS for 146 ER+ Breast Carcinomas. As described in Methods and Materials, expression levels of 142 candidate genes were determined using LCM-procured cells from 247 breast cancer biopsies.

For ER- cancers, Table 13 shows gene expression levels of five hormone-receptor pairs exhibiting significance for interaction for OS. The interaction term for HCRT-HCRTR2 was significant for ER- breast cancer and also significant for breast cancers without regard to receptor status. Interestingly, the hormone-receptor pairs significant for interaction in ER+ cancers reappeared with different genes but from the same group. For example, INHA-ACVR2B was significant in ER+ breast cancers and inhibin beta C (INHBC) and activin receptor type IB (ACVR1B) was significant in ER- breast cancers. This is also similar for the hormone gonadotrophin, which expressed type II significant for ER+ cancers and type II significant for ER- cancers. Three out of the five pairs showing significance for ER- cancers typically produce their hormone in the hypothalamus and regulate a variety of other peptide/protein hormones. These three hormones, CRH, GNRH1

and HCRT, should be examined more closely with the hormones they regulate to verify any autocrine pathways in ER- breast cancers.

OS Interaction ER-				
Gene Symbol	**β**	**HR**	**95% CI (HR)**	**adj. p**
IAPP & CALCR	3.93	50.99	(2.95,881.85)	0.20
INHBC & ACVR1B	0.46	1.59	(1.15,2.2)	0.20
CRH & CRHR1	-1.9	0.15	(0.04,0.53)	0.20
GNRH1 & GNRHR	2.77	15.9	(2.29,110.59)	0.20
HCRT & HCRTR2	1.5	4.47	(1.37,14.55)	0.30

Table 13. Summary of the Interaction Term for Hormone and Receptor Gene Pairs with Expression Levels Associated with OS for 101 ER- Breast Carcinomas. As described in Methods and Materials, expression levels of 142 candidate genes were determined using LCM-procured cells from 247 breast cancer biopsies.

For PR+ cancers, six pairs of hormone-receptor combinations showed significance for OS (Table 14) Amylin (IAPP) and calcitonin receptor (CALCR) exhibited the largest effect size for statistical interaction for PR+ cancers and previously in ER- cancers. Because elevated levels of ER promote elevated levels of PR, one would not expect genes that are expressed in ER- cancers to be expressed in PR+ cancers. This anomaly should be investigated as a new finding in order to rule out as simple randomness. POMC was identified for the first time in these interaction models with its partner melanocortin 5 receptor (MC5R).

OS Interaction PR+

Gene Symbol	β	HR	95% CI (HR)	adj. p
IAPP & CALCR	3.73	41.61	(2.68,645.8)	0.23
RLN2 & RXFP1	0.64	1.91	(1.2,3.02)	0.23
CGA & TSHR	0.27	1.31	(1.08,1.58)	0.23
GHRH & GHRHR	0.58	1.79	(1.16,2.75)	0.23
EDN1 & EDNRA	-1.1	0.33	(0.14,0.78)	0.27
POMC & MC5R	2.78	16.07	(1.69,153.35)	0.30

Table 14. Summary of the Interaction Term for Hormone and Receptor Gene Pairs with Expression Levels Associated with OS for 151 PR+ Breast Carcinomas. As described in Methods and Materials, expression levels of 142 candidate genes were determined using LCM-procured cells from 247 breast cancer biopsies.

For LCM-procured breast carcinoma cells from PR- cancers, four hormone and receptor pairs revealed significance for an association for OS (Table 15). Orexin (HCRT) along with two of its cognate receptors, HCRTR1 and HCRTR2, synergistically increased the risk for their hazard functions. An interesting result is the significance of the pair cortistatin (CORT) and SSTR4 in ER- cancers. In ER- cancers, POMC and PRL are overexpressed with relative gene expression means of 0.38 and 0.24, respectively, while CORT and SSTR4 are under expressed with relative gene expression means of -0.2 and -0.1, respectively. CORT has been shown to have an inhibitory effect upon the production of POMC and PRL (Córdoba-Chacón et al., 2011). If elevated gene expressions of POMC and/or PRL in ER- cancer cells are in a signally pathway to cause progression or death from the disease, then a therapeutic treatment for ER- cancers may be to find ways to increase CORT and SSTR4 gene expression.

OS Interaction PR-

Gene Symbol	β	HR	95% CI (HR)	adj. p
HCRT & HCRTR2	1.87	6.48	(2.09,20.07)	0.14
CRH & CRHR1	-1.79	0.17	(0.04,0.63)	0.24
HCRT & HCRTR1	2.31	10.11	(1.81,56.46)	0.24
CORT & SSTR4	1.14	3.14	(1.38,7.12)	0.24

Table 15. Summary of the Interaction Term for Hormone and Receptor Gene Pairs with Expression Levels Associated with OS for 96 PR- Breast Carcinomas. As described in Methods and Materials, expression levels of 142 candidate genes were determined using LCM-procured cells from 247 breast cancer biopsies.

For ER+/PR+ cancers, three pairs of hormone-receptor combinations showed significance for estimating the relationship with PFS (Table 16). All three pairs typically produce their hormones in the anterior pituitary gland. The hormone-receptor pair, pituitary adenylate cyclase-activating peptide (ADCYAP1) and pituitary adenylate cyclase-activating peptide receptor (ADCYAP1R), indicated the highest significance among these pituitary produced hormone pairs. ADCYAP1 has been tested as a methylation biomarker for cervical and endometrial cancer (Jung et al., 2011; Wentzensen et al., 2014)

PFS Interaction ER+/PR+

Gene Symbol	β	HR	95% CI (HR)	adj. p
ADCYAP1 & ADCYAP1R1	-0.65	0.52	(0.36,0.76)	0.08
FSHB & FSHR	24.77	5.72E+10	(2153,1.52E+18)	0.23
LHB & LHCGR	9.58	14473	(15,1.35E+7)	0.23

Table 16. Summary of the Interaction Term for Hormone and Receptor Gene Pairs with Expression Levels Associated with PFS for 118 ER+/PR+ Breast Carcinomas. As described in Methods and Materials, expression levels of 142 candidate genes were determined using LCM-procured cells from 247 breast cancer biopsies.

When carcinomas expressing ER+/PR- status, six gene pairs were found significant with three expressing synergistic and three expressing antagonistic effects on the hazard function (Table 17). The effect sizes are large for these six pairs but caution should be taken since the sample size is relatively small.

OS Interaction ER+/PR-

Gene Symbol	β	HR	95% CI (HR)	adj. p
SST & SSTR4	8.72	6096.72	(11.48,3237235.64)	0.24
INHA & ACVR2B	-7.11	0	(0,0.13)	0.24
RLN1 & RXFP1	5.67	289.4	(3.81,22009.88)	0.24
EPO & EPOR	14.1	1335702.39	(20,8.79E+10)	0.24
EDN1 & EDNRB	-3.83	0.02	(0,0.44)	0.24
VIP & VIPR1	-3.61	0.03	(0,0.39)	0.24

Table 17. Summary of the Interaction Term for Hormone and Receptor Gene Pairs with Expression Levels Associated with OS for 28 ER+/PR- Breast Carcinomas. As described in Methods and Materials, expression levels of 142 candidate genes were determined using LCM-procured cells from 247 breast cancer biopsies.

For ER-/PR+, two gene pairs were significant for an association with PFS and 12 gene pairs were significant for an association with OS (Table 18). There should be no surprise to find the pancreatic hormone amylin (IAPP) and its cognate pair calcitonin receptor (CALCR) as the most significant pair for assessing the relationship with both PFS and OS since its statistical interaction was found to be the most significant separately for ER- and PR+ cancers. The heart hormone atrial-natriuretic peptide (NPPA) and each of its three receptors (NPR1, NPR2 and NPR3) had a significant synergistic effect on the hazard for both predicting OS. Yet, caution should be taken with these results due to the small sample size of 33 ER-/PR+ cancers.

A. PFS Interaction ER-/PR+

Gene Symbol	β	HR	95% CI (HR)	adj. p
IAPP & CALCR	7.49	1781.76	(15.68,202470.3)	0.22
NPPA & NPR3	3.58	35.86	(3.12,411.6)	0.23

B. OS Interaction ER-/PR+

Gene Symbol	β	HR	95% CI (HR)	adj. p
IAPP & CALCR	6.71	818.18	(9.23,72540.58)	0.17
INHBA & ACVR1	0.92	2.52	(1.35,4.7)	0.17
NPPA & NPR3	3.96	52.38	(3.41,804.52)	0.17
INHBA & ACVR1B	-0.76	0.47	(0.24,0.89)	0.24
INHBC & ACVR1B	1.09	2.96	(1.27,6.92)	0.24
INHBC & ACVR2B	0.98	2.66	(1.23,5.74)	0.24
INHBE & ACVR2A	-1.12	0.33	(0.13,0.81)	0.24
RLN2 & RXFP1	0.85	2.33	(1.14,4.76)	0.24
CGA & TSHR	0.29	1.34	(1.04,1.72)	0.24
NPPA & NPR1	11.18	71334.66	(3.98,1.27E+9)	0.24
NPPA & NPR2	8.94	7617.82	(4.57,1.27E+7)	0.24
CORT & SSTR5	10.36	31462.61	(3.73,2.65E+8)	0.24

Table 18. Summary of the Interaction Term for Hormone and Receptor Gene Pairs with Expression Levels Associated with either PFS (A) or OS (B) for 33 ER-/PR+ Breast Carcinomas. As described in Methods and Materials, expression levels of 142 candidate genes were determined using LCM-procured cells from 247 breast cancer biopsies.

For ER-/PR- cancers, two pairs of hormone-receptor genes showed expression levels that exhibited significance for predicting OS (Table 19). HCRT and HCRTR2 were found to be significant in both ER- and PR- cancers separately. Surprisingly a new gene pair, arginine vasopressin (AVP) and cullin 5 (CUL5) appeared with an antagonistic effect on its hazard. CUL5 has been shown to inhibit cell proliferation (Burnatowska-Hledin et al., 2004; Van Dort et al., 2003).

OS Interaction ER-/PR-

Gene Symbol	β	HR	95% CI (HR)	adj. p
HCRT & HCRTR2	1.97	7.21	(1.87,27.79)	0.24
AVP & CUL5	-1.44	0.24	(0.09,0.62)	0.24

Table 19. Summary of the Interaction Term for Hormone and Receptor Gene Pairs with Expression Levels Associated with OS for 68 ER-/PR- Breast Carcinomas. As described in Methods and Materials, expression levels of 142 candidate genes were determined using LCM-procured cells from 247 breast cancer biopsies.

