January 2014

Genomic Predictor Of Residual Risk Of Recurrence After Chemotherapy In High Risk Estrogen Receptor Positive Breast Cancers

Sabrina Khan
Yale School of Medicine, sabrina.khan@yale.edu

GENOMIC PREDICTOR OF RESIDUAL RISK OF RECURRENCE AFTER CHEMOTHERAPY IN HIGH RISK ESTROGEN RECEPTOR POSITIVE BREAST CANCERS

A Thesis Submitted to the

Yale University School of Medicine

in Partial Fulfillment of the Requirements for the

Degree of Doctor of Medicine

by

Sabrina Khan, MPH

2014

ACKNOWLEDGEMENTS

Thank you to Drs. Christos Hatzis and Lajos Pusztai for all your teaching and guidance in helping me develop and complete this thesis. I would also like to acknowledge Dr. Thomas Karn for collaborating on this project. Finally, I would like to thank the Office of Student Research at Yale School of Medicine for grant funding support.

TABLE OF CONTENTS

GENOMIC PREDICTOR OF RESIDUAL RISK OF RECURRENCE AFTER CHEMOTHERAPY IN HIGH RISK ESTROGEN RECEPTOR POSITIVE BREAST CANCERS.

Sabrina Khan, Christos Hatzis and Lajos Pusztai. Section of Medical Oncology, Department of Internal Medicine, Yale University, School of Medicine, New Haven, CT.

ABSTRACT

Gene signature based prognostic tests can help improve adjuvant treatment decisions in early stage estrogen receptor positive (ER) breast cancers. Available tests in the clinic include Oncotype DX recurrence score (RS), PAM50 molecular class, and the Genomic Grade Index (GGI), which can identify high risk tumors that are likely to recur and have less favorable survival when treated with surgery and endocrine therapy alone. These high risk patients are recommended to also receive chemotherapy to improve their chance of survival. A subset of these "high risk" tumors is highly sensitive to adjuvant chemotherapy due to their high proliferation rates, and will be cured. We hypothesized that a new gene signature test ACES, which predicts treatment sensitivity to both endocrine therapy and chemotherapy and identifies tumors with excellent distant relapse free survival (RFS), could further stratify the currently "high risk" ER positive cancers into two groups: ACES predicted low and high residual risk after chemotherapy.

This is a retrospective cohort study, and samples size and power are limited by the number of available specimens. Three independent ER positive breast cancer cohorts – ACES Discovery Cohort (n=176), ACES Validation Cohort 1 (n=123), and a new Validation Cohort 2 (n=127) – were used to assess the ability of ACES to identify patients who were initially considered to be high risk for recurrence (by high RS, Luminal B subtype by PAM50, or high GGI) but became low risk after receiving adjuvant chemotherapy. The ACES algorithm was applied to the baseline high risk groups and cases were re-stratified into ACES predicted treatment sensitive and treatment insensitive groups. RFS and absolute risk reduction (ARR) of relapse were the main outcome measures compared between the ACES stratified groups.

In all three cohorts, cases that were high risk at baseline but predicted to be treatment sensitive by ACES showed a trend toward improved RFS. Cases with high risk by Oncotype DX high RS showed significant difference in RFS by ACES risk strata (p=0.048 and p=0.033) in validation cohort 1 and combined validation cohorts. Among these high RS tumors, n=11-13 (28-35%) were predicted to be treatment sensitive, which had RFS of 92-100% (95% CI: 54-100%) at 4-years. The ARR at 4-years was 0-41% (95% CI: -21-60%) and increased by 10-years to 19% (95% CI: 3-30%) favoring the treatment sensitive groups. Cases with high GGI in the discovery cohort also showed significant differences in RFS by ACES risk strata (p=0.004); the 45 (50%) high GGI cases who were predicted to be treatment sensitive had a RFS of 81% (95% CI: 60-92%) with ARR of 23% (95% CI: -2-51%). For these high RS and high GGI tumors, ACES remained an independent predictor of RFS in multivariate Cox regression analysis including age, T-stage, and lymph node involvement at diagnosis (p=0.072 and 0.017 respectively). Among Luminal B cancers, ACES was significantly associated with RFS only in the multivariate model of both validation cohorts (p=0 and 0.013).

This analysis provides evidence to suggest that ACES may further risk stratify high RS and high GGI tumors into low and high residual risk groups after adjuvant chemotherapy and endocrine therapy. The clinical relevance is that if ACES is adequately validated: (i) patients with low residual risk by ACES can be safely treated with current adjuvant chemotherapies and reassured, (ii) patients with high residual risk despite best current adjuvant chemotherapies could be encouraged to enter clinical trials that aim to improve the efficacy of current adjuvant therapies. Before ACES can be adopted for routine use it would require validation in an adequately powered prospective trial, and the results presented in this thesis suggest that future validation of the ACES algorithm as residual risk prediction tool should be pursued.

INTRODUCTION

Breast cancer represents at least three clinically important and molecularly distinct disease subtypes. Estrogen Receptor (ER) positive breast cancers express estrogen receptors and their growth is stimulated by estrogen. ER (and progesterone receptor [PR]) negative cancers are not dependent on estrogen stimulation and have distinct molecular features and epidemiologic risk factors. The third subtype is the Human Epidermal Growth Factor Receptor-2 (HER2) positive breast cancer, which overexpress HER2 due to gene amplification. HER2 positive breast cancers may be subdivided into HER2 positive/ER positive and HER2 positive/ER negative subtypes. Breast cancers that do not express ER or PR, and are HER2 negative, and are called triple negative or basal-like (a molecular subtype). The HER2 negative/ER positive cancers are comprised of two major molecular subtypes, Luminal A and Luminal B, based on differences in proliferation rate and gene expression profiles.

The different breast cancer subtypes differ in their clinical course (i.e. they have different patterns of relapse and overall survival) and require different therapeutic strategies [1-4]. Over 90% of newly diagnosed breast cancers present as clinical stage I, II or III[1] disease (i.e. localized to the breast or lymph nodes) and are potentially curable with multi-modality therapy [5]. Stage IV is metastatic breast cancer and is generally considered to be an incurable disease.

The focus of this thesis is on stage I-II, ER positive/HER2 negative breast cancers. The standard of care for these cancers includes surgery followed by adjuvant (i.e. postoperative) anti-estrogen (also called endocrine) therapy, with or without adjuvant chemotherapy to eradicate micro-metastatic disease. Almost all patients with ER positive

[1] Stage is defined according to American Joint Committee on Cancer Staging 2010

cancers receive adjuvant endocrine therapy [3] because it improves survival and causes only modest toxicity. However, which ER positive patients should receive adjuvant chemotherapy in addition to endocrine therapy used to be a decision making challenge.

If all ER positive breast cancer patients were treated with adjuvant chemotherapy, about 85% would be over-treated either because they were already cured by surgery and endocrine therapy, or because they had a chemotherapy resistant cancer [6]. To improve patient selection for adjuvant chemotherapy, several efforts have been made to identify patients with such a good prognosis that they would not require further adjuvant chemotherapy. Clinico-pathological factors such as age, comorbidities, tumor size and lymph node involvement are used to help make treatment decisions, based on estimating the risk of recurrence, but they remain imprecise. These anatomical-pathological factors also do not account for the cancer's sensitivity to endocrine therapy or chemotherapy.

Background

Prognostic and Predictive Genetic Signatures

This thesis examines the ability of a multi-gene test to re-stratify initially high risk ER positive breast cancers who receive adjuvant endocrine and chemotherapies into low or high residual risk categories. Prognostic factors are associated with the risk of recurrence of the primary tumor. Predictive factors are associated with the efficacy of a drug or therapeutic regimen [7]. Pure prognostic factors include tumor size and nodal involvement. In contrast, factors such as grade and proliferation rate of the tumor are both prognostic and predictive. ER expression or amplification of the HER-2 gene are primarily predictive markers for anti-estrogen and HER-2 targeted therapies,

respectively. The ACES gene signature that is the focus of this thesis includes endocrine therapy and chemotherapy predictive as well as prognostic components.

In the past 10 years, several multi-gene prognostic tests were introduced into the clinic. These tests categorize newly diagnosed, stage I-II ER positive breast cancers into “low risk” and “high risk” groups at the time of diagnosis. Low risk refers to excellent long-term survival with surgery and endocrine therapy alone. High risk indicates high rates of recurrence of over 15% at 10 years, and less favorable survival, if treated with only surgery and adjuvant endocrine therapy.

Multi-gene prognostic tests are treatment decision making aids, which are independent of, and complementary to the use of clinico-pathological factors and patient preference. Practice guidelines from groups such as the National Comprehensive Cancer Network (NCCN), the American Society for Clinical Oncology (ASCO), and the St. Gallen International Expert Consensus all agree on the general principle that molecular testing can aid in risk stratification [3, 8, 9]. Multi-gene prognostic and predictive tests are multivariate prediction models that use the semi-quantitative expression values of multiple genes to calculate a risk score. They are particularly useful to clinicians when clinico-pathological factors do not clearly point towards whether the patient will benefit from the addition of chemotherapy or not [10]. Several studies have shown that adjuvant treatment recommendations for early stage ER+ breast cancers change about 30% of the time after molecular tests results become available, compared to decisions made entirely based on anatomical-pathological variables [11-13].

Primary prognostic predictors that estimate prognosis in the absence of any systemic therapy include MammaPrint (by Agendia) [11, 13, 14]. Residual-risk

predictors that estimate prognosis after receiving adjuvant endocrine therapy include Oncotype DX (by Genomic Health) [6], PAM50 molecular subtype classifier (called Prosigna by NanoString) [15], and the Genomic Grade Index (GGI) (called MapQuant DX by Ipsogen/Qiagen). Each of these latter group of tests were developed and their performance characteristics defined in clinical studies that included patients who received adjuvant endocrine therapy, but not chemotherapy [6]. There are several other less well standardized prognostic and predictive molecular tests as well as protein marker tests [3]. The tests are collectively referred to as first generation prognostic signatures [16, 17]. They have been independently validated [18], and largely derive their risk stratification power from measuring the proliferation rate of cancers and ER-regulated gene expression [19]. Three of these tests will be utilized in this thesis: Oncotype DX, GGI and PAM50. It is important to note that we use genomic proxy-versions of these tests (i.e. the same genes and same formulas are used as in the commercial assays but gene expression measurements are done with a different platform, Affymetrix gene chips) and not the actual commercially marketed versions.

Genomic Grade Index

GGI was discovered by finding genes which were differentially expressed within histologic grade 3 tumors when compared with histologic grade 1 tumors. GGI consists of 97 genes detected by microarray analysis, which classify a tumor into high grade or low grade, and can also be used to reassign lower or higher grade to intermediate grade 2 cancers. The majority of genes in the GGI are associated with tumor proliferation and cell cycle regulation. In the pivotal validation study, patients with high GGI had 55%

recurrence free survival (RFS) at 10 years with surgery alone, compared to 88% of low GGI cases. GGI has been particularly useful in re-classifying pathologic grade 2 tumors, since they represent intermediate risk tumors, making clinical treatment decisions difficult without further data. High GGI classification, even among grade 2 tumors, has been associated with a significantly higher risk of recurrence in patients treated with adjuvant endocrine therapy [20]. GGI has also been shown to add prognostic information to standard clinico-pathological variables (e.g. age, tumor size, nodal status) [21]. A prospective study demonstrated that GGI is feasible to implement in clinical practice and often changed clinical treatment decisions [22, 23]. GGI was approved by European Community (CE) marking, with the assay conducted in non-centralized laboratories [3].

PAM50

The PAM50 assay measures the expression of 50 cancer genes and 5 control genes using the Nanostring mRNA quantification technology, to assign molecular subtypes including Luminal A, Luminal B, HER2, or basal-like status [24]. PAM50 combined with a proliferation score and tumor size produces a risk of recurrence (ROR) score that predicts risk in 10 years. PAM50 has been validated with large datasets, including by combining data of the Austrian Breast and Colorectal Cancer Study Group 8 (ABCSG-8 clinical trial) and the Arimidex, Tamoxifen Alone or in Combination (transATAC) study, with sample size of above 2,400 patients. Another large study utilized data from the National Cancer institute of Canada, Clinical Trial Group (NCIC CTG MA.12 trial). These studies showed that PAM50 provides independent prognostic

information compared to clinical factors and routine immunohistochemistry markers, in ER positive cancers treated with endocrine therapy [3, 25-27].

The Luminal B subtype as classified by the PAM50 assay, has significantly lower relapse free survival compared to Luminal A and basal-like tumors, when treated with adjuvant Tamoxifen [15]. At 8 years post-surgery, Luminal A tumors had 88% overall survival compared with 76% for Luminal B [28]. Survival from the time of distant relapse for Luminal B tumors is less than 2 years [4]. The assay holds European Union clearance, and approval by the FDA as of 2013 [3].