Multivariate Cox Regression for the Hormone of PFS and OS without Regard to ER and PR Status of the Cancer Biopsy

Two-variable models were employed to evaluate gene expression results for the hormones while controlling for all the levels of relative gene expression for their cognate receptors. In Table 20, the beta coefficient for the hormone term was analyzed for PFS without regard to sex-hormone receptor status. Among the 20 gene pairs in the Table 20, the hormones with multiple receptors had multiple appearances. Each of the hormones were significant for univariate analyses for PFS with no significant effect of their betas when accounting for the statistical presence of its receptor.

In Table 21, the beta coefficient for the hormone term was analyzed for OS without regard to sex-hormone receptor status. The strength of the signal from POMC regardless its cognate receptor dominated the significance for an association for OS. All gene pairs significant for OS were significant for PFS including secretin (SCT) with secretion receptor (SCRT) and pancreatic polypeptide (PPY) with neuropeptide Y receptor (NPY4R).

PFS Hormone

Two Gene Model	β for Hormone	HR	95% CI (HR)	adj. p
POMC + MC1R	0.54	1.71	(1.37,2.15)	<0.001
POMC + MC2R	0.56	1.74	(1.39,2.18)	<0.001
POMC + MC3R	0.54	1.72	(1.37,2.15)	<0.001
POMC + MC4R	0.54	1.72	(1.37,2.15)	<0.001
POMC + MC5R	0.51	1.66	(1.32,2.1)	<0.001
SST + SSTR2	0.56	1.75	(1.29,2.39)	0.007
SST + SSTR5	0.55	1.74	(1.28,2.36)	0.007
SST + SSTR3	0.51	1.67	(1.24,2.25)	0.01
SST + SSTR4	0.54	1.71	(1.25,2.34)	0.01
PPY + NPY4R	0.96	2.6	(1.35,5.02)	0.05
CRH + CRHR1	0.54	1.72	(1.17,2.55)	0.07
REN + ATP6AP2	0.6	1.82	(1.13,2.93)	0.13
PRL + PRLR	0.32	1.37	(1.06,1.77)	0.15
SCT + SCTR	0.57	1.78	(1.06,2.97)	0.21
EDN1 + EDNRA	0.39	1.48	(1.04,2.1)	0.21
EDN1 + EDNRB	0.4	1.49	(1.05,2.11)	0.21
RLN1 + RXFP1	-0.21	0.81	(0.67,0.99)	0.28
RLN1 + RXFP3	-0.2	0.82	(0.67,1)	0.30
RLN2 + RXFP1	-0.13	0.88	(0.77,1)	0.30
PTH + PTH2R	-1.26	0.28	(0.08,0.99)	0.30

Table 20. Summary of the Hormone Term from Two Variable Cox Models of Hormone and Receptor Gene Pairs with Expression Levels Associated with PFS. As described in Methods and Materials, expression levels of 142 candidate genes were determined using LCM-procured cells from 247 breast cancer biopsies.

OS Hormone				
Two Gene Model	**β for Hormone**	**HR**	**95% CI (HR)**	**adj. p**
POMC + MC1R	0.52	1.68	(1.34,2.1)	<0.001
POMC + MC2R	0.53	1.71	(1.36,2.13)	<0.001
POMC + MC3R	0.52	1.67	(1.34,2.1)	<0.001
POMC + MC4R	0.52	1.68	(1.34,2.1)	<0.001
POMC + MC5R	0.48	1.62	(1.28,2.05)	0.002
SCT + SCTR	0.77	2.17	(1.24,3.8)	0.13
PPY + NPY4R	0.88	2.4	(1.16,4.97)	0.30

Table 21. Summary of the Hormone Term from Two Variable Cox Models of Hormone and Receptor Gene Pairs with Expression Levels Associated with OS. As described in Methods and Materials, expression levels of 142 candidate genes were determined using LCM-procured cells from 247 breast cancer biopsies.

Multivariate Cox Regression for the Genes of Hormones of PFS and OS without Regard to ER and PR Status of the Cancer Biopsy

Two-variable models were employed to evaluate gene expression results for the receptors while controlling for all the levels of relative gene expression for their cognate hormones. In Table 22, the beta coefficient for the receptor term was analyzed for PFS without regard to sex-hormone receptor status. Receptors appeared in nearly the same order they appeared in univariate analysis for PFS with no significant effect on their betas when accounting for the statistical presence of its hormone.

In Table 23, the beta coefficient for the receptor term was analyzed for OS without regard to sex-hormone receptor status. Similar to PFS, the receptors appeared in the same order as they did for significance in univariate Cox regression with no effect seen with the statistical presence of its cognate hormone.

This interpretation continues into the subtypes of ER/PR and for the sake of brevity will not be shown here.

PFS Receptor				
Two Gene Model	**β for Receptor**	**HR**	**95% CI (HR)**	**adj. p**
CALCA + CALCR	0.44	1.55	(1.2,2)	0.04
IAPP + CALCR	0.46	1.58	(1.22,2.03)	0.04
IGF1 + IGF1R.clone	-0.17	0.84	(0.75,0.94)	0.06
INS + IGF1R.clone	-0.17	0.85	(0.76,0.94)	0.07
CORT + SSTR1	0.62	1.86	(1.24,2.8)	0.07
SST + SSTR3	0.93	2.53	(1.2,5.35)	0.10
INHA + ACVR2A	0.31	1.37	(1.08,1.73)	0.10
INHA + ACVR2B	0.35	1.42	(1.07,1.89)	0.10
INHBA + ACVR2A	0.29	1.34	(1.06,1.7)	0.10
INHBA + ACVR2B	0.35	1.42	(1.08,1.87)	0.10
INHBB + ACVR2A	0.33	1.39	(1.09,1.77)	0.10
INHBC + ACVR2A	0.32	1.38	(1.08,1.77)	0.10
INHBC + ACVR2B	0.35	1.42	(1.08,1.87)	0.10
INHBE + ACVR2A	0.29	1.34	(1.06,1.7)	0.10
INHBE + ACVR2B	0.35	1.42	(1.08,1.86)	0.10
AVP + AVPR2	0.91	2.49	(1.19,5.23)	0.10
CORT + SSTR3	0.96	2.61	(1.25,5.44)	0.10
VIP + VIPR1	-0.28	0.75	(0.6,0.94)	0.10
NPY + NPY1R	-0.17	0.85	(0.73,0.97)	0.12
INHBB + ACVR2B	0.32	1.38	(1.04,1.81)	0.13
ADCYAP1 + VIPR1	-0.25	0.78	(0.62,0.97)	0.13
GHRL + GHSR	-1.14	0.32	(0.12,0.88)	0.14
NPY + NPY6R	-1	0.37	(0.15,0.9)	0.14
INHBB + ACVR1B	0.27	1.31	(1.02,1.67)	0.15
GH1 + GHR	0.43	1.54	(1.04,2.28)	0.15
AGT + RXFP3	1.23	3.42	(1.09,10.76)	0.15
INS + INSR.AL365454	-0.25	0.78	(0.61,1)	0.17
INHBA + ACVR1	-0.27	0.76	(0.58,0.99)	0.17
INHBA + ACVR1B	0.26	1.29	(1.01,1.65)	0.17
INHBC + ACVR1	-0.28	0.75	(0.57,0.99)	0.17
INHBC + ACVR1B	0.26	1.29	(1.01,1.65)	0.17
INHBE + ACVR1B	0.25	1.29	(1.01,1.65)	0.17
INHA + ACVR1B	0.25	1.28	(1,1.65)	0.17
INHBE + ACVR1	-0.27	0.76	(0.58,1)	0.17

Table 22. Representative Summary of the Receptor Term from Two Variable Cox Models of Hormone and Receptor Gene Pairs with Expression Levels Associated with PFS. As described in Methods and Materials, expression levels of 142 candidate genes were determined using LCM-procured cells from 247 breast cancer biopsies.

OS Receptor

Two Gene Model	β for Receptor	HR	95% CI (HR)	adj. p
CALCA + CALCR	0.44	1.55	(1.17,2.06)	0.12
IAPP + CALCR	0.46	1.58	(1.19,2.1)	0.12
INS + IGF1R.clone	-0.17	0.84	(0.75,0.95)	0.12
SST + SSTR3	1.11	3.02	(1.34,6.83)	0.12
IGF1 + IGF1R.clone	-0.16	0.85	(0.75,0.96)	0.12
INHBB + ACVR2A	0.38	1.46	(1.11,1.93)	0.12
CORT + SSTR3	1.1	3.02	(1.35,6.73)	0.12
NPY + NPY6R	-1.4	0.25	(0.09,0.69)	0.12
INHA + ACVR2A	0.35	1.42	(1.08,1.87)	0.13
INHBA + ACVR2A	0.35	1.43	(1.08,1.88)	0.13
INHBC + ACVR2A	0.36	1.43	(1.08,1.89)	0.13
INHBE + ACVR2A	0.34	1.41	(1.08,1.85)	0.13
VIP + VIPR1	-0.31	0.74	(0.57,0.95)	0.17
NPY + NPY1R	-0.2	0.82	(0.69,0.97)	0.17
INS + IGF1R	-0.19	0.83	(0.69,1)	0.18
IGF1 + IGF1R	-0.18	0.83	(0.69,0.99)	0.18
AGT + RXFP3	1.29	3.65	(1.04,12.78)	0.18
INHA + ACVR1	-0.35	0.7	(0.51,0.98)	0.18
INHA + ACVR1B	0.3	1.35	(1.01,1.8)	0.18
INHA + ACVR2B	0.36	1.44	(1.04,1.99)	0.18
INHBA + ACVR1B	0.3	1.35	(1.01,1.79)	0.18
INHBA + ACVR2B	0.35	1.41	(1.03,1.94)	0.18
INHBB + ACVR1B	0.29	1.34	(1.01,1.77)	0.18
INHBC + ACVR1B	0.29	1.33	(1,1.77)	0.18
INHBC + ACVR2B	0.33	1.4	(1.02,1.91)	0.18
INHBE + ACVR2B	0.32	1.37	(1,1.88)	0.18
ADCYAP1 + VIPR1	-0.27	0.77	(0.6,0.99)	0.18
ADCYAP1 + VIPR2	-0.13	0.88	(0.78,0.99)	0.18
GHRL + GHSR	-1.35	0.26	(0.08,0.86)	0.18
VIP + VIPR2	-0.13	0.87	(0.78,0.99)	0.18
IAPP + RAMP2	-0.18	0.84	(0.7,1)	0.19
INHBC + ACVR1	-0.31	0.74	(0.54,1)	0.19
INHBA + ACVR1	-0.29	0.75	(0.55,1.01)	0.21

Table 23. Representative Summary of the Receptor Term from Two Variable Cox Models of Hormone and Receptor Gene Pairs with Expression Levels Associated with OS. As described in Methods and Materials, expression levels of 142 candidate genes were determined using LCM-procured cells from 247 breast cancer biopsies.