Oncotype DX

The most well studied of the first generation prognostic signatures is Oncotype DX, and thus it is the test most commonly used in the clinical setting [1, 8, 29]. Oncotype DX is a 21-gene signature measured by RT-PCR to calculate a recurrence score (RS). Five of the 21 genes are reference genes, used to normalize the expression levels of the 16 genes related to breast cancer. These 16 genes include 5 genes that are related to proliferation, 4 that capture ER transcriptional activity, 2 represent genes on the HER2 amplicon, 2 are involved with invasion, and 3 genes have less well defined biological roles. Oncotype DX RS classifies tumors into three risk categories; RS $\leq$18 is low risk of recurrence, RS $\geq$ 30 is high risk, and those in between are intermediate risk [2, 3].

The prognostic performance of Oncotype DX was assessed in several large retrospective studies, where sample sizes within a single study included up to 1372 patients. Several of these studies analyzed tumor samples from completed clinical trials including the National Surgical Adjuvant Breast and Bowel Project (NSABP) B-14 and

B-20 studies [6, 30], the Eastern Cooperative Oncology Group (ECOG) 2197 study [10], the Kaiser Permanente study [31], and the Arimidex, Tamoxifen Alone or in Combination (ATAC) study [32].

High RS was shown to be strongly associated with breast cancer recurrence and death [31]. When treated with endocrine therapy alone, the 10-year distant relapse free survival (DRFS) for high RS, lymph node negative tumors ranged from 69-75%. In contrast, DRFS for low RS tumors ranged from 93-96%, and the difference was highly statistically significant in all studies [6, 33, 34]. Unlike the previous two tests GGI and PAM50 which are sold as test kits to be performed by molecular pathology laboratories, Oncotype DX is a proprietary test performed in a single commercial laboratory [3].

High risk ER positive breast cancer and adjuvant chemotherapy

High risk ER positive breast cancers include the Oncotype DX high RS cases, the Luminal B molecular subtype determined by the PAM50 test, and high GGI. It has become clear that cancers identified as high risk by molecular tests are also the same cancers that are often very sensitive to chemotherapy, likely due to their high proliferation rate. High proliferation rate has been associated with greater response to chemotherapy, which is most clearly demonstrated by neoadjuvant studies, where the adjuvant chemotherapy is administered before surgery and tumor response can be directly measured [35]. When treated with neoadjuvant chemotherapy, Oncotype DX high RS cases achieve considerably greater rates of clinical complete response (CR)[2] and pathologic complete response rates compared to low or intermediate RS cases [35, 36].

[2] Clinical or pathologic complete response (CR or pCR) is defined as no cancer remaining in the primary tumor bed or within regional lymph nodes after chemotherapy.

When adjuvant chemotherapy was added to the treatment of high RS patients, there was a 28% absolute reduction in distant recurrence[3] compared to treatment with adjuvant endocrine therapy alone, in the NSABP B-20 trial [30]. Similar results were found in the SWOG 8814 clinical trial – selective benefit from chemotherapy[4] among the high RS patients, but not among the low RS patients [37].

In order to generate the highest level of evidence on the utility of Oncotype DX for chemotherapy treatment decisions, two prospective randomized clinical trials are ongoing: Trial Assigning Individualized Options for Treatment (TAILORx) [38] and Rx for Positive Node, Endocrine Responsive Breast Cancer (RxPONDER) [39]. In the TAILORx, low RS patients are not given chemotherapy, while high RS receive chemotherapy. Intermediate RS patients are randomly assigned to the chemotherapy arm or the one without chemotherapy. Over 11,000 patients have been recruited in TAILORx. The RxPONDER trial is similarly trying to determine the Oncotype DX RS threshold where chemotherapy is beneficial. This trial will also be comparing the RS to the ROR of PAM50 [2, 3, 34, 38, 39]. The trial results will not be available for several years.

For high GGI breast cancers, this high risk status has also been associated with increased sensitivity to a neoadjuvant chemotherapeutics (including paclitaxel, fluorouracil, adriamycin, and cyclophosphamide). High GGI cases have demonstrated greater rates of pathologic complete response (pCR or RCB-0) and lower residual cancer burden (RCB-I) after neoadjuvant chemotherapy. However, when only ER positive patients with high GGI were analyzed in the same study, survival remained poor even

[3] Patients in the NSABP B-20 trial who were randomized to the Tamoxifen and chemotherapy arm received either cyclophosphamide, methotrexate and 5-fluorouracil (CMF) or methotrexate and 5-fluorouracil (MF).

[4] Patients in SWOG 8814 in the Tamoxifen and chemotherapy arm received cyclophosphamide, doxorubicin, and fluorouracil (CAF).

after chemotherapy. This may be due to low endocrine therapy sensitivity of high GGI tumors, despite having greater sensitivity to chemotherapy [16, 40].

Luminal B breast cancers, identified by PAM50, have also been shown to have greater responsiveness to neoadjuvant chemotherapy, compared to Luminal A tumors [15]. Luminal B tumors are associated with high proliferation, and tend to have high Ki-67 (a nuclear marker for cell proliferation) expression, which can be detected by immunohistochemistry. Tumors with high Ki-67 treated with chemotherapy in addition to endocrine therapy have shown improved disease-free survival, when compared to tumors treated with endocrine therapy alone. In contrast, low Ki-67 tumors have not demonstrated a change in disease free-survival when chemotherapy was added [41].

Because of the consistency of these results, high risk ER positive patients today routinely receive chemotherapy in addition to anti-estrogen therapy to reduce their risk of recurrence. However, what their residual risk is after completion of both endocrine and chemotherapies remain uncertain. It is likely that many patients revert to low risk.

ACES Algorithm

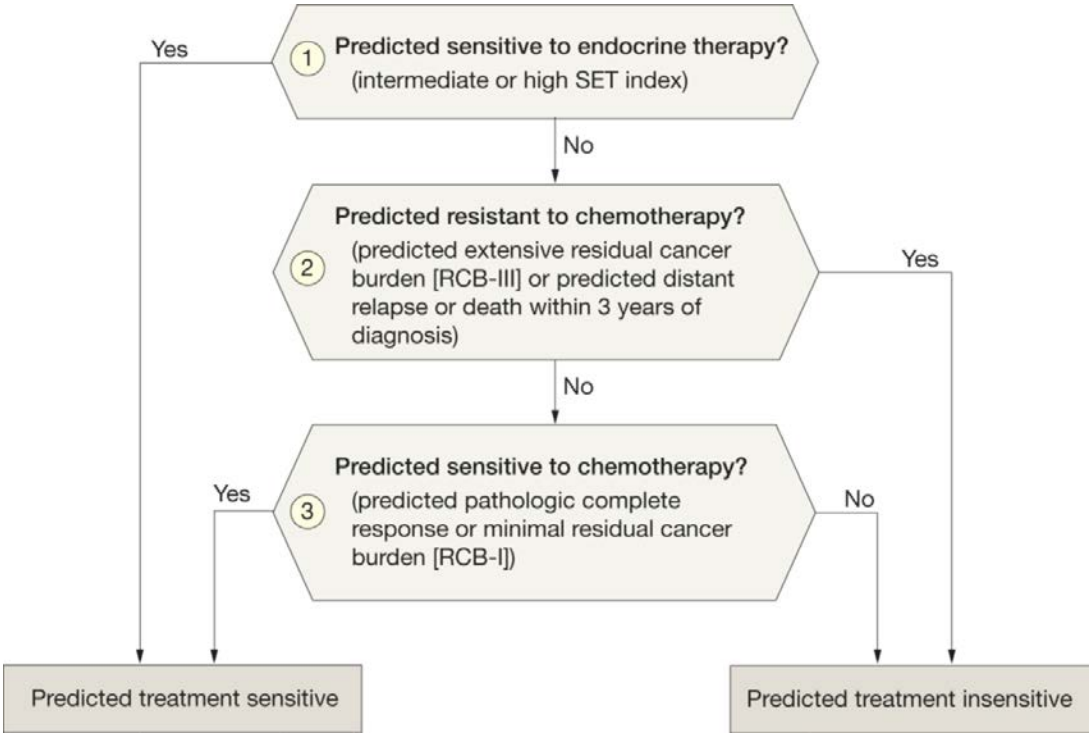
The ACES multi-gene survivor predictor was developed by Hatzis, Pusztai and colleagues to identify both ER positive and ER negative patients who have high excellent DRFS after endocrine therapy and chemotherapy. ACES captures information about sensitivity to endocrine therapy, sensitivity and resistance to chemotherapy and general prognostic information in the absence of any systemic therapy. The algorithm has been validated on an independent patient cohort. The ACES predictor is a combination of four separate multi-gene scores (Figure 1) [42]. The first score predicts sensitivity to

endocrine therapy, which is based on the Sensitivity to Endocrine Therapy (SET) Index. The SET Index, also discovered by Pusztai, Hatzis and colleagues, is a previously published 165-gene set associated with ER, which can predict survival after endocrine therapy or after combined endocrine therapy and chemotherapy [43]. If a tumor is determined to have an intermediate or high SET index, the ACES algorithm will classify it as treatment sensitive. If the SET index is low, then a second gene signature score which predicts resistance to chemotherapy is utilized by ACES. If the chemotherapy resistance score predicts extensive residual cancer (RCB-III), which usually indicates high risk of distant relapse or death within three years of diagnosis, the tumor is classified as treatment insensitive. Seventy-three genes are used to predict RCB-III response in ER positive cancers, and thirty-three other genes predict early relapse or death, despite neoadjuvant chemotherapy. If there is no predicted resistance to chemotherapy by the above criteria, the ACES algorithm analyzes if the tumor is sensitive to chemotherapy. Sensitivity is assessed by predicting pathologic complete response (pCR) or minimal residual cancer burden (RCB-I) after neoadjuvant chemotherapy, which is determined by thirty-nine genes. If the tumor is predicted to be sensitive to chemotherapy at this point, then it is considered overall treatment sensitive; otherwise, it is classified as treatment insensitive. The algorithm is summarized in Figure 1, and the genes utilized for ER positive cases are shown in Appendix Figure A3 [42].

Purpose of this study

The purpose of this thesis is to test the utility of a combined prognostic and endocrine therapy and chemotherapy sensitivity multi-gene signature called ACES, to

FIGURE 1. ACES algorithm (From Hatzis et al [42])



identify among the currently “high risk” ER positive/HER2 negative cancers, those patients who become low risk after receiving adjuvant chemotherapy. The secondary risk stratification value of ACES for ER positive cancers considered “high risk” by other genomic signatures (such as Oncotype DX, PAM50, and GGI), treated with both adjuvant endocrine therapy and chemotherapy, has not been previously reported.

The ACES algorithm is not optimally suited to guide the use of adjuvant chemotherapy because all patients in the discovery and validation cohort (used to discover and validate the ACES algorithm) received both endocrine therapy and chemotherapy. Hence, ACES cannot easily distinguish if endocrine therapy, or the combination of both endocrine therapy and chemotherapy, contributes to good survival.

In contrast, Oncotype DX, PAM50, and GGI were developed from patients who only received endocrine therapy, and therefore cannot inform about low or high residual risk if chemotherapy is also used. The goal of this thesis is to determine if ACES can provide complementary information once it has been decided that chemotherapy is indicated based on Oncotype DX or PAM50 or the GGI score. This thesis employs ACES to address the clinical question: “for which high risk ER positive patient is adjuvant chemotherapy sufficient, and which patient remains high risk despite receiving adjuvant chemotherapy”?

Furthermore, the goal of the original report that described ACES was to develop a survival predictor for ER positive and ER negative breast cancers respectively, but no attempts were made to further categorize ER positive patients into low and high risk groups using existing classification methods. The discovery and validation cohorts for the ACES predictor included ER positive patients who are of low, intermediate and high RS classification by Oncotype DX, as well as Luminal A and Luminal B molecular subtypes by PAM50, and both low and high GGI. Good performance on the combined low and high risk groups does not necessarily imply that ACES performs equally well in both patient subsets.

Hypothesis

A subset of stage I-II ER positive breast cancers that are currently categorized as “high risk” based on poor outcome with adjuvant endocrine therapy alone, are no longer high risk after receiving adjuvant chemotherapy. We hypothesize that a gene signature ACES, which accounts for treatment sensitivity to both endocrine therapy and

chemotherapy, will be able to re-stratify the currently “high risk” ER positive cancers into two groups: ACES predicted low residual risk after adjuvant chemotherapy, and ACES predicted high residual risk despite adjuvant chemotherapy.

Aim

The aim of this thesis is to determine if the ACES genomic predictor can re-stratify ER positive breast cancers called “high risk” by the commonly used prognostic assays Oncotype DX, PAM50 and GGI, into low and high residual risk categories after treatment with adjuvant chemotherapy.