Kaplan-Meier Plots Discriminating Hormone-Receptor Pairs Whose Expression Levels Predict Clinical Outcomes

With 115 possible pairs of genes for peptide/protein hormones and their cognate receptors, 59 exhibited gene expression levels that were significant for the non-parametric log-rank test for PFS and 41 were significant for the non-parametric log-rank test for OS without regard to receptor status (i.e., ER+, ER-, PR+, PR- and ER/PR combinations). Figure 5 has a representative sample of the most significant gene pairs from the log-rank test. Three examples are listed for cancers with regard to receptor status and one example for ER+/PR+, ER-/PR-, ER+, ER-, PR+ and PR-. There is a complexity to comparing these Kaplan-Meier plots to our interaction models. The main effects of the hormone and receptor in the interaction models are a stronger driver of the behavior of the four categories: above/below the median expression for the hormone and above/below the median expression for the receptor. Kaplan-Meier plots are different from our previous analyses with univariate and multivariate Cox regression because the magnitude of the relative gene expression isn't considered in the analysis once cancers are stratified into their respective groups. One aspect to consider when evaluating these plots is the behavior of cancers that exhibit both elevated hormone and receptor gene expression levels.

Without regard to ER/PR status, the PFS expressed in Kaplan-Meier plots for SST & SSTR1 and SST & SSTR3 show a contrasting difference between cancers exhibiting the worst prognosis that express mRNA levels of the hormone and receptor above the median and cancers with the best prognosis that express

the mRNA levels of hormone and receptor below the median (Figure 4). Three plots discriminated cancers expressing mRNA above the median for the hormone and receptor as the worst prognostic group, which included the OS without regard to ER/PR status for POMC and MC5R, the OS for patients with ER+ cancers for SST and SSTR2 and the PFS for patients with ER-/PR- cancers for ADCYAP1 and VIPR2. For ER+/PR+ cancers, EDN1 & EDNRA was significant for OS for the interaction term from multivariate Cox regression. EDN1 & EDNRA was also significant for OS for KM plots with cancers expressing both the hormone and receptor below the median resulting in the best prognosis. Gene expression results shown in the Kaplan-Meier plots for AVP & AVPR2 (ER-, OS), MLN & MLNR (PR+, PFS) and PTH & PTH2R (PR-, PFS), are representative of discordant pairs (i.e., above the median expression of the hormone gene and below the median expression of the receptor gene) as the group of patients with the poorest prognosis. Although the molecular basis for this predicted clinical behavior is unknown currently, the gene expression patterns warrant further research to determine if they may be useful in predicting clinical outcomes.

Multivariable Gene Expression Model fitted using the LASSO

The use of regularization techniques such as LASSO allowed us to evaluate gene expression levels of all 142 gene candidates to a model by penalizing the number of parameters in the model. The analyses were performed for predicting either PFS or OS without regard to either ER or PR receptor status, as well as in relationship to the four different combinations of ER/PR. Using the loss function to decide which model is the best, only three groups showed the minimal loss function

not equal to all zeroes for beta coefficients. In other words, all but three models exhibited their best fitted models having coefficients with all zeroes, which translates to having no significant variables associated with PFS and OS when a penalty is put on the size of the beta coefficients.

The LASSO fitted and cross-validated model for predicting breast cancer outcome expressing ER+/PR+ status for predicting OS was composed of expression levels of seven hormone genes (IGF1, INHBB, GNRH2, PENK, CALCA, GAL and PTMS) and those for nine receptor genes (SSTR2, SSTR3, SSTR4, PRLHR, EPOR, PRLR, CUL5, NPY2R and NPY5R) (Table 24). Strikingly, expression results of each of the eight genes showing statistical significance in univariate Cox regression analyses for predicting OS of ER+/PR+ cancers reappeared in the model derived from LASSO. The LASSO derived model of gene expression results for predicting PFS of ER-/PR- cancers was composed of five hormone genes (AGT, POMC, GH1, PRL and PTH) and five receptor genes (CALR, SSTR1, AGTRAP, AVPR1A and AVPR2). Expression levels of these ten genes were all significant for predicting PFS of ER-/PR- cancers from univariate analyses. The LASSO derived model for prediction of PFS of ER- breast cancers was composed of four hormone genes (POMC, GH1, LHB and SCT) and two receptor genes (CALCR and AVPR1A). Expression levels of these six genes were all significant for assessing PFS of ER- breast cancers using univariate analyses.

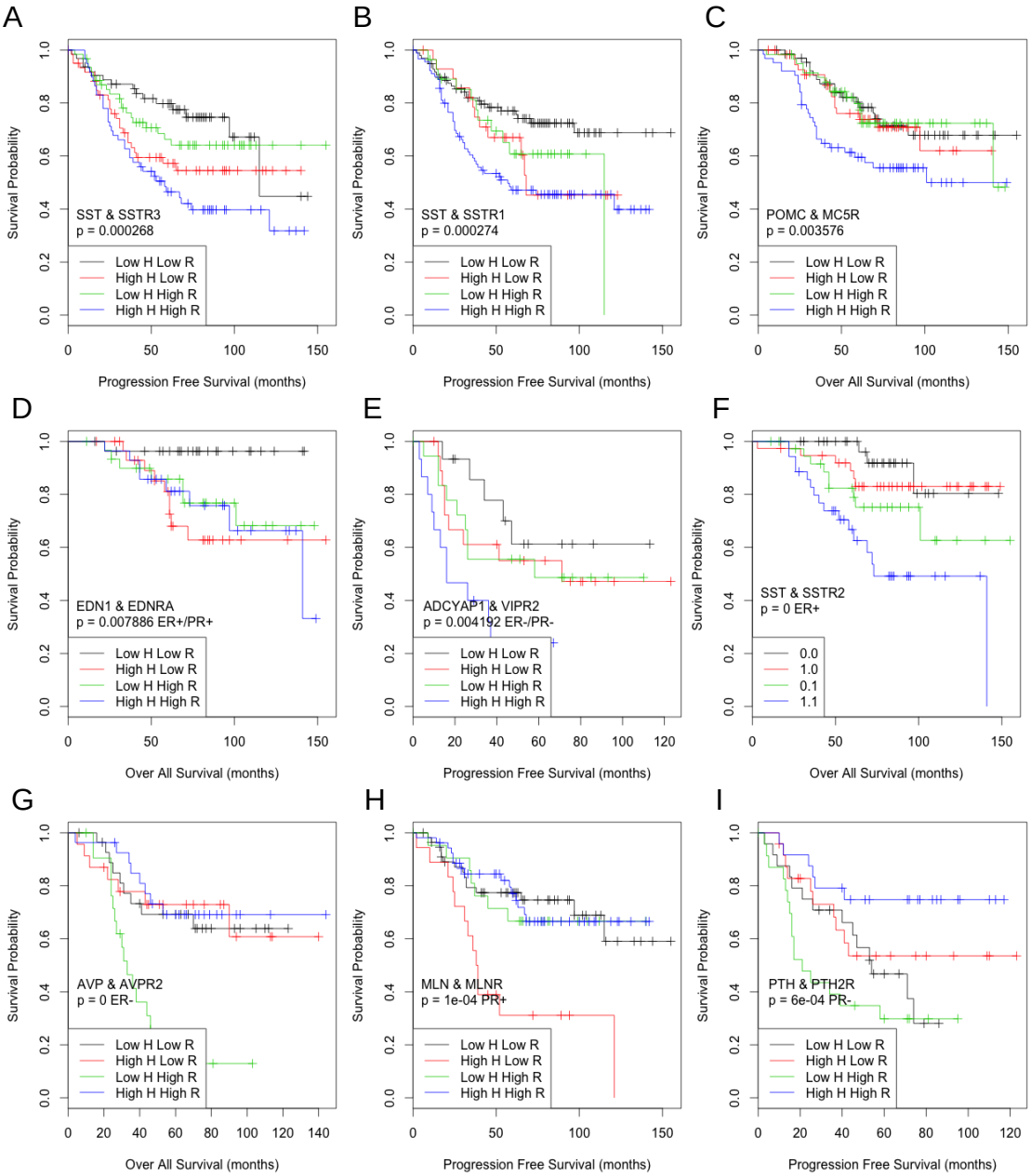

Figure 5. Representative Kaplan-Meier Plots Comparing Above and Below the Median Relative Gene Expression for Peptide/Protein Hormones and Their Cognate Receptors. Cancers with above the median expression for the hormone and the receptor have the worst prognosis (A (n=247), B (n=247), C (n=247), E (n=68), F (n=146)). ER+/PR+ cancers with below the median expression for both the hormone and receptor have the best prognosis. (D (n=118)) Cancers with below the median expression for the hormone and above the median expression for the receptor have the worst prognosis (G (n=101), I (n=96)). Cancers with above the median expression for the hormone and below the median expression for the receptor have the worst prognosis (H (n=151)).

A. OS ER+/PR+		B. PFS ER-/PR-		C. PFS ER-	
Gene Symbol	**β**	**Gene Symbol**	**β**	**Gene Symbol**	**β**
SSTR2	2.01	CALCR	0.17	CALCR	0.35
SSTR3	1.04	SSTR1	0.03	POMC	0.31
SSTR4	0.17	AGT	0.14	GH1	-0.40
IGF1	-0.31	AGTRAP	-0.09	LHB	0.05
INHBB	-0.59	POMC	0.31	SCT	0.13
GNRH2	0.39	GH1	-0.38	AVPR1A	0.23
PRLHR	-0.78	PRL	0.10		
PENK	-0.06	AVPR1A	0.12		
EPOR	0.12	AVPR2	0.21		
PRLR	0.42	PTH	-0.07		
CUL5	-0.41				
CALCA	1.27				
GAL	0.08				
NPY2R	0.34				
NPY5R	-0.03				
PTMS	0.19				

Table 24. Significant Genes and Corresponding Beta Coefficients for LASSO Performed on 142 Peptide/Protein Hormones and Their Cognate Receptors With and Without Receptor Status of Breast Carcinomas. ER+/PR+ cancers for OS (**A**), ER-/PR- cancers for PFS (**B**) and ER- cancers for PFS (**C**) were significant models from LASSO with a maximized likelihood from an L1 penalty and cross-validation.

Discussion

The expansion of our analysis with multivariate techniques intended to exhaust all the relationships hormones and receptors may have on each other. Two variable models proved to be no better at predicting PFS and OS than their respective univariate models. However, the interaction models gave us a myriad of questions to be answered, which is exactly what an exploratory based study should do.

Interaction models allowed us to explore the idea of endocrine autonomy where a cancer cell produces hormone proteins that bind to its own receptor proteins. This self-promoting loop could be associated with signaling pathways for proliferation. The idea that cancer cells have autonomous mechanisms to promote proliferation is not new (Hanahan & Weinberg, 2000). In addition, we know of a few peptide proteins which have been researched as biomarkers for early cancer detection (Assiri, Kamel, & Hassanien, 2015; Bae, Schaab, & Kratzsch, 2015). Yet to the best of our knowledge no one as looked at these peptide/protein hormones and their cognate receptors collectively.

What we have revealed with the novel findings from significant gene pairs for interaction are potential gene pairs to be used in a molecular model to predict cancer prognosis or potential targets for therapy. Throughout our exhaustive search through related literature, we found no studies that employed the use of statistical interaction to develop their molecular signatures to predict cancer behavior. For future studies we would like to investigate combinations of these significant genes for interaction to develop a unique molecular signature. Hormone therapy treatment with tamoxifen blocks the estrogen receptor site, thus disrupting the function of the receptor signaling inside the cell. If further investigation reveals a pathway for proliferation through one of our many hormone-receptor pairs, then specific targets could be developed to block those receptor sites.

The Kaplan-Meier plots offered us the best visual understanding of the relationship between hormones and their cognate receptors. The distinction drawn from cancers expressing high levels of hormones and receptor genes and those

expressing low levels of hormone and receptor genes suggests a strong dose-response for predicting PFS and OS (i.e., (A) SST and SSTR3 and (B) SST and SSTR1 in Figure 5). Significant Kaplan-Meier plots with a single group exhibiting a worse prognosis spawn exploratory ideas such as why do patients with cancers expressing high levels of the hormone motilin (MLN) and low levels of its cognate receptor MLNR have a significantly worse prognosis compared to the other three groups (Figure 5 (H)). The best way to embark on an investigation of one of these anomalies is to culminate univariate and multivariate Cox regressions, ER/PR boxplots and Kaplan-Meier plots for the hormone and receptor. Thus, these results serve as a platform for further investigation.

LASSO was performed on all 142 peptide/protein hormones and their cognate receptors to evaluate which genes were being expressed with the strongest signals for predict PFS and OS. The only non-zero models materialized from ER/PR subtypes. This suggests that the role of the sex hormones play a larger role in the behavior of peptide/protein hormones and receptors in predicting PFS and OS than trying to predict PFS and OS without considering the ER/PR receptor status. Receptor status is essential to determining the type of treatment a patient receives. Further examination of our findings may provide better prognosis for breast cancer patients across multiple biomarkers especially with the molecular signatures developed in Chapter IV.

CHAPTER IV

META-ANALYSIS AND MOLECULAR SIGNATURES

To strengthen our findings of specific genes whose expression predicts clinical behavior of breast cancer, we extended our analyses by examining our results with those reported in several public databases. The unique nature of the HRL dataset, having been developed using RNA extracted from Laser Capture Micro-dissected breast carcinoma cells, made direct comparisons to results of each of the four public databases examined difficult since none of them used LCM techniques in their studies. Therefore, it was reasoned that the meta-analysis would be conducted using various approaches of combining the HRL dataset with the public databases to determine the relationships of expression of genes for the 142 peptide/protein hormone and their cognate receptors related to clinical outcomes of breast carcinoma patients (Figure 1 Flow Diagram).

Methods and Material

Public Databases

One of the widely accepted practices is to externally validate the significant findings of our study using public datasets. Four public databases, TRANSBIG, VDX, MAINZ and UNT, were chosen based on the strength of their studies and their accessibility through Gene Omnibus Expression (GEO). Gene

expression values and clinical measurements were downloaded using the Bioconductor (version 3.2) packages 'breastCancerTRANSBIG', 'breastCancerVDX', 'breastCancerMAINZ' and 'breastCancerUNT'. Microarray data of relative gene expression levels determined with the AffymetrixU133a GeneChip were reported by TRANSBIG, VDX, MAINZ and UNT. TRANSBIG contained results from 198 lymph-node negative patients treated at five different centers. (Buyse et al., 2006) (Desmedt et al., 2007) VDX reported the relative gene expression results from 286 lymph-node negative patients with primary breast cancer. (Wang et al., 2005) The MAINZ database contained results from 200-node negative patients. (Schmidt et al., 2008) while UNT reported gene expression levels of 189 lymph-node negative patients. (Sotiriou et al., 2006) Each of the four public databases employed in these investigations contained the levels of estrogen receptor expressed in each cancer biopsy.

Prior to performing meta-analysis, the relative expression levels of each of the 142 candidate genes were normalized to a mean of zero and standard deviation of one within each of the five datasets. The relative gene expression results of the HRL database and those of the four public databases, in various combinations, were concatenated into individual datasets. No weight was given to any dataset nor to any subset of patients based on their characteristics (e.g., ER status of the cancer). Using each of the combined datasets, univariate Cox regression and multivariate Cox regression analyses were performed with gene expression results of each of the 142 candidate genes using only the clinical outcome PFS/RFS and the status of the biomarker, ER.

Molecular Signatures and Multivariable Cox Model

The field of 142 candidate genes was constricted to a smaller specialized group by a predefined criterion. In order for a gene to meet the criterion it must be significant ($p < 0.05$) for univariate Cox regression in the HRL data analysis and at least one other public database. This technique gives a higher preference for the HRL dataset, which is the only database to use LCM on its tissue biopsies.

These genes were selected as the special candidates for a molecular signature to be modeled with training and test sets using the meta-analyses dataset. Initially, the patients were placed into subsets based on their ER status since ER status is a strong predictor of clinical outcome. Patients were randomly selected into a 70% training set and the other 30% of patients were placed in a test set. Equal proportions of ER+ and ER- patients were selected to reflect the same proportion in the entire population. This randomization was independently repeated 1000 times.

Each training set was evaluated using Least Absolute Shrinkage and Selection Operator (LASSO) with an L1 penalty determined by the optL1 command in the *penalized* package. An L1 penalty shrinks the beta coefficients of non-significant genes to zero. The genes in each model were tallied after 1000 iterations. The number of gene occurrences in the training models was then compared to the maximum gene occurrence in permuted data sets. The permuted sets were constructed by randomly reordering the patient's clinical outcomes against their gene expression values. This approach established a baseline for the noise in the data as well as identified the genes with the strongest signals. Any of

the genes within the training set that contain a higher frequency than the maximum frequency in the permuted set were considered to be significant. (S. A. Andres et al., 2015)

Linear predictors were determined with the training set models and applied to the remaining 30% (test sets). A linear predictor is constructed for each patient by the summation of the relative gene expressions for each gene in the model multiplied by each gene's beta coefficient. The linear predictors were segregated into above/below the median subsets. Kaplan-Meier plots were constructed for the median splits using all models that had predictors in them. In order to visually compare these plots an alternative Kaplan-Meier plot was constructed which permuted the clinical outcomes (progression free survival, recurrence) against their corresponding linear predictors. These alternative plots should have no recognizable pattern between the median and tertiary split subsets. The distributions of the p-values from the non-zero Cox models, median splits, and tertiary subsets are shown in boxplots.

In order to externally validate the genes in our molecular signatures (MS), the four public databases were divided into 1000 training (70%) and test (30%) sets. The beta coefficients determined by Cox regression with the genes in each of the molecular signatures were applied to their corresponding test set. These linear predictors were evaluated against the clinical outcomes of the test set using the concordance index (C-index). Boxplots were employed on the 1000 C-indices to illustrate their distributions. The R command *rcorrcens* in R package *Hmisc*

calculated the C-index for each linear predictor and set of clinical outcomes (RFS for TRANSBIG and UNT and DMFS for VDX and MAINZ).

Results

Meta-analysis of HRL and Four Public Databases

Univariate Cox regression performed on 1126 breast cancer biopsies (HRL and four public databases combined) for each of the 142 candidate genes resulted in 12 genes exhibiting unadjusted p-values < 0.07. After p-values were adjusted for multiple comparisons using the Benjamin-Hochberg method, no genes met the 0.3 discovery threshold. Two limitations of this analysis are the clinical outcome being measured by MAINZ and VDX ignore local metastasis as an event and none of the public databases utilized LCM to avoid the convolution of surrounding non-cancerous tissues in their analyses. Despite limitations and loss of significance after controlling for the false discovery rate, RLN2, RLN1, VIPR1 and ACVR1B appear later in our 14-gene molecular signature for cancers without regard to receptor status.

Univariate Cox regression performed on 737 ER+ breast cancer biopsies for each of the 142 candidate genes resulted in eight genes exhibiting an association with PFS/DMFS and unadjusted p-values less than 0.04. Univariate Cox regression performed on 378 ER- breast cancer biopsies for each of the 142 candidate genes resulted in 13 genes exhibiting an association with PFS/DMFS and unadjusted p-values less than 0.07. After adjusting p-values for multiple comparisons, the expression of genes in neither ER+ nor ER- cancers exhibited

Gene Symbol	β	HR	95% CI (HR)	p	adj. p
RLN2	-0.14	0.87	(0.78,0.96)	0.01	0.69
GHR	0.12	1.13	(1.03,1.25)	0.01	0.69
FSHB	-0.12	0.89	(0.81,0.98)	0.02	0.69
VIPR2	-0.11	0.89	(0.81,0.99)	0.02	0.69
ADCYAP1	0.1	1.11	(1,1.23)	0.04	0.69
VIPR1	-0.1	0.91	(0.83,1)	0.05	0.69
TRH	-0.1	0.9	(0.81,1)	0.06	0.69
RLN1	-0.09	0.91	(0.83,1)	0.06	0.69
SSTR3	0.09	1.1	(1,1.21)	0.06	0.69
ACVR1B	0.1	1.1	(1,1.22)	0.06	0.69
TMSB10	0.09	1.1	(1,1.21)	0.06	0.69
THPO	0.09	1.1	(1,1.21)	0.06	0.69

Table 25. Summary of the Genes from Meta-analysis without Regard to Receptor Status (n=1126) with Expression Levels Associated with PFS/DMFS. As described in Methods and Materials, expression levels of 142 candidate genes were determined using LCM-procured cells from 247 breast cancer biopsies.

adjusted p-values less than the 0.3 threshold. The unadjusted univariate associated gene, GRP, for ER+ cancers appears in the 3-gene molecular signature for ER+ breast cancers while the unadjusted univariate associated genes, IGF1 and POMC, for ER- cancers appear in the 9-gene molecular signature for ER- breast cancers (Figure 7).