METHODS

Study Design

This is a retrospective cohort study. The ACES predictor was applied to assess residual risk in high risk ER-positive breast cancer cases treated with systemic endocrine and chemotherapy. ER positive cases were first assessed from the cohort used to develop the ACES predictor (cohort 1). The next evaluation of risk stratification was done in the independent validation cohort for ACES (cohort 2). Finally, to assess the generalizability of the results, a blinded independent validation on a third cohort (cohort 3) was performed. For cohort 3, gene expression data without any information on patient outcomes was received, risk categories (by the Oncotype DX RS, PAM50 and GGI) was assigned, predictions by ACES for the high risk subsets were calculated, and finally the predictions were sent back to our collaborator (Dr. Thomas Karn, Goethe-University, Frankfurt, Germany) who plotted survival curves by ACES treatment sensitivity

category. After this blinded independent validation, the patient outcome data was received from cohort 3 in order to perform a pooled analysis combining cohorts 2 and 3 for improved power.

Power Analysis

Assuming a 5-year DRFS of 90% for the ACES predicted "treatment sensitive" strata and 60% for the ACES "treatment insensitive" strata, the estimated hazard ratio (HR) is:

$$HR = \log(0.9)/\log(0.6) = 0.206.$$

The assumption was also made that 30% of the ER positive high risk cases (by high RS, Luminal B subtype, and high GGI), would be re-assigned to ACES "treatment sensitive" (i.e. low residual risk) category after treatment with chemotherapy. Then, the number of events (N) required for the log-rank test to detect significance of this HR at a 0.05 one-sided significance with 80% power, is:

$$N = (1.64 + 0.84)^2/(0.3 * 0.7 * \log(0.206)^2) = 12.$$

The overall event rate in the ER positive cases is about 15%, and would be expected to be even higher in the high risk group. Assuming a 20% event rate at 5 years, a cohort of N=60 would have 80% power to detect a HR of 0.206. This power estimation suggested that if the above assumptions held true, the retrospective study with a sample size limited by availability, would have sufficient power to detect existing significant effects.

Datasets

Three independent cohorts of ER positive/HER2 negative tumors were used.

A. Cohort 1: Discovery Cohort: This cohort was used to discover the ACES algorithm. 310 tumor biopsies of newly diagnosed Stage I to III, invasive breast cancers, obtained prior to any systemic treatment, were collected as part of a prospective international multicenter biomarker discovery study from 2000 to 2006. The samples were obtained by fine-needle aspiration or core biopsy. Tumor messenger RNA (mRNA) hybridization to oligonucleotides was performed with Affymetrix Human Gene U133A GeneChip microarrays. All patients received entirely neoadjuvant taxane-anthracycline chemotherapies, and others classes of chemotherapy were added. Patients received endocrine therapy if ER positive. The cohort included 176 ER positive cases, which were analyzed in this thesis.

B. Cohort 2: Validation Cohort 1: This cohort was used to validate the ACES algorithm, and also published in 2011 [42]. This is an independent cohort to the above discovery cohort. It contains 198 HER2 negative invasive breast cancer patients, for whom biopsy samples were obtained by fine-needle aspiration or core biopsy, from 2002 to 2009. Gene expression profiling was performed at the same laboratory as for cohort 1 using Affymetrix Human Gene U133A GeneChips. One hundred and twenty three patients had ER positive cancers, which are included in the current analysis. All of these patients were treated with sequential taxane and anthracycline chemotherapy, most receiving it as neoadjuvant therapy. All patients also received endocrine therapy [42].

C. Cohort 3: Validation Cohort 2: The third cohort is independent of the previous two cohorts. The aim for including Cohort 3 was to 1) provide an additional cohort for independently validating that ACES can stratify high risk ER-positive/HER2-negative breast cancers in a blinded manner from a different institution, and 2) if the previous aim

could be demonstrated, to combine the two validation cohorts (cohort 2 and cohort 3) to increase the power for detecting a significant stratification by ACES.

This cohort was obtained after searching the published literature and public databases, which included the Gene Expression Omnibus (GEO), and by contacting researchers. Dr. Thomas Karn, based in Goethe-University Frankfurt in Germany, agreed to collaborate. In validation cohort 2, 22% of patients received AC (cyclophosphamide and doxorubicin), 46% EC (epirubicin and cyclophosphamide), 10% CMF (cyclophosphamide, methotrexate and 5-fluorouracil), and 22% TAC (docetaxel, doxorubicin and cyclophosphamide)[5], and all ER positive patients received endocrine therapy. Some of the cases in this cohort are from GeparTrio, a completed multi-center randomized trial by the German Breast Group [44].

Gene expression profiling was performed with Affymetrix U133A and 2.0 gene chips in Dr. Karn's laboratory. Raw intensity files (.CEL files) of 252 cases were provided without any clinical information. Outcome data was not provided initially in order to create a blinded validation study. ER positive/HER2 negative cases were included in our analysis only if microarray based ER and HER2 determination matched the clinical ER and HER2 status provided by Dr. Karn. Thus, 127 cases from the German cohort were analyzed for this thesis.

Data Processing and Generation of Predictions

Microarray data processing: All the raw data (.CEL) files from microarrays were processed using Bioconductor (www.bioconductor.org) and R (www.r-project.org,

[5] Types of anthracyclines include doxorubicin and epirubicin, and types of taxanes include docetaxel.

version 2.10.1) and normalized using custom R programs (packages) developed and provided by Dr. Hatzis.

Standardization of microarray dataset for cohort 3 (i.e. validation cohort 2): Since tumor sample preparation and microarray protocols differed between cohort 3 and cohorts 1 and 2, normalization was done of the genomic indices calculated for cohort 3. Normalization served to transform data within cohort 3, in order to make measurements between the different microarray datasets comparable [45]. The four gene signatures or subcomponents of the ACES algorithm (Figure 1) are each associated with a quantitative score and a numerical threshold which defines its predicted class. The four sub-scores were each separately normalized. Five normalization strategies were used, outlined below (Appendix Tables A4 and A5).

- *Normalization Strategy 1:* Distributions (or proportions of cases) of subcomponents of ACES were matched between validation cohort 2, and the combined discovery and validation cohort 1.
- *Normalization Strategy 2:* Since the original ACES study demonstrated that T-stage was significantly associated with DRFS within a multivariate model which included ACES and ER status [42], normalization was done to account for potentially prognostic variables. In normalization strategy 2, ACES subcomponent distributions were matched within T-stage stratified cases (T1/T2 versus T3/T4 tumors) of validation cohort 2.
- *Normalization Strategy 3:* Similarly, ACES subcomponent distributions were matched within lymph node stratified cases (lymph node positive versus negative) of validation cohort 2.

- *Normalization Strategy 4:* ACES subcomponent distributions were matched within grade stratified cases (grade 1 or 2, versus grade 3).
- *Normalization Strategy 5:* Multivariate linear regression models were built to adjust for differences between the microarray datasets by adjusting for imbalances in disease spectra between validation cohort 2, and the combined development cohort and validation cohort 1. The ACES subcomponents were the response variables, and the predictor variables were age, T-stage, nodal status, grade and cohort (1 and 2 versus 3). Any significant effect from cohort 3 was estimated from the model and subtracted from the corresponding subcomponent score.
- *Normalization of baseline risk classifiers:* GGI values were normalized in cohort 3 by redefining the numerical threshold that determines high versus low GGI. Oncotype DX RS was not normalized as its microarray based thresholds have not been validated. PAM50 was not normalized as this assay does not involve numerical thresholds but determines molecular subtype by closest similarity to prototypical expression patterns, or subtype centroids.

Definition of Risk Groups: "High risk" cases were defined as: high RS classification by Oncotype DX, Luminal B molecular class by PAM50, or high grade by GGI. Standard, previously published methods were used when applying the prognostic predictors to the three cohorts [42]. Since the Oncotype DX is a proprietary assay of Genomic Health and is performed by using a polymerase chain reaction (PCR) assay, a published genomic surrogate version of the test was used. The surrogate version uses gene expression values of the same 21 genes as in the proprietary assay, but is generated by Affymetrix gene expression arrays [46].

Generation of ACES predictions: The ACES algorithm was applied to each of the "high risk" groups (high RS, Luminal B and high GGI) within the three cohorts. For validation cohort 2, the ACES algorithm was applied after each of the five normalization strategies was carried out. The "high risk" cases were stratified by ACES into predicted "treatment sensitive" and "treatment insensitive" strata.

Statistical Analysis

Assessment of Predictor Performance: Kaplan-Meier relapse-free survival (RFS) curves were plotted of each "high risk" group by response strata predicted by ACES. RFS was defined as the time from initial biopsy at diagnosis until relapse was diagnosed. Distant relapse was the outcome of the discovery and validation cohort 1, while any relapse (local or distant) was the outcome of validation cohort 2 (because distant relapse free survival was not made available). Observations were right censored at the time of loss to follow-up. The ACES predicted "treatment sensitive" and "treatment insensitive" strata were compared by the log-rank test. Multivariate Cox regression models were used to adjust the risk associated with ACES for other clinical prognostic variables: age, T-stage, and nodal status at time of diagnosis.

The ACES predicted strata were also compared by calculating the Absolute Risk Reduction (ARR) at 4 years (and 10 years in validation cohort 2); the associated 95% confidence interval (CI) was calculated under bootstrap [42]. The performance of ACES in predicting RFS was assessed by calculating the algorithm's sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), diagnostic positive and negative likelihood ratios (LR+ and LR-) and Odds Ratio (which equals LR+/LR-) with

associated 95% CI [42].

The data analysis was done using a combination of R, SAS 9.3 and Excel 2010.The statistical analysis was conducted by me under the primary supervision of Dr. Hatzis, and also by Dr. Pusztai. As the outcome data of validation cohort 2 was initially blinded, Dr. Karn performed the initial survival analysis of this dataset.

RESULTS

Pre-treatment characteristics

Patient demographics and tumor characteristics for the three cohorts (discovery cohort, n=176; validation cohort 1, n=123; and validation cohort 2, n=127) are shown on Table 1. The discovery cohort and validation cohort 1 were similar in age, grade, T-stage, nodal status, AJCC stage and Progesterone Receptor (PR) status. Comparison of validation cohort 2 with the combined discovery and validation cohort 1 showed a significant difference in T-stage or tumor size (p=0.000), nodal status (p=0.001) and pathologic grade (p=0.041) distribution. There were more T1 stage tumors, less lymph node positive cases, and more pathologic grade 1 or 2 tumors within validation cohort 2, but no difference was seen in age. Normalization strategies 2, 3, 4 and 5 of validation cohort 2 adjusted for the impact of these clinical prognostic variables on RFS (Appendix Table A4-A5).

High risk classification by Oncotype DX, PAM50 and GGI

The GGI assigned the highest number of patients to high risk category in all three cohorts; n=90 (51%) in the discovery cohort, n=64 (52%) in validation cohort 1, and

TABLE 1. Pre-treatment Characteristics

Characteristic	**Discovery Cohort** N (% of 176)	**Validation Cohort 1** N (% of 123)	**Validation Cohort 2** N (% of 127)	**Odds Ratio†**	**p-value†**
Age					
<50	90 (51)	64 (52)	63 (50)	1.08	0.751
≥50	86 (49)	59 (48)	64 (50)		
Mean (SD)	51 (10)	50 (10)	51 (10)		
T-stage					
1	26 (15)	8 (7)	43 (34)	0.25	**0.000**
2	94 (53)	54 (44)	69 (54)		
3	28 (15)	41 (33)	9 (7)		
4	21 (12)	19 (15)	6 (5)		
Unknown	7 (4)	1 (1)	0		
Nodal Status					
Negative	61 (35)	46 (37)	69 (54)	0.47	**0.001**
Positive	113 (64)	77 (63)	57 (45)		
Unknown	2 (1)	0	1 (1)		
AJCC stage					
I	1 (1)	1 (1)	-	-	-
II	103 (59)	51 (41)			
III	71 (40)	45 (37)			
Unknown	1 (1)	26 (21)			
Grade					
1	19 (11)	10 (8)	12 (9)	0.61	**0.041**
2	100 (57)	53 (43)	83 (65)		
3	46 (26)	54 (44)	32 (25)		
Unknown	11 (6)	6 (5)	0		
PR status					
Negative	43 (24)	30 (24)	-	-	-
Positive	130 (74)	93 (76)			
Indeterminate	3 (2)	0			

† Odds ratio and p-value reflects the results of Fisher's exact test when Validation Cohort 2 is compared with the combined Discovery Cohort and Validation Cohort 1; age, T-stage, lymph node involvement and pathologic grade at time of diagnosis were compared between the cohorts. Abbreviations: PR, progesterone receptor; SD, standard deviation; AJCC, American Joint Committee on Cancer

n=54 (43%) in validation cohort 2 (Table 2). The number of high RS and Luminal B cases were smaller within each cohort, ranging from n=37 (29%) to n=39 (32%) for high RS, and from n=18 (14%) to n=40 (23%) for Luminal B. This study was determined to have 80% power in detecting statistical significance when it exists, for a sample size of at least 60. Therefore, for high GGI cases, there was sufficient statistical power to detect

TABLE 2. Distribution of risk classification by Oncotype DX, PAM50 and GGI

Group	Discovery Cohort N (% of 176)	Validation Cohort 1 N (% of 123)	Validation Cohort 2 N (% of 127)	Odds Ratio†	p-value†
Oncotype DX					
High RS	39 (22)	39 (32)	37 (29)	0.89	0.648
Intermediate RS	21 (12)	9 (7)	31 (24)		
Low RS	116 (66)	75 (61)	59 (46)		
PAM50					
Luminal B	40 (23)	33 (27)	18 (14)	1.72	**0.055**
Luminal A	99 (56)	55 (45)	94 (74)		
Basal	11 (6)	15 (12)	2 (2)		
HER2	11 (6)	9 (7)	5 (4)		
Normal	15 (9)	11 (9)	8 (6)		
GGI					
High	90 (51)	64 (52)	54 (43)‡	1.21	0.349
Low	86 (49)	59 (48)	73 (57)‡		

† Odds ratio and p-value reflects the results of Fisher's exact test when Validation Cohort 2 is compared with the combined Discovery Cohort and Validation Cohort 1. ‡ Normalized. Abbreviations: RS, Recurrence Score; GGI, Genomic Grade Index

the HR=0.206, but the study was underpowered for Luminal B and high RS groups due to the lower prevalence of high risk predictions by these methods. Furthermore, in validation cohort 2, all high risk categories contained less than 60 cases. Therefore, the two validation cohorts were combined to increase sample size and thus the ability to detect statistical significance when present.

There were more Luminal B cases in the discovery and validation cohort 1, than in validation cohort 2 (p=0.055), but the distribution of high RS and high GGI cases did not differ (Table 2). The overlap between the three high risk groups is shown on Table 3. The majority of Luminal B tumors, 85-100%, are also classified as high GGI, within all three cohorts. The next highest level of overlap is seen among high RS tumors which are also high GGI, ranging from approximately 50-70% within the cohorts. The overlap between the other high risk groups is lower, ranging from 5% to 36%. It has been reported previously that Oncotype DX classifies most Luminal B tumors into high RS

TABLE 3. Overlap in risk prediction the three baseline risk classifiers

High Risk Group	Discovery Cohort	Validation Cohort 1	Validation Cohort 2
High RS	N (% of High RS cases)		
Luminal B	10 (26)	2 (5)	4 (11)
High GGI	28 (72)	23 (59)	18 (49)
Luminal B	N (% of Luminal B cases)		
High RS	10 (25)	2 (6)	4 (22)
High GGI	37 (93)	28 (85)	18 (100)
High GGI	N (% of High GGI cases)		
High RS	28 (31)	23 (36)	18 (33) ‡
Luminal B	37 (41)	28 (44)	18 (33) ‡

‡ Normalized. Abbreviations: RS, Recurrence Score; GGI, Genomic Grade Index

categories [24]. In this dataset, about 25% of Luminal B cases are also high RS in the discovery cohort, but only about 5-11% are so in validation cohorts.

ACES prediction of residual risk

The ACES algorithm (Figure 1) classified 22-50% of "high risk" cases (high RS, Luminal B, and high GGI) as treatment sensitive. This proportion did not differ significantly from the proportion of all ER positive tumors (regardless of baseline risk) in the cohorts predicted to be ACES treatment sensitive. The Kaplan-Meier relapse free survival plots of the high risk cases re-stratified by ACES, are shown in Figure 2. The Kaplan-Meier survival plots of validation cohort 2, after normalization, are shown in Appendix Figure A1. The median follow-up time was 3 years for both the discovery cohort and validation cohort 1, and the range was from 0-7 years. For validation cohort 2, median time of follow-up was 5 years, with range from 0.4-10 years.

In all "high risk" categories within all cohorts (discovery cohort, validation cohort

1, and normalized validation cohort 2), the ACES predicted treatment sensitive strata showed a trend toward improved RFS when compared to the ACES predicted treatment insensitive groups. Within validation cohort 2, consistent trends were only seen after normalization utilizing stratification by T-stage (normalization strategy 2; Appendix Figure A1 and Table A6); therefore, this normalized dataset was pooled with validation cohort 1. Within Luminal B cases of the discovery cohort, the trend of ACES predicted sensitive strata having improved RFS was only present before four years. No such time dependence was seen in other groups.

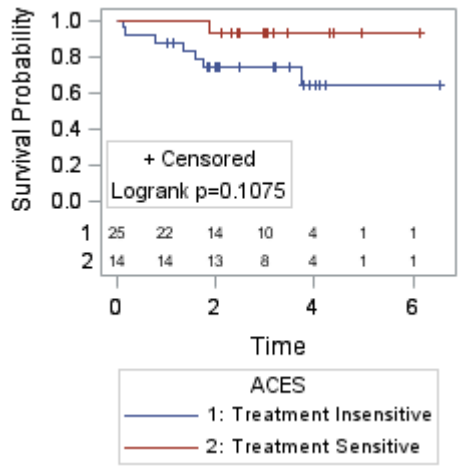
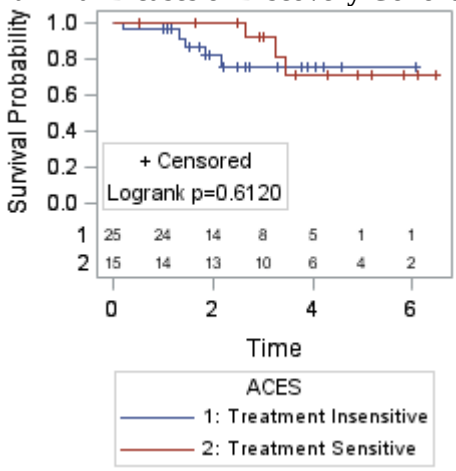
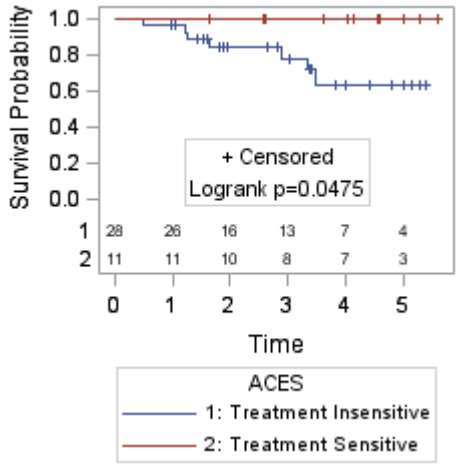
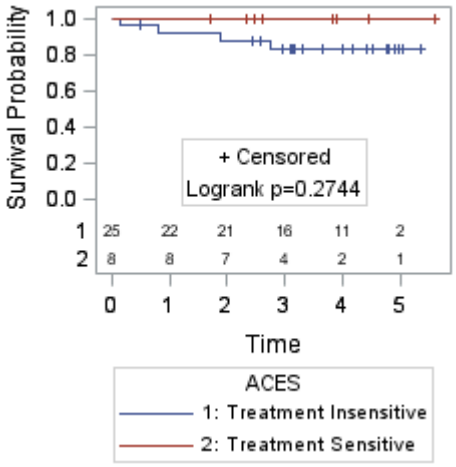
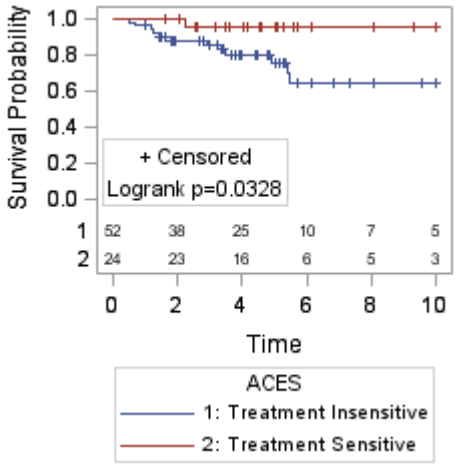
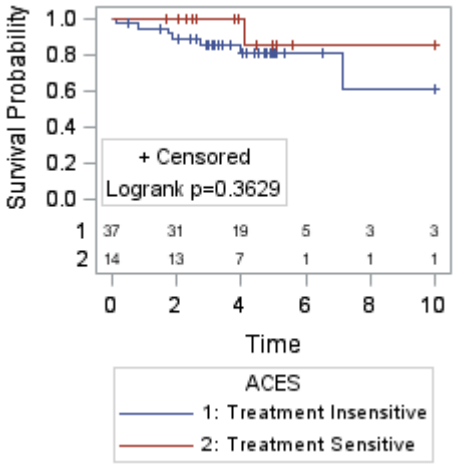
- High RS

A significant difference in RFS was seen between the ACES predicted treatment sensitive and insensitive strata among high RS cases of validation cohort 1 and the combined validation cohorts (p=0.048 and p=0.033). High RS cases in the discovery cohort had a corresponding log-rank test p-value of 0.108. There were 0-1 (0-8%) relapses within the ACES predicted treatment sensitive groups, and 5-7 (21-28%) relapses in the predicted treatment insensitive groups of all cohorts (Table 4). The 4-year RFS in the validation cohorts was 92-100% (95% CI: 54-100%) in the ACES predicted sensitive strata of the validation cohorts. In validation cohort 1, the 4-year RFS in the predicted treatment insensitive strata was 64 (95% CI: 36-82%) in the predicted insensitive strata, with an ARR in relapse at 4 years of 41% (95% CI: 14-60%) favoring the predicted treatment sensitive strata. The ARR at 4-year for validation cohort 2 was 0% (95% CI: -21-17%), but by 10 years, the ARR rose to 19% (95% CI: 3-30%).

- Luminal B

Statistically significant differences in RFS were not seen between the ACES

Figure 2. Kaplan Meier Estimates of Relapse Free Survival

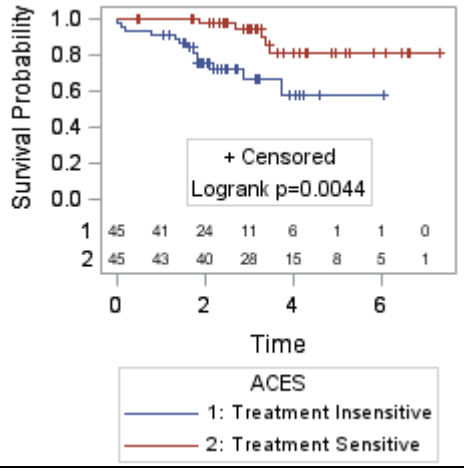
High RS	Luminal B
High RS cases of Discovery Cohort	**Luminal B cases of Discovery Cohort**
	
High RS cases of Validation Cohort 1	**Luminal B cases of Validation Cohort 1**
	
High RS cases of Validation Cohorts 1/2†	**Luminal B cases of Validation Cohorts 1/2†**
	

Number of cases at risk shown within plot. †Normalized using stratification by T-stage.

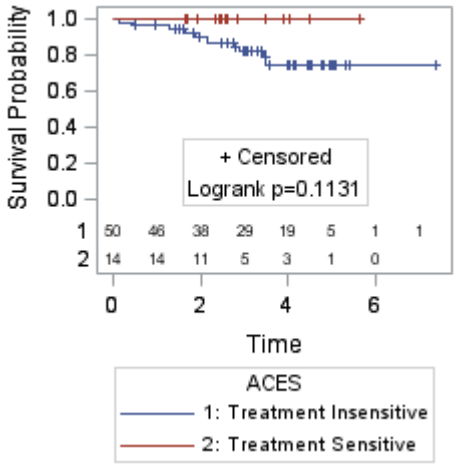
Figure 2 (continued). Kaplan Meier Estimates of Relapse Free Survival

High GGI

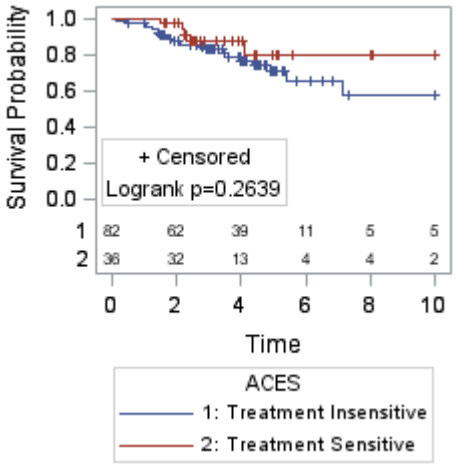
High GGI cases of Discovery Cohort



High GGI cases of Validation Cohort 1



High GGI cases of Validation Cohorts 1 and 2†



Number of cases at risk shown within plot. †Normalized using stratification by T-stage.

predicted treatment sensitive and insensitive strata among Luminal B cases. The ARR at 4 years ranged from 0-17% (95% CI: 1-38%) in the validation cohorts, and was -5% (95% CI: -25-10%) in the discovery cohort. In the discovery cohort, although lower risk of relapse was seen before 4 years in the ACES predicted treatment sensitive group, this trend did not continue after this time. In both validation cohorts, there were 0-1 (0-17%) relapses in the treatment sensitive strata, and 3-4 (16-25%) relapses in the treatment insensitive strata. The RFS at 4-years for the treatment sensitive groups was 100% (95% CI: 100-100%) in both validation cohorts, and ranged from 80-83% (95% CI: 39-95%) in the predicted treatment insensitive groups. The ARR at 10-years was 20%[6] in validation cohort 2, rising from 0% (9 to 36%) at 4-years.