A. PFS/DMFS ER+

Gene Symbol	β	HR	95% CI (HR)	p	adj. p
RLN2	-0.19	0.82	(0.72,0.94)	0.01	0.45
THPO	0.15	1.16	(1.03,1.31)	0.01	0.45
PRLR	0.15	1.17	(1.02,1.33)	0.02	0.45
SST	0.15	1.16	(1.02,1.32)	0.02	0.45
C19orf80	-0.15	0.86	(0.76,0.98)	0.02	0.45
ACVR1B	0.14	1.15	(1.02,1.31)	0.03	0.45
GRP	-0.14	0.87	(0.77,0.98)	0.03	0.45
ATP6AP2	0.15	1.16	(1.02,1.33)	0.03	0.45

B. PFS/DMFS ER-

Gene Symbol	β	HR	95% CI (HR)	p	adj. p
IGF1	0.22	1.25	(1.06,1.46)	0.01	0.63
PTMS	0.22	1.25	(1.05,1.49)	0.01	0.63
AVPR2	0.18	1.2	(1.02,1.4)	0.02	0.63
GHR	0.16	1.18	(1.02,1.37)	0.03	0.63
POMC	0.15	1.16	(1.01,1.33)	0.03	0.63
TSHB	-0.19	0.83	(0.7,0.99)	0.04	0.63
VIPR2	-0.17	0.84	(0.71,0.99)	0.04	0.63
GALR3	-0.17	0.84	(0.71,0.99)	0.04	0.63
LEPR	0.16	1.17	(1,1.37)	0.05	0.63
RAMP2	0.15	1.16	(1,1.36)	0.05	0.63
C19orf80	0.15	1.16	(0.99,1.36)	0.06	0.63
INSR	0.14	1.14	(0.99,1.32)	0.06	0.63
CAP2	-0.15	0.86	(0.74,1.01)	0.06	0.63

Table 26. Summary of the Meta-analysis for (**A**) ER+ (n=737) and (**B**) ER- (n=378) Breast Cancers with Expression Levels Associated with PFS/DMFS. As described in Methods and Materials, expression levels of 142 candidate genes were determined using LCM-procured cells from 247 breast cancer biopsies.

Kaplan-Meier Plots Demonstrating Results from Meta-Analysis

The Kaplan-Meier plots in Figure 6 compare the difference between breast cancers reported in the public databases and in the HRL database expressing a gene above and below the median. The three most significant genes are shown for the 1126 cancers without regard to receptor status, 737 ER+ cancers and 378 ER- cancers (11 cancers had missing data on ER status). For breast cancers without regard to receptor status, patients with cancers that expressed ADCYAP1 and SST above the median and VIPR2 below the median had the worst prognosis. Noteworthy, SST and VIPR2 appear in the 14-gene molecular signature (Figure 7). Patients with ER+ breast cancers expressing SST and SSTR2 above the median and RAMP2 below the median tend to have a shorter time for a recurrence of cancer. Interestingly, SSTR2 also appears in the 3 –gene molecular signature. Patients with ER- breast cancers expressing C19orf80 above the median and TRH and FSHB below the median exhibited the worst prognosis.

Gene Signatures Predicting Clinical Behavior of Breast Carcinomas

Significant genes from univariate Cox regression, either PFS or OS, were analyzed for their ability to predict risk of breast cancer recurrence as a set with 1000 LASSO training/test sets. The genes that occurred in more models than in the permuted set, which serves as the null distribution, are considered gene signature candidates until externally compared with results in public databases, as described in Methods and Materials. A 14-gene subset was identified as a candidate molecular signature for predicting clinical outcome without regard to the sex hormone receptor status of the breast cancer biopsies (Figure 7). When sex-

hormone receptor status was considered, a three-gene signature composed of NPY1R, SSTR2 and GRP was deciphered for ER+ breast cancer biopsies and a candidate 9-gene signature was identified for ER- breast cancer biopsies.

The 14-gene signature was composed of expression patterns for eight receptor genes and six hormone ligand genes. Of these 14 genes, only NPY1R was also identified in the 3-gene signature in ER+ breast cancers (Figure 7). Interestingly, three of the genes, POMC, CALCR and PRL, of the signature deciphered without regard for ER status were also found in the 9-gene signature identified in ER- breast cancers. Surprisingly, in contrast to the composition of the 14-gene signature, only two of the 9 genes of the molecular signature associated with ER- breast cancer direct the synthesis of hormone receptors. No genes reflecting peptide/protein hormone-receptor pairs were detected in the 14-gene signature nor in the 3-gene signature. It is noteworthy, that the gene for the hormone prolactin (PRL), which plays a central role in breast physiology, and prolactin's cognate receptor PRLR appeared in the 9-gene signature for ER- biopsies.

Concordance indices evaluated the predictive ability of the genes in each of the three molecular signatures on each of the four public databases. Figure 8 summarizes the distribution of the C-indices determined by 1000 training and test sets. The 9-gene MS performed better than the 14-gene and 3-gene MS in VDX, UNT and MAINZ and as well as the 3-gene model in TRANSBIG. The median C-index for the 9-gene model was greater than 0.6 for two public databases, MAINZ

and UNT. Excitingly, the best predictive MS is in the 9-gene model for ER- cancers, since patients with ER- cancers have very limited treatment options.

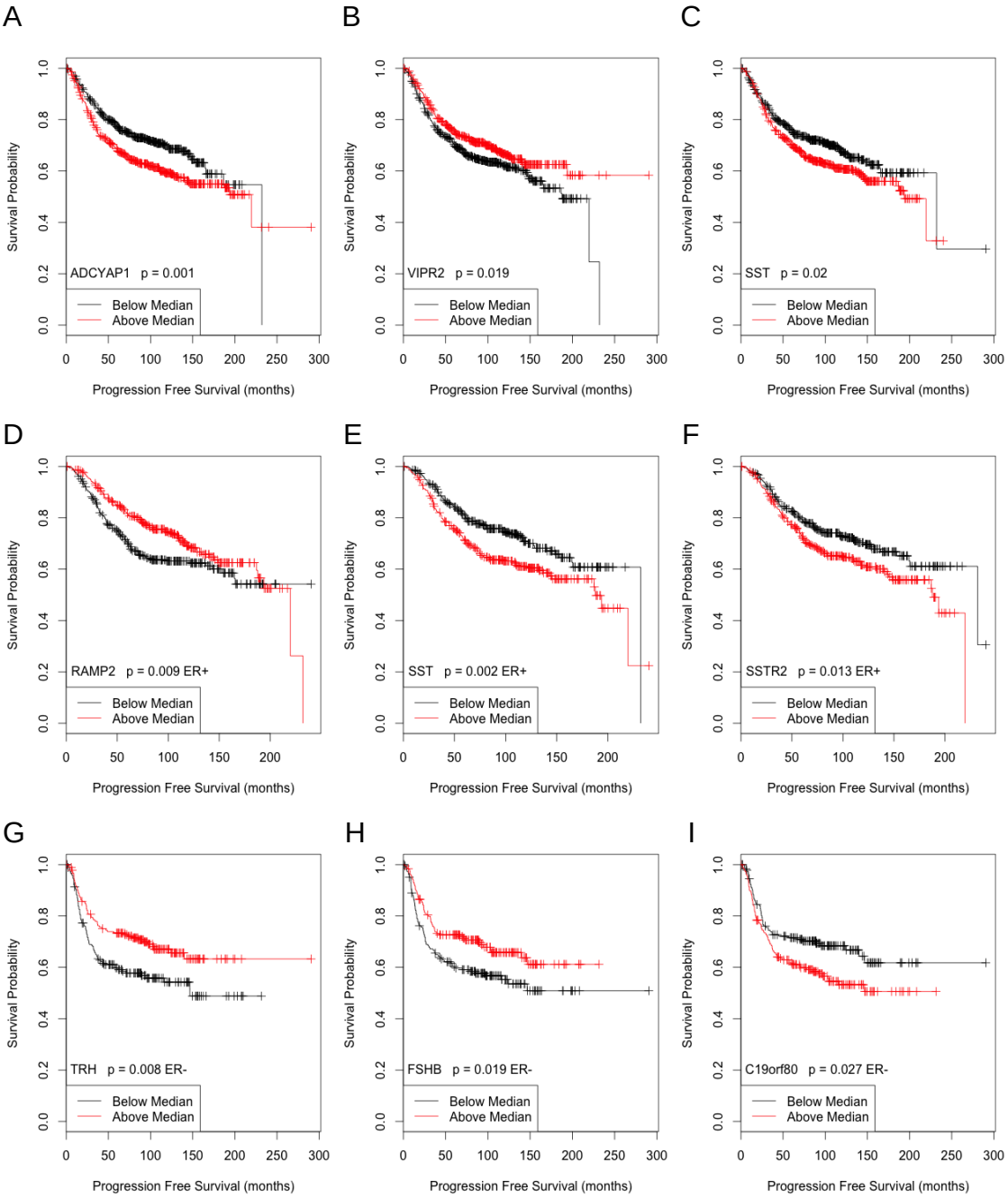

Figure 6. Representative Kaplan-Meier Plots Comparing Above and Below the Median Relative Gene Expression of 142 Peptide/Protein Hormones or Their Cognate Receptors with and without Regard to ER Status for PFS from the Meta-analysis. Without regard to ER status, 1126 breast carcinomas expressing ADCYAP1 (**A**) and SST (**C**) above the median and VIPR2 (**B**) below the median had the worst prognosis. 737 ER+ breast carcinomas expressing RAMP2 (**D**) below the median and SST (**E**) and SSTR2 (**F**) above the median had the worst prognosis. 378 ER- breast carcinomas expressing TRH (**G**) and FSHB (**H**) below the median and C19orf80 (**I**) above the median had the worst prognosis.