- High GGI

A significant difference in RFS was seen between the ACES predicted strata within high GGI cases of the discovery cohort (p=0.004). Such a significant difference was not found in validation cohort 1 (p=0.113), but the 14 high GGI cases predicted to be treatment sensitive in this cohort did not experience any relapses while 10 (20%) had a relapse in the predicted treatment insensitive strata. The ARR at 4-years was 25% (95% CI: 18-38%) in validation cohort 1. The ARR at 4-years was 0% (95% CI: -24-23%) in validation cohort 2, but the ARR at 10-years increased to 16%[7].

Clinical Factors associated with Relapse Free Survival

Clinical factors associated with relapse free survival were explored with multivariate Cox Proportional Hazards regression models (Appendix Table A1). The

[6] 95% Confidence Interval could not be obtained under bootstrap.
[7] 95% Confidence Interval could not be obtained under bootstrap

TABLE 4. Survival Analysis

High Risk Group	Cohort 1	Cohort 2	Cohort 3†
High RS			
ACES Rx Insensitive			
No.	25	28	24
No. Relapses (Event Rate, %)	7 (28)	7 (25)	5 (21)
No. Censored	18	21	19
4-yr RFS (95% CI)	64 (35 to 83)	64 (36 to 82)	92 (71 to 98)
10-yr RFS (95% CI)	-	-	73 (45 to 88)
ACES Rx Sensitive			
No.	14	11	13
No. Relapses (Event Rate, %)	1 (7)	0	1 (8)
No. Censored	13	11	12
4-yr RFS (95% CI)	93 (59 to 99)	100 (100 to 100)	92 (54 to 99)
10-yr RFS (95% CI)	-	-	92 (54 to 99)
ARR at 4-yr	29 (3 to 46)	41 (14-60)	0 (-21 to 17)
ARR at 10-yr	-	-	19 (3 to 30)
p-value of Log-Rank test	0.108	**0.048**	**0.033**‡
Luminal B			
ACES Rx Insensitive			
No.	25	25	12
No. Relapses (Event Rate, %)	5 (20)	4 (16)	3 (25%)
No. Censored	20	21	9
4-yr RFS (95% CI)	76 (51 to 89)	83 (61-93)	80 (39 to 95)
10-yr RFS (95% CI)	-	-	60 (16 to 86)
ACES Rx Sensitive			
No.	15	8	6
No. Relapses (Event Rate, %)	3 (20)	0	1 (17%)
No. Censored	12	8	5
4-yr RFS (95% CI)	71 (34 to 90)	100 (100 to 100)	100 (100 to 100)
10-yr RFS (95% CI)	-	-	80 (20 to 97)
ARR at 4-yr	-5 (-25 to 10)	17 (0.8 to 38)	0 (9 to 36)
ARR at 10-yr	-	-	20 (*)
p-value of Log-Rank test	0.612	0.274	0.363‡
High GGI			
ACES Rx Insensitive			
No.	45	50	32
No. Relapses (Event Rate, %)	13 (29)	10 (20)	10 (31)
No. Censored	32	40	22
4-yr RFS (95% CI)	58 (34 to 76)	75 (58 to 86)	80 (60 to 91)
10-yr RFS (95% CI)	-	-	56 (30 to 75)
ACES Rx Sensitive			
No.	45	14	22
No. Relapses (Event Rate, %)	5 (11)	0	5 (23)
No. Censored	40	14	17
4-yr RFS (95% CI)	81 (60 to 92)	100 (100 to 100)	80 (55 to 92)
10-yr RFS (95% CI)	-	-	72 (44 to 88)
ARR at 4-yr	23 (-2 to 51)	25 (18 to 38)	0 (-24 to 23)
ARR at 10-yr	-	-	16 (*)
p-value of Log-Rank test	**0.004**	0.113	0.264‡

‡p-value reflects that of the combined validation cohort 1 and 2. †Normalized using stratification of cohort by T-stage (See Strategy 2 in Methods). *Values could not be obtained under bootstrap. Abbreviations: RS, Recurrence Score; GGI, Genomic Grade Index; No., number; RFS, relapse-free survival; Rx, treatment; CI, confidence interval; ARR, absolute risk reduction (where positive number indicates lower risk in the ACES predicted treatment sensitive strata); yr, year

models included ACES, age, clinical T-stage and clinical nodal status at time of at diagnosis. Age was greater than or equal to 50 (versus less than 50), T-stage was T3/T4 (versus T1/T2), node positive (versus negative) at time of diagnosis, and ACES predicted strata was treatment sensitive (versus treatment insensitive).

Within high GGI cases, the ACES algorithm was significantly associated with RFS in the discovery cohort and validation cohort 1 (p=0.017 and p=0 respectively). Being predicted treatment sensitive by ACES within high GGI cases indicated a relapse rate of about one-third that of the treatment insensitive group (HR = 0.3) in the discovery cohort. The HR was 10^{-9} in validation cohort 1, which reflects that almost all events (relapses) are associated with the ACES predicted treatment insensitive strata. Positive nodal involvement was significantly associated with RFS in validation cohort 2 (HR=4, p=0.042) among high GGI tumors.

Within high RS cases, the ACES algorithm was associated significantly with RFS in validation cohort 1 (HR=10^{-9}, p=0). In the combined validation cohorts, HR was 0.1 with p of 0.072. No significant effect of ACES was seen in validation cohort 2 alone, but in this cohort, T-stage was extremely strongly associated with RFS (HR=10^{-9}, p=0). Lymph node involvement was also associated with RFS in validation cohort 2 and the combined validation cohorts (HR=5-9, p=0.047 and 0.042).

In Luminal B tumors of validation cohort 1 and 2, ACES was significant with HR=10^{-9} – 0.15 (p=0 and p=0.013). In Luminal B cases of the discovery cohort, none of the model's variables were significant predictors of RFS. In the initial analysis of the ACES algorithm in 2011, a multivariate Cox regression model showed that ACES, T-stage and ER status were associated with DRFS [42], but not lymph node involvement. In

this analysis, nodal status was a significant predictor of RFS among both high RS and high GGI tumors only within validation cohort 2. T-stage was a significant predictor of RFS only within validation cohort 2.

Performance of ACES algorithm

Performance of the ACES algorithm in accurately predicting relapse free survival at 4-years is shown in Table 5. A positive test is being predicted treatment insensitive by ACES. The sensitivity and NPV of ACES was higher compared to the sensitivity and positive predictive value (PPV). Sensitivity ranges from 60-100% (95% CI: 9-100%) for all three high risk groups. In tests with high sensitivity, a negative test (ACES predicted treatment sensitive) is associated with no event (relapse). The important performance parameter here is the NPV, which is the probability of no relapse among those predicted to be treatment sensitive by ACES, and ranges 71-100% (95% CI: 48-100%). Specificity is lower ranging from 25-56% (95% CI: 13-67%). The positive predictive value (PPV) of ACES (probability of relapse among those predicted as treatment insensitive), is low at 8-44% (95% CI: 0-64%).

The LR- indicates how many times more likely it is for those without relapse to have an ACES treatment sensitive prediction, than it is for those with relapse. The LR- was significant (less than 1) in high RS cases in the discovery cohort and validation cohort 1, among Luminal B cases of the validation cohorts, and within high GGI cases of validation cohort 1; LR- was 0 (95% CI: 0.01-0.01). The LR+ indicates how much more likely it is for a patient with relapse to have an ACES predicted treatment insensitive tumor, compared to a patient without relapse. The LR+ ranged from

TABLE 5. Performance of the ACES algorithm in predicting Relapse Free Survival

Risk Category	**Discovery Cohort** Value or % (95% CI)	**Validation Cohort 1** Value or % (95% CI)	**Validation Cohort 2†** Value or % (95% CI)
High RS			
Sensitivity	88 (47 to 100)	100 (59 to 100)	67 (9 to 99)
Specificity	42 (25 to 61)	34 (19 to 53)	35 (20 to 54)
PPV	36 (6 to 56)	36 (8 to 56)	8 (0 to 19)
NPV	93 (80 to 100)	100 (100 to 100)	92 (77 to 100)
LR+	1.6 (0.65 to 2.84)	1.8 (0.45 to 5.10)	1.0 (0 to 3.06)
LR-	**0.2 (0.01 to 0.37)**	**0 (0.01 to 0.01)**	1.0 (0.01 to 3.74)
OR	7.3 (1.8 to 229)	Infinite (45 to 510)	1 (0 to 244)
Luminal B			
Sensitivity	63 (24 to 91)	100 (40 to 100)	100 (16 to 100)
Specificity	38 (21 to 56)	28 (13 to 47)	38 (15 to 65)
PPV	24 (3 to 41)	17 (0.3 to 31)	20 (0 to 42)
NPV	71 (48 to 100)	100 (100 to 100)	100 (100 to 100)
LR+	0.8 (0.55 to 1.32)	1.3 (0.71 to 2.39)	1.7 (0.50 to 2.82)
LR-	1.0 (0.41 to 2.5)	**0 (0.01 to 0.01)**	**0 (0.01 to 0.01)**
OR	0.8 (0.31 to 3.34)	Infinite (71 to 239)	Infinite (50 to 282)
High GGI			
Sensitivity	72 (47 to 90)	100 (69 to 100)	60 (26 to 88)
Specificity	56 (43 to 67)	26 (15 to 40)	41 (26 to 57)
PPV	42 (16 to 60)	25 (10 to 38)	44 (15 to 64)
NPV	81 (68 to 98)	100 (100 to100)	72 (53 to 97)
LR+	1.7 (0.63 to 4.2)	1.3 (0.46 to 1.77)	0.98 (0.28 to 2.19)
LR-	0.5 (0.11 to 1.27)	**0 (0.01 to 0.01)**	0.98 (0.01 to 2.59)
OR	3.2 (0.83 to 23)	Infinite (46 to 177)	1.0 (0.22 to 22)

†Normalized using stratification of cohort by T-stage (See Strategy 2 in Methods). Abbreviations: PPV, positive predictive value; NPV, negative predictive value; LR+, Positive Likelihood Ratio; LR-, Negative Likelihood Ratio; OR, Odds Ratio.

0.8-2.0 (95% CI: 0-8). The associated OR (LR+/LR-) of the groups with significant LR- was thus undefined, but indicates that there is increase in the odds of having a relapse when predicted treatment insensitive by ACES.

The performance measures (sensitivity, specificity, PPV, NPV, LR+, LR-) of all ER positive cases were calculated and compared to that of the high risk cases. The only significant differences found were among Luminal B tumors of validation cohort 2. The

NPV is greater, and the LR- is lower with a higher OR, among Luminal B cases of validation cohort 2 compared to all ER positive cases of the cohort; in the composite ER positive cases, NPV was 88% (95% CI: 79-98%), LR- was 0.8 (95% CI: 0.4-1.5), and OR was 1.4 (95% CI: 0.4-2.8).

Distribution of subcomponents of ACES

The distribution of the four subcomponents of the ACES algorithm (SET Index**,** predicted RCB-0/I, RCB-III and early relapse/death gene signatures shown in Figure 1) within the high risk groups was compared to all ER positive cases in its corresponding cohort (Appendix Table A2). One purpose of this comparison was to see if the high risk tumors have fewer proportion of cases that are sensitive to endocrine therapy (according to SET Index), lower rates of pathologic response (according to the RCB-0/I or RCB-III gene signatures) and/or greater rates of death or early relapse within 3 years of diagnosis. No such differences were seen in Validation Cohort 2 (after normalization).

Among Luminal B tumors, the only significant difference in the distribution of ACES subcomponents seen was in the SET Index, when compared with all ER positive cases. SET Index classification of high/intermediate status (i.e. high/intermediate sensitivity to endocrine therapy) alone is sufficient to classify cases as overall ACES predicted treatment sensitive (Figure 1). There were only 0-1 (0-3%) Luminal B cases classified as SET high/intermediate in the discovery and validation cohort 1. In contrast, more tumors [14-45 (11-26%)] were classified as SET high/intermediate within the composite ER positive cohorts which contain low risk tumors as well (p=0.001 and p=0.042 in the discovery and validation cohort 1 respectively).

Among high GGI cases, there were also fewer tumors with SET high/intermediate classification when compared to the proportion of all ER positive cases with SET high/intermediate status in validation cohort 1 (p=0.003). Fewer high GGI cases had extensive residual cancer burden compared to all ER positive cases in the discovery cohort; this pattern was not seen in validation cohort 1. No other ACES subcomponent showed a different distribution among high GGI cases as compared with all ER positive cases.

Within the high RS tumors, a greater proportion of cases were predicted to have early relapse/death, than within all ER positive cases (p=0.022) in the development cohort; such a difference was not seen in validation cohort 1. In high RS cases, no other ACES subcomponent showed a different distribution from the composite ER positive groups.

The high risk cases were also stratified by each of the four subcomponents of the ACES algorithm (Appendix Figure A2). High RS tumors predicted to have high or intermediate SET class showed trends toward improved RFS compared with tumors with low SET class, in the development and combined validation cohorts. None of the 11 SET-high/intermediate tumors experienced relapse, while 8 -13 (19-23%) of SET-low tumors did. High RS tumors predicted to be at high risk for early relapse/death (within 3 years of diagnosis) by this ACES subcomponent, demonstrated a trend toward lower RFS. Stratification by RCB-0/I and RCB-III predictors did not show consistent trends in all cohorts. None of the subcomponents demonstrated statistically significant stratification by the log-rank test.

Among Luminal B tumors, there was a significant stratification of cases by the

early relapse/death predictor ($p < 0.0001$) in the discovery cohort. However, in the combined validation cohorts, while tumors with high risk for early relapse/death had higher probability of RFS before 6 years compared with low risk tumors, the trend was not present after this time. Only 1-2 (3-4%) Luminal B tumors were predicted to be SET-High/Intermediate class; hence the risk stratification power of SET Index could not be assessed. The RCB-0/I and RCB-III predictors within Luminal B cases did not show consistent trends in all cohorts.

In high GGI tumors, there was a statistically significant stratification of cases by the SET index in the development and combined validation cohorts ($p=0.0316$ and $p=0.0318$ respectively). High risk tumors by the early relapse/death predictor also showed significant stratification in the development cohort ($p < 0.0001$), but not in the validation cohorts. Again, the stratification by RCB-0/I or RCB-III predictors were not prominent. Therefore, the dominant subcomponents that drive the stratification power of the overall ACES algorithm in predicting RFS are likely SET Index and the early relapse/death predictors.

Performance of gene signature for predicting pathologic response

Two of the subcomponents of the ACES algorithm predict post-chemotherapy pathologic complete response and minimal residual cancer burden (RCB-0/I), or extensive residual cancer burden (RCB-III) (Figure 1). The performance of these two gene signatures in predicting actual pathologic response in the high risk groups of the discovery and validation cohort 1[8] are shown in Appendix Table A3. There was limited data available for validation cohort 1. The sensitivity, specificity, PPV and NPV were

[8] Actual pathologic response data was not available for Validation Cohort 2 for comparison.

variable, ranging from 13-96% (95% CI: 2-100%). The LR- was significant for High RS, Luminal B cases and high GGI cases and the RCB-0/I predictor, and within high RS and Luminal B cases and the RCB-III predictor [0-0.4 (95% CI: 0-0.99)]. These performance measures were similar to that of the corresponding entire ER positive cohort.

DISCUSSION

This thesis demonstrated the potential of the ACES algorithm to classify "high risk" (high RS, Luminal B and high GGI) invasive breast cancers into low and high residual risk strata after treatment with chemotherapy and endocrine therapy. A statistically significant difference in relapse free survival was found between the ACES predicted treatment sensitive and insensitive strata, within high RS cases of validation cohort 1 and in the combined validation cohorts (log rank test p=0.048 and p=0.033), and within high GGI cases of the discovery cohort (p=0.004). These results were confirmed by Cox regression analysis after adjusting for the effects of other clinical covariates age, T-stage and nodal status at time of diagnosis (for high RS, HR=10^{-9} to 0.15, p=0 and 0.072 in validation cohort 1 and the combined validation cohorts; and for high GGI, HR=10^{-9} to 0.29, p=0 and 0.017 in validation cohort 1 and the discovery cohort). Among Luminal B tumors, multivariate Cox regression also showed a significant association of ACES with RFS in both validation cohorts (HR=10^{-9} to 0.15, p=0 and 0.013).

High GGI cases of the discovery cohort had the largest sample size and most recurrence events compared to all other risk groups. This may be a reason why a strong statistically significant risk re-stratification could be detected in this cohort. In other high risk group/cohort combinations, trends showed toward lower residual risk in the ACES

predicted treatment sensitive strata, but statistical significance was not consistently reached. These groups may not have had the adequate number of cases needed to detect statistical significance. In the power calculations for this thesis proposal, it was estimated that the study would have 80% power to detect statistical significance with a sample size of 60, and event size of at least 12. In the high GGI cases of the discovery cohort, sample size was 90 and 18 relapses occurred, meeting the required numbers estimated in the power analysis. The next largest sample size of 64 was within high GGI cases in the validation cohort 1, but only 10 events occurred in this group.

A second validation cohort was therefore utilized to further assess the performance of ACES. One general limitation of this approach is that when heterogeneous microarray platforms are used in the analysis of tumors, the same numerical measurements may become harder to compare across platforms. The ACES algorithm was developed and validated using tumors assessed with Affymetrix Human Gene U133A GeneChip microarrays. Not all cases in validation cohort 2 utilized the U133A platforms. Small differences in sample collection and preparation protocols can also lead to large deviations in microarray results [47]. Therefore, the use of different microarray platforms likely introduced a bias in the performance of ACES within validation cohort 2.

A potential solution employed by this thesis was normalization to make measurements between the different microarrays comparable. Blinded assessment of different normalization strategies showed that standardization of the distributions of individual subcomponents within strata defined by T-stage appeared promising. High RS, Luminal B, and high GGI cases within this normalized validation cohort 2 data

demonstrated consistent trends in lower residual risk in the ACES predicted treatment sensitive strata. Stratification by T-stage or tumor size may have been successful due to the association of T-stage with relapse; this association was demonstrated in the initial study which described the ACES algorithm [42]. By controlling for the effect of T-stage, the association of ACES with relapse may have become easier to detect.

The difference in relapse between ACES predicted treatment sensitive and insensitive strata within the normalized validation cohort 2 was not statistically significant however; the sample size (18-54) and event size (1-10) were below the numbers needed in this study to detect statistical significance. Therefore, the normalized data of validation cohort 2 was pooled with validation cohort 1 to increase the size of the study to detect statistical significance if present. Among high RS cases of the pooled dataset, a significantly different relapse rate was found between the ACES predicted treatment sensitive and insensitive strata ($p=0.033$). Among Luminal B cases, the sample size of 51 and event size of 8 remained too small for detection of statistical significance even after pooling the validation cohorts; the trend of these Luminal B cases toward higher residual risk of relapse in ACES predicted treatment insensitive cases was seen.

For high GGI cases of the pooled validation cohorts, which had the highest sample size and event rates, significance was not reached. This may be related to two other biases in assessing the performance of ACES in validation cohort 2. One bias may be secondary to the chemotherapeutic regimen used. The ACES algorithm has previously been shown to predict well only in taxane-anthracycline containing regimens, and not to be effective in cisplatin treated patients [42]. All patients in the discovery and validation cohort 1 were uniformly treated with a combined anthracycline and taxane regimen, in

addition to other chemotherapeutic agents. In the ER positive/HER2 negative cases of validation cohort 2, only 22% received combined taxane and anthracycline; 68% had an anthracycline but not taxane, and 10% did not receive either of these two classes of chemotherapy.

A second potential bias in assessing the performance of ACES in validation cohort 2 is that the outcome in this cohort was any form of relapse, which includes a local, regional or distant relapse. In the discovery and validation cohort 1, the outcome assessed was distant relapse. The ACES algorithm was developed to predict distant relapse, not local or regional ones [42] The molecular and pathologic drivers for local and regional relapses are typically different and may relate predominately to local disease control factors, such as surgical margins and radiation therapy.

Despite these biases, ACES demonstrated a relatively high NPV in predicting RFS [71-100% (95% CI: 48-100%)]. The clinical relevance of this thesis is that upon adequate validation of ACES: (i) patients with low residual risk by ACES can be safely treated with current adjuvant chemotherapies and reassured about their prognosis, (ii) patients who remain at substantial risk for relapse or death despite receiving the current standard of care adjuvant therapies, we be advised to seek out clinical trials that aim to improve the efficacy of current therapies. Further molecular characterization of these truly high risk cancers could also lead to the discovery of new drug targets for the very patient population who needs novel therapies in order to improve their survival [3, 24, 48]. Costs may be potentially saved if research trials are recruiting and enrolling only the specific subset of patients for whom improvement in therapy is needed. Costs may also

be reduced if patients are not being over-treated with therapeutic regimens of limited value to them [49].

This thesis placed the ACES gene signature into the context of commonly used clinical predictors of prognostic risk – the Oncotype DX as well as PAM50 and GGI assays for comparison – and utilized three independent cohorts to attempt to generate evidence. This study provides the initial evidence to suggest that the ACES algorithm may further risk stratify high RS and high GGI tumors into those with low and high residual risk after adjuvant chemotherapy and endocrine. For Luminal B cases, trends consistently showed the ability to ACES to identify those patients who remain at high risk, and ACES was significantly associated with RFS after controlling for other prognostic clinical variables. Before ACES can be adopted for routine use it would require independent validation in an adequately powered prospective trial and require adoption of the technology into a certified[9] clinical molecular pathology laboratory with standardized operating procedure for the test. The results presented in this thesis suggest that future validation of the ACES algorithm as residual risk prediction tool should be pursued.

REFERENCES

1. Sotiriou, C. and L. Pusztai, *Gene-expression signatures in breast cancer.* N Engl J Med, 2009. **360**(8): p. 790-800.
2. Gradishar, W.e.a. *NCCN Clinical Practice Guidelines in Oncology Breast Cancer*. NCCN Guidelines, 2013. **Version 1.2014**, 180.
3. Harbeck, N., et al., *Molecular and protein markers for clinical decision making in breast cancer: Today and tomorrow.* Cancer Treat Rev, 2013.
4. Kennecke, H., et al., *Metastatic behavior of breast cancer subtypes.* J Clin Oncol, 2010. **28**(20): p. 3271-7.

[9] Certified by governmental Clinical Laboratory Improvement Amendments

5. Edge, S.B., D.R.; Compton, C.C.; Fritz, A.G.; Greene, F.L.; Trotti, A., *AJCC Cancer Staging Manual*. Vol. 7th ed. . 2010: Springer. 649.
6. Paik, S., et al., *A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer.* N Engl J Med, 2004. **351**(27): p. 2817-26.
7. Hayes, D.F., B. Trock, and A.L. Harris, *Assessing the clinical impact of prognostic factors: when is "statistically significant" clinically useful?* Breast Cancer Res Treat, 1998. **52**(1-3): p. 305-19.
8. Harris, L., et al., *American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer.* J Clin Oncol, 2007. **25**(33): p. 5287-312.
9. Goldhirsch, A., et al., *Thresholds for therapies: highlights of the St Gallen International Expert Consensus on the primary therapy of early breast cancer 2009.* Ann Oncol, 2009. **20**(8): p. 1319-29.
10. Goldstein, L.J., et al., *Prognostic utility of the 21-gene assay in hormone receptor-positive operable breast cancer compared with classical clinicopathologic features.* J Clin Oncol, 2008. **26**(25): p. 4063-71.
11. Bueno-de-Mesquita, J.M., et al., *Use of 70-gene signature to predict prognosis of patients with node-negative breast cancer: a prospective community-based feasibility study (RASTER).* Lancet Oncol, 2007. **8**(12): p. 1079-87.
12. Kelly, C.M., et al., *Utility of oncotype DX risk estimates in clinically intermediate risk hormone receptor-positive, HER2-normal, grade II, lymph node-negative breast cancers.* Cancer, 2010. **116**(22): p. 5161-7.
13. Buyse, M., et al., *Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer.* J Natl Cancer Inst, 2006. **98**(17): p. 1183-92.
14. van 't Veer, L.J., et al., *Gene expression profiling predicts clinical outcome of breast cancer.* Nature, 2002. **415**(6871): p. 530-6.
15. Parker, J.S., et al., *Supervised risk predictor of breast cancer based on intrinsic subtypes.* J Clin Oncol, 2009. **27**(8): p. 1160-7.
16. Liedtke, C., et al., *Genomic grade index is associated with response to chemotherapy in patients with breast cancer.* J Clin Oncol, 2009. **27**(19): p. 3185-91.
17. Hess, K.R., et al., *Pharmacogenomic predictor of sensitivity to preoperative chemotherapy with paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide in breast cancer.* J Clin Oncol, 2006. **24**(26): p. 4236-44.
18. Iwamoto, T., et al., *First generation prognostic gene signatures for breast cancer predict both survival and chemotherapy sensitivity and identify overlapping patient populations.* Breast Cancer Res Treat, 2011. **130**(1): p. 155-64.
19. Tordai, A., et al., *Evaluation of biological pathways involved in chemotherapy response in breast cancer.* Breast Cancer Res, 2008. **10**(2): p. R37.
20. Sotiriou, C., et al., *Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis.* J Natl Cancer Inst, 2006. **98**(4): p. 262-72.
21. Metzger-Filho, O., et al., *Genomic grade adds prognostic value in invasive lobular carcinoma.* Ann Oncol, 2013. **24**(2): p. 377-84.
22. Metzger-Filho, O., et al., *Genomic Grade Index (GGI): feasibility in routine practice and impact on treatment decisions in early breast cancer.* PLoS One, 2013. **8**(8): p. e66848.
23. Metzger Filho, O., M. Ignatiadis, and C. Sotiriou, *Genomic Grade Index: An important tool for assessing breast cancer tumor grade and prognosis.* Crit Rev Oncol Hematol, 2011. **77**(1): p. 20-9.
24. Eroles, P., et al., *Molecular biology in breast cancer: intrinsic subtypes and signaling pathways.* Cancer Treat Rev, 2012. **38**(6): p. 698-707.