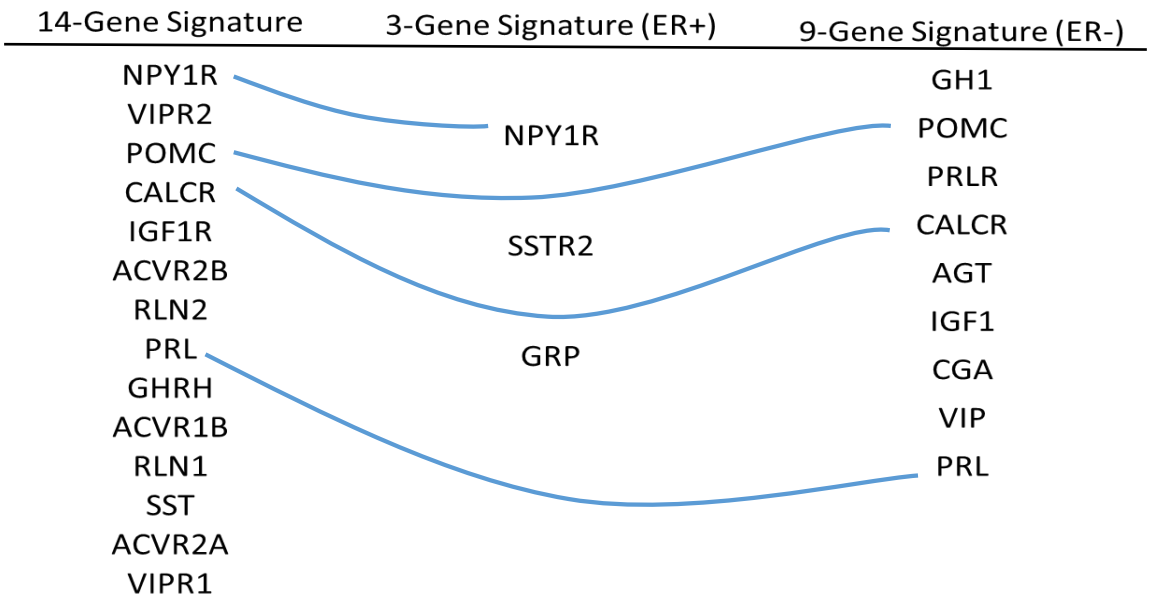

Figure 7. Composition of Molecular Signatures Derived Showing Genes Common to Each

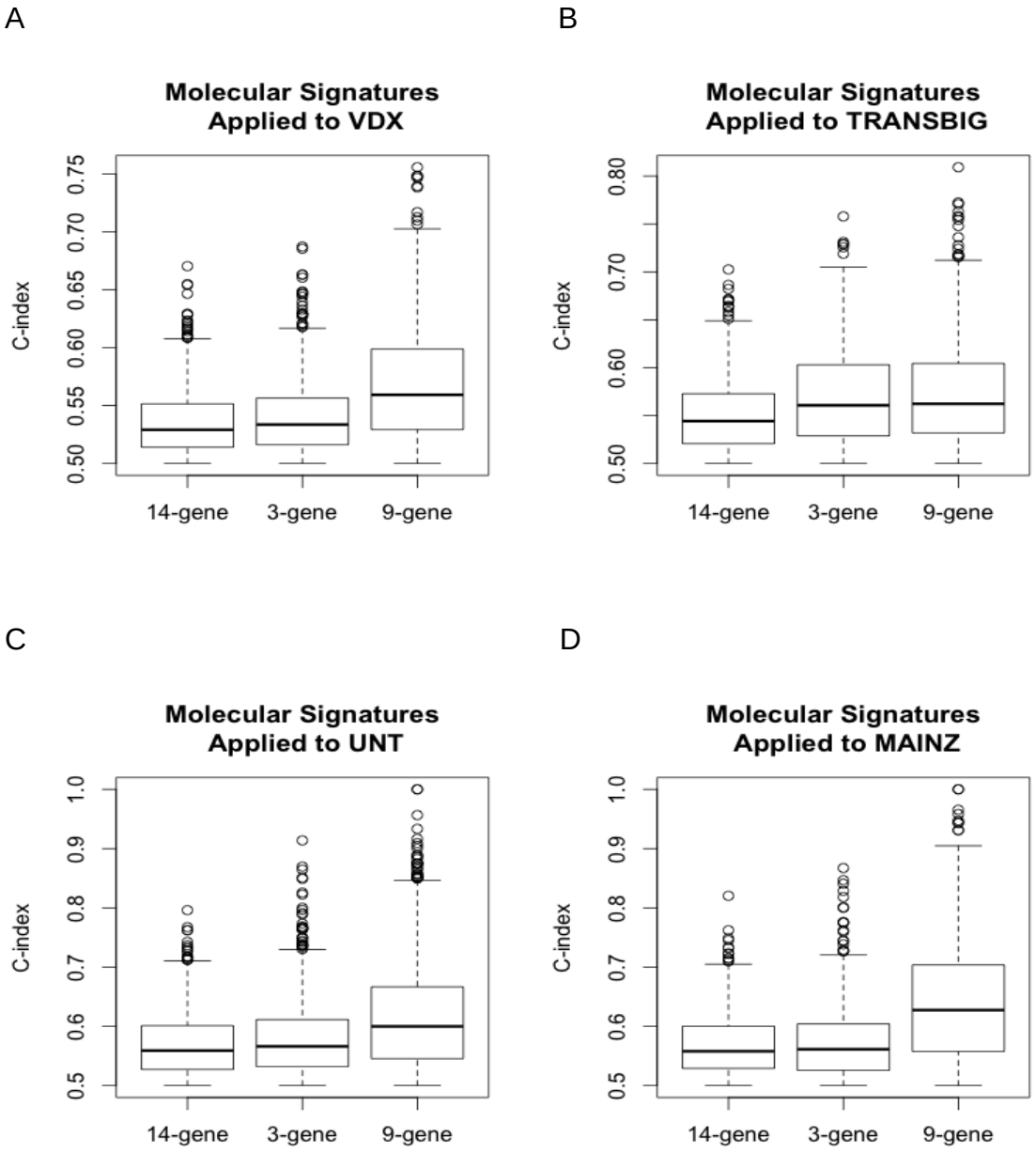

Figure 8. Summary of 1000 C-indices of Each of the Three Molecular Signatures on Each of the Four Public Databases, (**A**) VDX, (**B**) TRANSBIG, (**C**) UNT and (**D**) MAINZ. The 14-Gene MS was applied to cancers without regard to receptor status. The 3-Gene MS and 9-Gene MS was applied to ER+ and ER- cancers, respectively. Validation of the genes in the three MS was performed by dividing public databases into 70% training and 30% test sets.

Discussion

As expressed earlier, our overarching goal is to determine the relationships of the expression of the genes for each of the peptide/protein hormones and that of their cognate receptors with clinical outcomes of breast cancer patients. Our hypothesis is that expression profiles of subsets of these genes for regulatory molecules may be used to predict risk of breast carcinoma recurrence in patients with primary disease. It appears that our approach described in Figure 1 is original considering the following resources: 1) only de-identified breast cancer tissues were processed by LCM for microarray, 2) gene expression levels reflected only mRNA of specific cell types, 3) 22,000 genes were determined in 247 primary breast carcinomas in a standardized fashion and 4) data were complemented by quantitative results of protein biomarker levels (ER, PR and HER-2). These unparalleled properties of each specimen were accompanied by clinical follow-up and patient outcome. A multitude of statistical analyses such as univariate and multivariate Cox regressions, Kaplan-Meier plots, boxplots and LASSO were utilized to predict a patient's clinical outcome.

Recall from Chapter II the importance POMC, GH1 and CALCR contributed to predicting PFS and OS for univariate Cox regression and Kaplan-Meier plots for ER- subtypes of breast carcinomas. SST and its cognate receptors diverge expression towards cancers exuding estrogen receptor proteins (ER+). POMC is the standout gene from univariate Cox regression with and without regard to receptor status while GH1 in ER- cancers displayed an exceptional differentiation between cancers expressing the gene at high and low levels. POMC has appeared

in many studies showing an association with a variety of diseases such as obesity and cancer (Clark, 2015; Mountjoy, 2015). Other studies have shown polymorphisms of GH1 to have a protective influence on breast cancer risk (Wagner et al., 2006). Although we did not examine polymorphism, we did independently show that elevated levels GH1 in breast carcinomas have a protective effect. These findings warrant further the investigation of these two genes with the HRL's Next Generation Sequencing (NGS) database.

Chapter III explored a myriad of multivariate predictors for progression of disease and survivorship of patients with primary breast cancers. A number of hormone-receptors exhibited a significant statistical interaction such as the pairs HCRT-HCRTR2 in ER- and PR- cancers, IAPP-CALCR in ER- cancers and FSHB-FSHR in ER+/PR+ cancers. Two variable models did not distinguish any better predictors for PFS and OS than univariate Cox regression results contributed. LASSO revealed the prominence of the role that sex hormone status performs in predicting PFS and OS for peptide/protein hormones and their receptors. The significant LASSO results of 16 genes for ER+/PR+ cancers, 10 genes for ER-/PR- cancers and 6 genes for ER- will be externally validated with the public databases as alternative molecular signatures in future studies.

We learned from Chapter IV that the significance of many genes in the meta-analysis evaporated when adjusted for multiple comparison were calculated. This loss in the number of significant genes happened throughout the study. The challenge of exploring the relative expression levels of 142 genes is not just the volume of work to be organized but discretizing statistical significance from the

inherent noise of highly correlated data found in microarray. Our original answer to the question of how to validate our discoveries since most of them could not overcome the loss of significance from adjusting for a false discovery rate was the use of public databases. The challenge was the lack of public databases that used LCM to ensure only cancer cells were being evaluated and the uncertainty of their consistency used in specimen collection and lab techniques.

Despite these challenges, our 9-gene molecular signature for ER- cancers performed better than the 14 gene and 3-gene MS in predicting the PFS/DMFS in all four public databases. The median concordance index of two public databases was over 60%. In summary, the resources and approach clearly support the integrity of the molecular signatures primary breast carcinoma to predict risk of recurrence. Additional steps in the clinical validation of these signatures may include the implementation of a clinical trial whereby the molecular signatures are used as biomarkers with and without regard to sex hormone receptor status of the breast carcinoma biopsy.

REFERENCES

Alldinger, I., Dittert, D., Peiper, M., Fusco, A., Chiappetta, G., Staub, E., . . . Ockert, D. (2005). Gene expression analysis of pancreatic cell lines reveals genes overexpressed in pancreatic cancer. *Pancreatology, 5*(4-5), 370-379.