25. Nielsen, T.O., et al., *A comparison of PAM50 intrinsic subtyping with immunohistochemistry and clinical prognostic factors in tamoxifen-treated estrogen receptor-positive breast cancer.* Clin Cancer Res, 2010. **16**(21): p. 5222-32.
26. Chia, S.K., et al., *A 50-gene intrinsic subtype classifier for prognosis and prediction of benefit from adjuvant tamoxifen.* Clin Cancer Res, 2012. **18**(16): p. 4465-72.
27. Bastien, R.R., et al., *PAM50 breast cancer subtyping by RT-qPCR and concordance with standard clinical molecular markers.* BMC Med Genomics, 2012. **5**: p. 44.
28. Martin, M., et al., *PAM50 proliferation score as a predictor of weekly paclitaxel benefit in breast cancer.* Breast Cancer Res Treat, 2013. **138**(2): p. 457-66.
29. Ward, S., et al., *Gene expression profiling and expanded immunohistochemistry tests to guide the use of adjuvant chemotherapy in breast cancer management: a systematic review and cost-effectiveness analysis.* Health Technol Assess, 2013. **17**(44): p. 1-302.
30. Paik, S., et al., *Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer.* J Clin Oncol, 2006. **24**(23): p. 3726-34.
31. Habel, L.A., et al., *A population-based study of tumor gene expression and risk of breast cancer death among lymph node-negative patients.* Breast Cancer Res, 2006. **8**(3): p. R25.
32. Dowsett, M., et al., *Prediction of risk of distant recurrence using the 21-gene recurrence score in node-negative and node-positive postmenopausal patients with breast cancer treated with anastrozole or tamoxifen: a TransATAC study.* J Clin Oncol, 2010. **28**(11): p. 1829-34.
33. Dowsett, M., et al., *Comparison of PAM50 risk of recurrence score with oncotype DX and IHC4 for predicting risk of distant recurrence after endocrine therapy*, in *J Clin Oncol*. 2013. p. 2783-90.
34. Sparano, J.A. and S. Paik, *Development of the 21-gene assay and its application in clinical practice and clinical trials.* J Clin Oncol, 2008. **26**(5): p. 721-8.
35. Gianni, L., et al., *Gene expression profiles in paraffin-embedded core biopsy tissue predict response to chemotherapy in women with locally advanced breast cancer.* J Clin Oncol, 2005. **23**(29): p. 7265-77.
36. Chang, J.C., et al., *Gene expression patterns in formalin-fixed, paraffin-embedded core biopsies predict docetaxel chemosensitivity in breast cancer patients.* Breast Cancer Res Treat, 2008. **108**(2): p. 233-40.
37. Albain, K.S., et al., *Prognostic and predictive value of the 21-gene recurrence score assay in postmenopausal women with node-positive, oestrogen-receptor-positive breast cancer on chemotherapy: a retrospective analysis of a randomised trial.* Lancet Oncol, 2010. **11**(1): p. 55-65.
38. *TAILORx. ClinicalTrials.gov Identifier: NCT00310180.*; Available from: http://clinicaltrials.gov/ct2/show/NCT00310180?term=TailoRx&rank=1.
39. *RxPONDER. ClinicalTrials.gov Identifier: NCT01272037*. Available from: http://clinicaltrials.gov/ct2/show/NCT01272037?term=NCT01272037&rank=1.
40. Naoi, Y., et al., *High genomic grade index associated with poor prognosis for lymph node-negative and estrogen receptor-positive breast cancers and with good response to chemotherapy.* Cancer, 2011. **117**(3): p. 472-9.
41. Criscitiello, C., et al., *High Ki-67 score is indicative of a greater benefit from adjuvant chemotherapy when added to endocrine therapy in Luminal B HER2 negative and node-positive breast cancer.* Breast, 2014. **23**(1): p. 69-75.
42. Hatzis, C., et al., *A genomic predictor of response and survival following taxane-anthracycline chemotherapy for invasive breast cancer.* JAMA, 2011. **305**(18): p. 1873-81.

43. Symmans, W.F., et al., *Genomic index of sensitivity to endocrine therapy for breast cancer.* J Clin Oncol, 2010. **28**(27): p. 4111-9.
44. Group., G.B. *GeparTrio.*; Available from: http://www.germanbreastgroup.de/studien/neoadjuvant/gepartrio/english-summary-.html.
45. Gentleman, R.e.a., *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. 2005: Springer. 473.
46. Haibe-Kains, e.a. "*genefu: a package for breast cancer gene expression analysis*", *Bioconductor package v.2.13*. 2013; Available from: http://www.bioconductor.org/packages/2.13/bioc/vignettes/genefu/inst/doc/genefu.pdf.
47. Consortium, M., et al., *The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements.* Nat Biotechnol, 2006. **24**(9): p. 1151-61.
48. Pusztai, L.H., C., *Breast cancer biomarkers beyond molecular classification.*, Y.U.S.o. Medicine, Editor. 2013. p. 6.
49. Simon, R., *Roadmap for developing and validating therapeutically relevant genomic classifiers.* J Clin Oncol, 2005. **23**(29): p. 7332-41.

APPENDIX

TABLE A1. Multivariate Cox Proportional Hazards Analysis of Association with Relapse Free Survival

High Risk Group	**Discovery Cohort**		**Validation Cohort 1**		**Validation Cohort 2†**		**Validation Cohort 1 and 2†**	
	HR (95% CI)	p-value	HR (95% CI)	p-value	HR (95% CI)	p-value	HR (95% CI)	p-value
High RS								
Age	1.45 (0.4-6.9)	0.604	0.45 (0.1-2.0)	0.298	3.38 (0.6-1.8)	0.160	1.53 (0.6-4.3)	0.415
T-stage	2.46 (0.5-11)	0.249	2.13 (0.4-1.2)	0.381	**3E-9 (6E-10-1E-8)**	**2E-16**	1.46 (0.4-5.0)	0.554
Node	1.17 (0.2-6.8)	0.863	3.66 (0.5-50)	0.326	**9.07 (1.1-7.6)**	**0.042**	**5.35 (1.0-28)**	**0.047**
ACES	0.16 (0.0-2.3)	0.174	**2E-9(6E-10-8E-9)**	**2E-16**	0.31 (0.0-2.6)	0.280	**0.15 (0.0-1.2)**	**0.072**
Luminal B								
Age	1.11 (0.3-4.2)	0.879	0.22 (0.0-2.7)	0.239	3.18 (0.3-32)	0.330	0.95 (0.2-4.1)	0.950
T-stage	0.65 (0.1-6.6)	0.715	2.75 (0.07-9.5)	0.521	0.65 (0.1-6.5)	0.720	0.98 (0.2-5.2)	0.980
Node	1.31 (0.1-18)	0.836	0.48 (0.0-1.0)	0.641	3.96 (0.5-29)	0.180	1.6 (0.3-9.9)	0.620
ACES	0.81 (0.16-4.1)	0.802	**3E-9(2E-10-5E-8)**	**2E-16**	**0.15 (0.0-0.7)**	**0.013**	0.4 (0.1-2.4)	0.310
High GGI								
Age	1.25 (0.5-3.4)	0.662	0.52 (0.14-1.9)	0.318	1.27 (0.5-0.6)	0.691	0.77 (0.3-1.7)	0.530
T-stage	1.55 (0.5-4.4)	0.414	4.52 (0.6-3.4)	0.143	1.33 (0.3-5.5)	0.694	0.70 (0.3-1.6)	0.396
Node	2.28 (0.6-9.3)	0.252	1.41 (0.2-8.1)	0.702	**3.70 (1.0-13)**	**0.042**	**2.7 (0.94-7.8)**	**0.066**
ACES	**0.29 (0.1-0.8)**	**0.017**	**4E-9 (1E-9-2E-8)**	**2E-16**	0.58 (0.2-2.0)	0.386	0.78 (0.3-2.1)	0.619

Age, T-stage, Node and ACES were binary variables where 1 was defined as age greater than or equal to 50 (versus less than 50), T-stage of T3/T4 (versus T1/T2), node positive (versus negative) at time of diagnosis, and ACES predicted treatment sensitive (versus insensitive).
†Normalized using stratification of cohort by T-stage (See Strategy 2 in Methods).
Abbreviations: HR, Hazard Ratio; RS, recurrence score; GGI, genomic grade index

TABLE A2. Distribution of sub-components of ACES algorithm

Group	Discovery Cohort	Validation Cohort 1	Validation Cohort 2†
High RS	N (% of High RS cases)		
SET-High/Intermediate	4 (11)	2 (5)	5 (14)
Predicted early relapse/death	13 (33)*	19 (49)	15 (41)
Predicted RCB-III	13 (33)	19 (49)	11 (30)
Predicted RCB-0/I	14 (36)	20 (51)	20 (54)
ACES Rx Sensitive	14 (36)	11 (28)	13 (35)
Luminal B	N (% of Luminal B cases)		
SET-High/Intermediate	1 (3)*	0*	2 (11)
Predicted early relapse/death	5 (13)	13 (39)	8 (44)
Predicted RCB-III	12 (30)	13 (39)	7 (39)
Predicted RCB-0/I	16 (40)	17 (52)	10 (56)
ACES Rx Sensitive	15 (38)	8 (24)	6 (33)
High GGI	N (% of High GGI cases)		
SET-High/Intermediate	16 (18)	0*	10 (19)
Predicted early relapse/death	17 (19)	28 (44)	20 (37)
Predicted RCB-III	27 (30)*	28 (44)	18 (33)
Predicted RCB-0/I	39 (43)	36 (56)	28 (52)
ACES Rx Sensitive	45 (50)	14 (22)	22 (41)
All ER positive/HER2 negative	N (% of all ER positive/HER2 negative cases)		
SET-High/Intermediate	45 (26)	14 (11)	36 (28)
Predicted early relapse/death	28 (16)	40 (33)	31 (24)
Predicted RCB-III	79 (45)	58 (47)	54 (43)
Predicted RCB-0/I	64 (36)	57 (46)	55 (43)
ACES Rx Sensitive	80 (45)	37 (30)	54 (43)

*Fisher's exact test shows that there is a statistically significant difference between this value and the corresponding one in the entire ER positive/HER2 negative cohort.

†Normalized using stratification of cohort by T-stage (See Strategy 2 in Methods).