Andres, S. A., Bickett, K. E., Alatoum, M. A., Kalbfleisch, T. S., Brock, G. N., & Wittliff, J. L. (2015). Interaction between smoking history and gene expression levels impacts survival of breast cancer patients. *Breast cancer research and treatment, 152*(3), 545-556.

Andres, S. A., Bickett, K. E., Alatoum, M. A., Kalbfleisch, T. S., Brock, G. N., & Wittliff, J. L. (2015). Interaction between smoking history and gene expression levels impacts survival of breast cancer patients. *Breast Cancer Res Treat, 152*(3), 545-556. doi:10.1007/s10549-015-3507-z

Andres, S. A., Brock, G. N., & Wittliff, J. L. (2013). Interrogating differences in expression of targeted gene sets to predict breast cancer outcome. *BMC Cancer, 13*(1), 1.

Andres, S. A., Edwards, A. B., & Wittliff, J. L. (2012). Expression of Urokinase-Type Plasminogen Activator (uPA), its Receptor (uPAR), and Inhibitor (PAI-1) in Human Breast Carcinomas and Their Clinical Relevance. *Journal of clinical laboratory analysis, 26*(2), 93-103.

Andres, S. A., & Wittliff, J. L. (2011). Relationships of ESR1 and XBP1 expression in human breast carcinoma and stromal cells isolated by laser capture microdissection compared to intact breast cancer tissue. *Endocrine, 40*(2), 212-221.

Andres, S. A., & Wittliff, J. L. (2012). Co-expression of genes with estrogen receptor-α and progesterone receptor in human breast carcinoma tissue. *Hormone molecular biology and clinical investigation, 12*(1), 377-390.

Assiri, A., Kamel, H. F., & Hassanien, M. F. (2015). Resistin, visfatin, adiponectin, and leptin: risk of breast cancer in pre-and postmenopausal saudi females and their possible diagnostic and predictive implications as novel biomarkers. *Disease markers, 2015*.

Austin, P. C., & Tu, J. V. (2004). Automated variable selection methods for logistic regression produced unstable models for predicting acute myocardial infarction mortality. *Journal of clinical epidemiology, 57*(11), 1138-1146.

Bae, Y. J., Schaab, M., & Kratzsch, J. (2015). Calcitonin as Biomarker for the Medullary Thyroid Carcinoma *Medullary Thyroid Carcinoma* (pp. 117-137): Springer.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 289-300.

Bradburn, M. J., Clark, T. G., Love, S. B., & Altman, D. G. (2003). Survival Analysis Part II: Multivariate data analysis – an introduction to concepts and methods. *British Journal of Cancer, 89*(3), 431-436. doi:10.1038/sj.bjc.6601119

Burnatowska-Hledin, M. A., Kossoris, J. B., Van Dort, C. J., Shearer, R. L., Zhao, P., Murrey, D. A., . . . Barney, C. C. (2004). T47D breast cancer cell growth is inhibited by expression of VACM-1, a cul-5 gene. *Biochemical and biophysical research communications, 319*(3), 817-825.

Buyse, M., Loi, S., van't Veer, L., Viale, G., Delorenzi, M., Glas, A. M., . . . Consortium, T. (2006). Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. *J Natl Cancer Inst, 98*(17), 1183-1192. doi:10.1093/jnci/djj329

Chen, A., Kaganovsky, E., Rahimipour, S., Ben-Aroya, N., Okon, E., & Koch, Y. (2002). Two Forms of Gonadotropin-releasing Hormone (GnRH) Are Expressed in Human Breast Tissue and Overexpressed in Breast Cancer A Putative Mechanism for the Antiproliferative Effect of GnRH by Down-Regulation of Acidic Ribosomal Phosphoproteins P1 and P2. *Cancer Research, 62*(4), 1036-1044.

Clark, A. J. (2015). 60 YEARS OF POMC: The proopiomelanocortin gene: discovery, deletion and disease. *Journal of molecular endocrinology*, JME-15-0268.

Córdoba-Chacón, J., Gahete, M. D., Pozo-Salas, A. I., Martínez-Fuentes, A. J., de Lecea, L., Gracia-Navarro, F., . . . Luque, R. M. (2011). Cortistatin is not a somatostatin analogue but stimulates prolactin release and inhibits GH and ACTH in a gender-dependent fashion: potential role of ghrelin. *Endocrinology, 152*(12), 4800-4812.

Cormier, E. M., Wolf, M. F., & Jordan, V. C. (1989). Decrease in estradiol-stimulated progesterone receptor production in MCF-7 cells by epidermal growth factor and possible clinical implication for paracrine-regulated breast cancer growth. *Cancer Res, 49*(3), 576-580.

Desmedt, C., Piette, F., Loi, S., Wang, Y., Lallemand, F., Haibe-Kains, B., . . . Consortium, T. (2007). Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. *Clin Cancer Res, 13*(11), 3207-3214. doi:10.1158/1078-0432.CCR-06-2765

Falzon, M., & Du, P. (2000). Enhanced Growth of MCF-7 Breast Cancer Cells Overexpressing Parathyroid Hormone-Related Peptide 1. *Endocrinology, 141*(5), 1882-1892.

Fisher, B., Brown, A., Wolmark, N., REDMOND, C., Wickerham, D. L., Wittliff, J., . . . PRAGER, D. (1987). Prolonging tamoxifen therapy for primary breast cancer: Findings from the National Surgical Adjuvant Breast and Bowel Project clinical trial. *Annals of internal medicine, 106*(5), 649-654.

Fisher, B., Dignam, J., Tan-Chiu, E., Anderson, S., Fisher, E. R., Wittliff, J. L., & Wolmark, N. (2001). Prognosis and treatment of patients with breast tumors of one centimeter or less and negative axillary lymph nodes. *Journal of the National Cancer Institute, 93*(2), 112-120.

Fisher, B., Redmond, C., Brown, A., Wickerham, D., Wolmark, N., Allegra, J., . . . Wittliff, J. (1983). Influence of tumor estrogen and progesterone receptor levels on the response to tamoxifen and chemotherapy in primary breast cancer. *Journal of Clinical Oncology, 1*(4), 227-241.

Fisher, B., Redmond, C., Brown, A., Wolmark, N., Wittliff, J., Fisher, E. R., . . . Wolter, J. (1981). Treatment of primary breast cancer with chemotherapy and tamoxifen. *New England Journal of Medicine, 305*(1), 1-6.

Fleisher, M., Dnistrian, A. M., Sturgeon, C. M., Lamerz, R., & Wittliff, J. L. (2002). Practice guidelines and recommendations for use of tumor markers in the clinic. *Tumor markers: physiology, pathobiology, technology and clinical applications*, 33-63.

Gingras, I., Desmedt, C., Ignatiadis, M., & Sotiriou, C. (2015). CCR 20th Anniversary Commentary: Gene-Expression Signature in Breast Cancer—Where Did It Start and Where Are We Now? *Clinical Cancer Research, 21*(21), 4743-4746.

Ginsburg, E., & Vonderhaar, B. K. (1995). Prolactin synthesis and secretion by human breast cancer cells. *Cancer Research, 55*(12), 2591-2595.

Goeman, J., Meijer, R., & Chaturvedi, N. (2014). L1 and L2 penalized regression models: R.

Goeman, J. J. (2010). L1 penalized estimation in the Cox proportional hazards model. *Biometrical journal, 52*(1), 70-84.

Greenspan, F. S., Gardner, D. G., & Shoback, D. (1997). *Basic & clinical endocrinology*: Appleton & Lange Stamford, CT.

Hammond, M. E. H., Hayes, D. F., Dowsett, M., Allred, D. C., Hagerty, K. L., Badve, S., . . . Hayes, M. (2010). American Society of Clinical Oncology/College of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer (unabridged version). *Archives of pathology & laboratory medicine, 134*(7), e48-e72.

Hanahan, D., & Weinberg, R. A. (2000). The hallmarks of cancer. *cell, 100*(1), 57-70.

Hanukoglu, I. (1992). Steroidogenic enzymes: structure, function, and role in regulation of steroid hormone biosynthesis. *J Steroid Biochem Mol Biol, 43*(8), 779-804. doi:10.1016/0960-0760(92)90307-5

Hong, Y., & Chen, S. (2011). Aromatase, estrone sulfatase, and 17β-hydroxysteroid dehydrogenase: Structure–function studies and inhibitor development. *Molecular and cellular endocrinology, 340*(2), 120-126.

Hoyer, D., Bell, G. I., Berelowitz, M., Epelbaum, J., Feniuk, W., Humphrey, P. P., . . . et al. (1995). Classification and nomenclature of somatostatin receptors. *Trends Pharmacol Sci, 16*(3), 86-88.

James, L. W. (1984). Steroid hormone receptors in breast cancer. *Cancer, 532*(2), 630.

Journal of receptor and signal transduction research. (1995). http://www.informaworld.com/openurl?genre=journal&issn=1079%2d9893

http://www.metapress.com/link.asp?id=108174

http://www.metapress.com/link.asp?id=113324

http://firstsearch.oclc.org

http://firstsearch.oclc.org/journal=1079-9893;screen=info;ECOIP Retrieved from http://www.informaworld.com/openurl?genre=journal&issn=1079%2d9893

http://www.metapress.com/link.asp?id=108174

http://www.metapress.com/link.asp?id=113324

http://firstsearch.oclc.org

http://firstsearch.oclc.org/journal=1079-9893;screen=info;ECOIP

Jung, S., Yi, L., Jeong, D., Kim, J., An, S., Oh, T.-J., . . . Kim, K. I. (2011). The role of ADCYAP1, adenylate cyclase activating polypeptide 1, as a methylation biomarker for the early detection of cervical cancer. *Oncology reports, 25*(1), 245-252.

Kaushal, S., & Sinha, M. (2015). WORLD JOURNAL OF PHARMACEUTICAL RESEARCH.

Kerr II, D. A., & Wittliff, J. L. (2011). A five-gene model predicts clinical outcome in ER+/PR+, early-stage breast cancers treated with adjuvant tamoxifen. *Hormones and Cancer, 2*(5), 261-271.

Kidd, L. R., Brock, G. N., VanCleave, T. T., Benford, M. L., Lavender, N. A., Kruer, T. L., & Wittliff, J. L. (2010). Angiogenesis-associated sequence variants relative to breast cancer recurrence and survival. *Cancer Causes & Control, 21*(10), 1545-1557.

Klein, J. P., & Moeschberger, M. L. (1997). Survival analysis techniques for censored and truncated data. Retrieved from http://site.ebrary.com/id/10005871

Klein, J. P., & Moeschberger, M. L. (2003). *Survival analysis : techniques for censored and truncated data* (2nd ed.). New York: Springer.

Kruer, T. L., Cummins, T. D., Powell, D. W., & Wittliff, J. L. (2013). Characterization of estrogen response element binding proteins as biomarkers of breast cancer behavior. *Clinical biochemistry, 46*(16), 1739-1746.

Langevin, S. M., Kratzke, R. A., & Kelsey, K. T. (2015). Epigenetics of lung cancer. *Translational Research, 165*(1), 74-90.

Lee, Y. S., Challis, B. G., Thompson, D. A., Yeo, G. S., Keogh, J. M., Madonna, M. E., . . . Farooqi, I. S. (2006). A POMC variant implicates beta-melanocyte-stimulating hormone in the control of human energy balance. *Cell Metab, 3*(2), 135-140. doi:10.1016/j.cmet.2006.01.006

Ma, X., Wang, W., Salunga, R., Tuggle, T., Stecker, K., Baer, T., . . . Wittliff, J. (2003). *Gene expression signatures associated with clinical outcome in breast cancer via laser capture microdissection.* Paper presented at the Breast cancer research and treatment.

McShane, L. M., Altman, D. G., Sauerbrei, W., Taube, S. E., Gion, M., Clark, G. M., & Statistics Subcommittee of, N. C. I. E. W. G. o. C. D. (2006). REporting recommendations for tumor MARKer prognostic studies (REMARK). *Breast Cancer Res Treat, 100*(2), 229-235. doi:10.1007/s10549-006-9242-8

Metzler, M. A., Venkatesh, S. G., Lakshmanan, J., Carenbauer, A. L., Perez, S. M., Andres, S. A., . . . Darling, D. S. (2015). A systems biology approach identifies a regulatory network in parotid acinar cell terminal differentiation. *PLoS One, 10*(4), e0125153.

Mountjoy, K. (2015). Pro-Opiomelanocortin (POMC) Neurones, POMC-Derived Peptides, Melanocortin Receptors and Obesity: How Understanding of this System has Changed Over the Last Decade. *Journal of neuroendocrinology, 27*(6), 406-418.

Norman, A. W., & Litwack, G. (1997). *Hormones*: Academic Press.

Pierce, J. G. (1982). *Protein and peptide hormones*. Stroudsburg, Pa. :: Hutchinson Ross Pub. Co.

Raloxifene Hydrochloride. Retrieved from http://www.cancer.gov/about-cancer/treatment/drugs/raloxifenehydrochloride

Rich, J. T., Neely, J. G., Paniello, R. C., Voelker, C. C., Nussenbaum, B., & Wang, E. W. (2010). A practical guide to understanding Kaplan-Meier curves. *Otolaryngol Head Neck Surg, 143*(3), 331-336. doi:10.1016/j.otohns.2010.05.007

Rodrigues-Ferreira, S., & Nahmias, C. (2015). G-protein coupled receptors of the renin-angiotensin system: new targets against breast cancer? *Frontiers in pharmacology, 6*.

Sachdev, D. (2008). Regulation of breast cancer metastasis by IGF signaling. *Journal of mammary gland biology and neoplasia, 13*(4), 431-441.

Schmidt, M., Bohm, D., von Torne, C., Steiner, E., Puhl, A., Pilch, H., . . . Gehrmann, M. (2008). The humoral immune system has a key prognostic impact in node-negative breast cancer. *Cancer Res, 68*(13), 5405-5413. doi:10.1158/0008-5472.CAN-07-5206

Sotiriou, C., Wirapati, P., Loi, S., Harris, A., Fox, S., Smeds, J., . . . Delorenzi, M. (2006). Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J Natl Cancer Inst, 98*(4), 262-272. doi:10.1093/jnci/djj052

Surveilance, Epidemiology, and End Results Program. Retrieved from http://seer.cancer.gov/statfacts/html/breast.html

Tamoxifen Citrate. Retrieved from http://www.cancer.gov/about-cancer/treatment/drugs/tamoxifencitrate

Tecimer, C., Doering, D., Goldsmith, L., Meyer, J., Abdulhay, G., & Wittliff, J. (2000). Clinical relevance of urokinase-type plasminogen activator, its receptor and inhibitor type 1 in ovarian cancer. *International Journal of Gynecological Cancer, 10*(5), 372-381.

Therneau, T. (2013). A package for survival analysis in S. R package version 2.37-4. *URL http://CRAN. R-project. org/package= survival. Box, 980032*, 23298-20032.

Tikk, K., Sookthai, D., Fortner, R. T., Johnson, T., Rinaldi, S., Romieu, I., . . . Clavel-Chapelon, F. (2015). Circulating prolactin and in situ breast cancer risk in the European EPIC cohort: a case-control study. *Breast Cancer Res, 17*, 49.

Toremifene. Retrieved from http://www.cancer.gov/about-cancer/treatment/drugs/toremifene

Van Dort, C., Zhao, P., Parmelee, K., Capps, B., Poel, A., Listenberger, L., . . . Clare, P. (2003). VACM-1, a cul-5 gene, inhibits cellular growth by a mechanism that involves MAPK and p53 signaling pathways. *American Journal of Physiology-Cell Physiology, 285*(6), C1386-C1396.

Wagner, K., Hemminki, K., Grzybowska, E., Klaes, R., Burwinkel, B., Bugert, P., . . . Pamula, J. (2006). Polymorphisms in genes involved in GH1 release and their association with breast cancer risk. *Carcinogenesis, 27*(9), 1867-1875.

Wang, Y., Klijn, J. G., Zhang, Y., Sieuwerts, A. M., Look, M. P., Yang, F., . . . Foekens, J. A. (2005). Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet, 365*(9460), 671-679. doi:10.1016/S0140-6736(05)17947-1

Wentzensen, N., Bakkum-Gamez, J. N., Killian, J. K., Sampson, J., Guido, R., Glass, A., . . . Rush, B. (2014). Discovery and validation of methylation markers for endometrial cancer. *International Journal of Cancer, 135*(8), 1860-1868.

Wittliff, J., Ma, X., Stecker, K., Salunga, R., Tuggle, J., Tran, Y., . . . Pistone, M. (2002). Gene expression profiles and tumor marker signatures of human breast carcinoma cells procured by laser capture microdissection. *Endocrine Soc Abs, 3*, 538.

Wittliff, J., Pasic, R., & Bland, K. (1998). Steroid and peptide hormone receptors: methods, quality control and clinical use. *The Breast: Comprehensive Management of Benign and Malignant Diseases. Philadelphia, PA, WB Saunders Co*, 458-498.

Wittliff, J. L. (2010). Laser Capture Microdissection and Its Use in Genomics & Proteomics, in Reliable Lab Solutions. In P. M. Conn (Ed.), *Techniques in Confocal Microscopy* (pp. 463-477): Elsevier Press.

Wu, M., & Singh, A. K. (2012). Single-cell protein analysis. *Current opinion in biotechnology, 23*(1), 83-88.

Yang, Y., & Yee, D. (2012). Targeting insulin and insulin-like growth factor signaling in breast cancer. *Journal of mammary gland biology and neoplasia, 17*(3-4), 251-261.

CURRICULUM VITAE

Michael W. Daniels

911 Mallard Creek Road

Louisville, KY 40207

(502) 693-4229 (cell)

(502) 852-5694 (HRL Office)

daniels.mike1220@gmail.com

Current Positions

Graduate Program, University of Louisville

Master of Science Biostatistics

Department of Public Health and Information Sciences

Graduate Research Assistant, University of Louisville

Department of Biochemistry & Molecular Genetics

Hormone Receptor Laboratory

- Commonwealth of Kentucky (State License #200078

Endocrinology & Chemistry)

- Health Care Financing Administration (Federal License – CLIA #18D0648476, Endocrinology)

Student Member, Institute for Molecular Diversity & Drug Design (IMD3), University of Louisville

Donor Center Coordinator, Kentucky Organ Donor Affiliates

Education

Master of Science Biostatistics

Aug. 2014 – Current (expected May 2016)

University of Louisville

Louisville, KY

GPA 3.85

Bachelor of Science Mathematics

Concentration: Pure Mathematics

January 2012 - May 2014

University of Louisville

Louisville, KY

GPA 3.25

Honors and Awards

Travel Award, IMD3 – Partial funding to attend the San Antonio Breast Cancer Symposium and co-present two posters.

Travel Award, School of Public Health – Partial funding to attend the San Antonio Breast Cancer Symposium and co-present two posters.

Work Experience

Feb. 2015 - Current

Graduate Research Assistant, Hormone Receptor Laboratory

- Use of various statistical software packages including R to analyze complex genomics and proteomics results in a comprehensive database of biochemical and clinical studies of human cancer.
- Development of presentation formats for meetings, grant applications and reports
- Supervise statistical studies of two undergraduate research fellows.

2011 - Current

Donor Center Coordinator, Kentucky Organ Donor Affiliates

- Accurately screen potential donors for medical suitability.
- Counsel potential donor families regarding their donation options.

- Obtain consent for appropriate organs and/or tissues and conduct medical/social history interviews with families.

National & International Research Meeting - Abstracts

Wittliff JL, **Daniels MW**, Brock GN. Expression of Genes for Peptide/Protein Hormones and Their Receptors in Breast Carcinomas as Biomarkers Predicting Risk of Recurrence, 2015 San Antonio Breast Cancer Symposium (SABCS), in press.

Sanders MA, **Daniels MW**, Wittliff JL. Expression of genes for aromatase inhibitor targets to discriminate invasive lobular from invasive ductal carcinomas of the breast using LCM-procured cells to complement endocrine biomarkers, 2015 San Antonio Breast Cancer Symposium (SABCS), in press.

Mentored Abstracts

Hameed ZR, Sereff SB, Wittliff JL, **Daniels MW**. Deciphering the Molecular Basis of Breast Cancer Behavior using Proteomics and Genomics, Posters-at-the-Capitol, in press.

Professional Organizations

American Statistical Association, member.

American Association for Cancer Research, member.

Biostatistics Club, active participant.

Military

Served nine years in Virginia National Guard and US Army Reserves as an enlisted infantryman.

Competed at a national M60 competition for all branches of US military service. (1st place in 2-mile run with 800 meter shot)

Mobilized for Operation Noble Eagle in 2002.