Abbreviations: RS, Recurrence Score; GGI, Genomic Grade Index; SET, Sensitivity to Endocrine Therapy; RCB, Residual Cancer Burden; ER, Estrogen Receptor; HER2, Human Epidermal Growth Factor Receptor

TABLE A3: Performance of RCB-0/I and RCB-III gene signatures in predicting actual pathologic response

Risk Category	**RCB-0/I Prediction Discovery Cohort** % (95% CI)	**RCB-0/I Prediction Validation Cohort 1** % (95% CI)	**RCB-III Prediction Discovery Cohort** % (95% CI)	**RCB-III Prediction Validation Cohort 1** % (95% CI)
High RS				
Sensitivity	71 (42-92)	40 (19-64)	80 (28-99)	67 (9-99)
Specificity	84 (64-95)	18 (2-52)	74 (56-87)	54 (34-72)
PPV	71 (42-92)	47 (23-72)	31 (9-61)	13 (2-40)
NPV	84 (64-95)	14 (2-43)	96 (80-100)	94 (70-100)
LR+	4.5 (0.1-undefined)	0.5 (0-0.99)	3.0 (0.02-undefined)	1.4 (0.005-undefined)
LR-	**0.3 (0-0.99)**	3.5 (0.004-undefined)	**0.3 (0-0.99)**	0.6 (0-1)
OR	13.1 (undefined)	0.1 (0-0.8)	11.1 (undefined)	2.3 (0.3-undefined)
Luminal B				
Sensitivity	71 (42-92)	46 (19-75)	100 (69-100)	100 (16-100)
Specificity	76 (55-91)	60 (15-95)	97 (82-100)	69 (41-89)
PPV	63 (35-85)	75 (35-97)	91 (59-100)	29 (4-71)
NPV	83 (61-95)	30 (7-65)	100 (88-100)	100 (72-100)
LR+	3.0 (0.01-undefined)	1.2 (0-undefined)	29 (52-undefined)	3.2 (0.08-undefined)
LR-	**0.4 (0-0.99)**	0.9 (0-1)	**0 (0.00-0.98)**	**0 (0.00-0.995)**
OR	7.9 (undefined)	1.3 (0.05-undefined)	Infinite (undefined)	Infinite (undefined)
High GGI				
Sensitivity	74 (55-88)	65 (44-83)	94 (71-100)	100 (48-100)
Specificity	73 (60-84)	54 (25-81)	89 (79-95)	68 (49-83)
PPV	61 (43-76)	73 (52-90)	67 (45-84)	31 (11-59)
NPV	84 (70-93)	44 (20-70)	98 (91-100)	100 (85-100)
LR+	2.8 (0.01-undefined)	1.4 (0.005-undefined)	8.2 (0.12-undefined)	3.1 (0.08-undefined)
LR-	**0.4 (0-0.99)**	0. (0-1)	0.07 (undefined-0.99)	**0 (0.00-0.996)**
OR	7.9 (undefined)	2.2 (0.2-undefined)	124 (undefined)	Infinite (undefined)

Abbreviations: RS, Recurrence Score; GGI, Genomic Grade Index; RCB-0/I, Pathologic Complete Response or Minimal Residual Cancer Burden; RCB-III, Extensive Residual Cancer Burden

TABLE A4. Normalization of Validation Cohort 2, by stratification with prognostic factors and adjusting ACES subscore thresholds

Characteristic	Discovery Cohort and Validation Cohort 1	Validation Cohort 2	Odds Ratio	p-value
T1 and T2	**N (% of 182)**	**N (% of 112)**		
SET-Low	140 (77)	33 (29) → 34 (30)	2.6 → 2.5	**0.000→ 0.000**
SET-Intermediate/High	42 (23)	79 (71) → 78 (69)	0.3 → 0.3	**0.000→ 0.000**
High Risk early death/relapse	32 (18)	30 (27) → 27 (24)	0.7 → 0.7	**0.154** → 0.307
Low Risk early death/relapse	150 (82)	82 (73) → 85 (76)	1.1 → 1.1	0.526 → 0.652
RCB-III	83 (46)	57 (51) → 45 (40)	0.9 → 1.1	0.601 → 0.586
Not RCB-III	99 (54)	55 (49) → 67 (60)	1.1 → 0.9	0.681 → 0.691
RCB-0/I	65 (36)	69 (62) → 50 (45)	0.0 → 0.8	0.580 → 0.368
Not RCB-0/I	117 (64)	43 (38) → 62 (55)	1.7 → 1.2	**0.017** → 0.492
T3 and T4	**N (% of 115)**	**N (% of 15)**		
SET-Low	99 (86)	9 (60) → 13 (87)	1.4 → 0.99	0.518 → 1
SET-Intermediate/High	16 (14)	6 (40) → 2 (13)	0.4 → 1.0	0.086 → 1
High Risk early death/relapse	36 (31)	5 (33) → 4 (27)	0.9 → 1.2	1 → 1
Low Risk early death/relapse	79 (69)	10 (67) → 11 (73)	1.0 → 0.9	1 → 1
RCB-III	52 (45)	9 (60) → 9 (60)	0.8 → 0.8	0.640 → 0.640
Not RCB-III	63 (55)	6 (40) → 6 (40)	1.4 → 1.4	0.632 → 0.633
RCB-0/I	55 (48)	10 (67) → 5 (33)	0.7 → 1.4	0.498 → 0.616
Not RCB-0/I	60 (52)	5 (33) → 10 (67)	1.6 → 0.8	0.464 → 0.655

"→" represents values after normalization of data by stratification with T-stage, lymph node involvement and grade, respectively within validation cohort 2. The Fisher's exact test was used to compare proportions of cases within the cohorts before and after normalization, and the resulting odds ratio and p-value are reported. Age (less than 50 versus greater than or equal to 50), T-stage (T1/T2 versus T3/T4), lymph node status (positive versus negative) and grade (1 or 2 versus 3) are at time of diagnosis.

Abbreviations: SET, Sensitivity to Endocrine Therapy; RCB, Residual Cancer Burden

TABLE A4 (continued). Normalization of Validation Cohort 2, by stratification with prognostic factors and adjusting ACES subscore thresholds

Characteristic	Discovery Cohort and Validation Cohort 1	Validation Cohort 2	Odds Ratio	p-value
Node Negative	**N (% of 106)**	**N (% of 69)**		
SET-Low	81 (76)	19 (28) → 49 (71)	2.8 → 1.1	**0.000** → 0.812
SET-Intermediate/High	25 (24)	50 (72) → 20 (29)	0.3 → 0.8	**0.000**→ 0.613
High Risk early death/relapse	19 (18)	16 (23) → 11 (16)	0.8 → 1.1	0.572 → 0.842
Low Risk early death/relapse	87 (82)	53 (77) → 58 (84)	1.1 → 1.0	0.816 → 1
RCB-III	40 (38)	35 (51) → 30 (43)	0.7 → 0.9	0.328 → 0.667
Not RCB-III	66 (62)	34 (49) → 39 (57)	1.3 → 1.1	0.437 → 0.800
RCB-0/I	45 (42)	45 (65) → 32 (46)	0.7 → 0.9	0.116 → 0.781
Not RCB-0/I	61 (58)	24 (35) → 37 (54)	1.7 → 0.7	0.098 → 0.797
Node Positive	**N (% of 190)**	**N (% of 57)**		
SET-Low	157 (83)	20 (35) → 43 (75)	2.4 → 1.1	**0.002** → 0.733
SET-Intermediate/High	33 (17)	37 (65) → 14 (25)	0.3 → 0.7	**0.000** → 0.354
High Risk early death/relapse	49 (26)	19 (33) → 16 (28)	0.8 → 0.9	0.425 → 0.869
Low Risk early death/relapse	141 (74)	38 (67) → 41 (72)	1.1 → 1.0	0.724 → 0.908
RCB-III	94 (49)	31 (54) → 24 (42)	0.9 → 1.2	0.701 → 0.592
Not RCB-III	96 (51)	26 (46) → 33 (58)	1.1 → 0.9	0.791 → 0.612
RCB-0/I	75 (39)	33 (58) → 19 (33)	0.7 → 1.2	0.146 → 0.663
Not RCB-0/III	115 (61)	24 (42) → 38 (67)	1.4 → 0.9	0.195 → 0.718

"→" represents values after normalization of data by stratification with T-stage, lymph node involvement and grade, respectively within validation cohort 2. The Fisher's exact test was used to compare proportions of cases within the cohorts before and after normalization, and the resulting odds ratio and p-value are reported. Age (less than 50 versus greater than or equal to 50), T-stage (T1/T2 versus T3/T4), lymph node status (positive versus negative) and grade (1 or 2 versus 3) are at time of diagnosis. Abbreviations: SET, Sensitivity to Endocrine Therapy; RCB, Residual Cancer Burden.

TABLE A4 (continued). Normalization of Validation Cohort 2, by stratification with prognostic factors and adjusting ACES subscore thresholds

Characteristic	Discovery Cohort and Validation Cohort 1	Validation Cohort 2	Odds Ratio	p-value
Grade 1 and 2	**N (% of 182)**	**N (% of 95)**		
SET-Low	136 (75)	25 (26) → 77 (81)	2.8 → 0.9	**0.000** → 0.703
SET-Intermediate/High	46 (25)	70 (74) → 18 (19)	0.3 → 1.3	**0.000** → 0.380
High Risk early death/relapse	30 (16)	20 (21) → 11 (12)	0.8 → 1.4	0.520 → 0.380
Low Risk early death/relapse	152 (84)	75 (79) → 84 (88)	1.1 → 0.9	0.777 → 0.781
RCB-III	90 (49)	58 (61) → 64 (67)	0.8 → 0.7	0.341 → 0.145
Not RCB-III	92 (51)	37 (39) → 31 (33)	1.3 → 1.5	0.306 → 0.080
RCB-0/I	62 (34)	57 (60) → 36 (38)	0.6 → 0.9	**0.013** → 0.712
Not RCB-0/I	120 (66)	38 (40) → 59 (62)	1.7 → 1.1	**0.030** → 0.839
Grade 3	**N (% of 100)**	**N (% of 32)**		
SET-Low	93 (93)	14 (44) → 26 (81)	2.1 → 1.1	**0.033** → 0.765
SET-Intermediate/High	7 (7)	18 (56) → 6 (19)	0.1 → 0.4	**0.000** → 0.103
High Risk early death/relapse	34 (34)	15 (47) → 13 (41)	0.7 → 0.8	0.446 → 0.697
Low Risk early death/relapse	66 (66)	17 (53) → 19 (59)	1.2 → 1.1	0.617 → 0.870
RCB-III	40 (40)	8 (25) → 10 (31)	1.6 → 1.3	0.317 → 0.694
Not RCB-III	60 (60)	24 (75) → 22 (69)	0.8 → 0.9	0.526 → 0.747
RCB-0/I	54 (54)	22 (69) → 16 (50)	0.8 → 1.1	0.512 → 0.864
Not RCB-0/I	46 (46)	10 (31) → 16 (50)	1.5 → 0.9	0.444 → 0.859

"→" represents values after normalization of data by stratification with T-stage, lymph node involvement and grade, respectively within validation cohort 2. The Fisher's exact test was used to compare proportions of cases within the cohorts before and after normalization, and the resulting odds ratio and p-value are reported. Age (less than 50 versus greater than or equal to 50), T-stage (T1/T2 versus T3/T4), lymph node status (positive versus negative) and grade (1 or 2 versus 3) are at time of diagnosis. Abbreviations: SET, Sensitivity to Endocrine Therapy; RCB, Residual Cancer Burden.

TABLE A5. Normalization of Validation Cohort 2, by multivariate linear regression model of prognostic factors

Multivariate Model	**p-value**				
	Age	**T-stage**	**Nodal Status**	**Grade**	**Cohort**
SET-Index = 1.3 –0.3(age) – 0.4(T-stage) – 0.1(nodal status) – 0.6(grade) + 0.7(cohort)	**0.020**	**0.013**	0.510	**0.000**	**0.000**
Early Death/Relapse Score = -0.6 + 0.03(age) + 0.1(T-stage) + 0.04(nodal) + 0.4(grade) + 024(cohort)	0.621	0.322	0.554	**0.000**	**0.000**
RCB-III Score = 0.6 – 0.3(age) + 0.5(T-stage) + 1.3(nodal status) – 2.6(grade) – 0.2(cohort)	0.550	0.399	**0.008**	**0.000**	0.426
RCB-0/I Score = -2.4 – 0.7(age) + 0.1(T-stage) – 2.0(nodal status) + 2.6(grade) + 1.5(cohort)	0.255	0.883	**0.003**	**0.000**	**0.000**

In this multivariate linear regression model, the dependent variable is quantitative and the dependent variables are binary: Age (less than 50 versus greater than or equal to 50), T-stage (T1/T2 versus T3/T4), lymph node status (positive versus negative) and grade (1 or 2 versus 3), all at time of diagnosis, and cohort (Discovery Cohort or Validation Cohort 1, versus Validation Cohort 2).

Abbreviations: SET, Sensitivity to Endocrine Therapy; RCB, Residual Cancer Burden.

TABLE A6. Survival Analysis of Validation Cohort 2

Normalization Method	**Log-Rank Test p-value**	**ACES Treatment Sensitive**		**ACES Treatment Insensitive**	
		No.	Event Rate (%)	No.	Event Rate (%)
Normalization Strategy 1 (Overall Matched Proportions)					
High RS	0.472	10	10	27	19
Luminal B	0.358	2	50	16	19
High GGI†	0.233	15	40	39	23
Normalization Strategy 2 (Stratification by T-stage)					
High RS	0.307	15	7	22	23
Luminal B	0.704	8	13	10	30
High GGI†	0.654	25	20	29	34
Normalization Strategy 3 (Stratification by Nodal Status)					
High RS	0.894	13	15	24	17
Luminal B	0.704	6	17	12	25
High GGI†	0.902	23	26	31	29
Normalization Strategy 4 (Stratification by Grade)					
High RS	0.755	10	20	27	15
Luminal B	0.631	3	33	15	20
High GGI†	0.245	15	40	39	23
Normalization Strategy 5 (Multivariate Model)					
High RS	0.931	12	17	25	16
Luminal B	0.295	3	0	15	27
High GGI†	0.989	17	29	37	27

†Normalized GGI.

Abbreviations: RS, Recurrence Score; GGI, Genomic Grade Index

FIGURE A1. Kaplan-Meier Estimates of Relapse Free Survival Stratified by ACES, in high risk cases of Validation Cohort 2

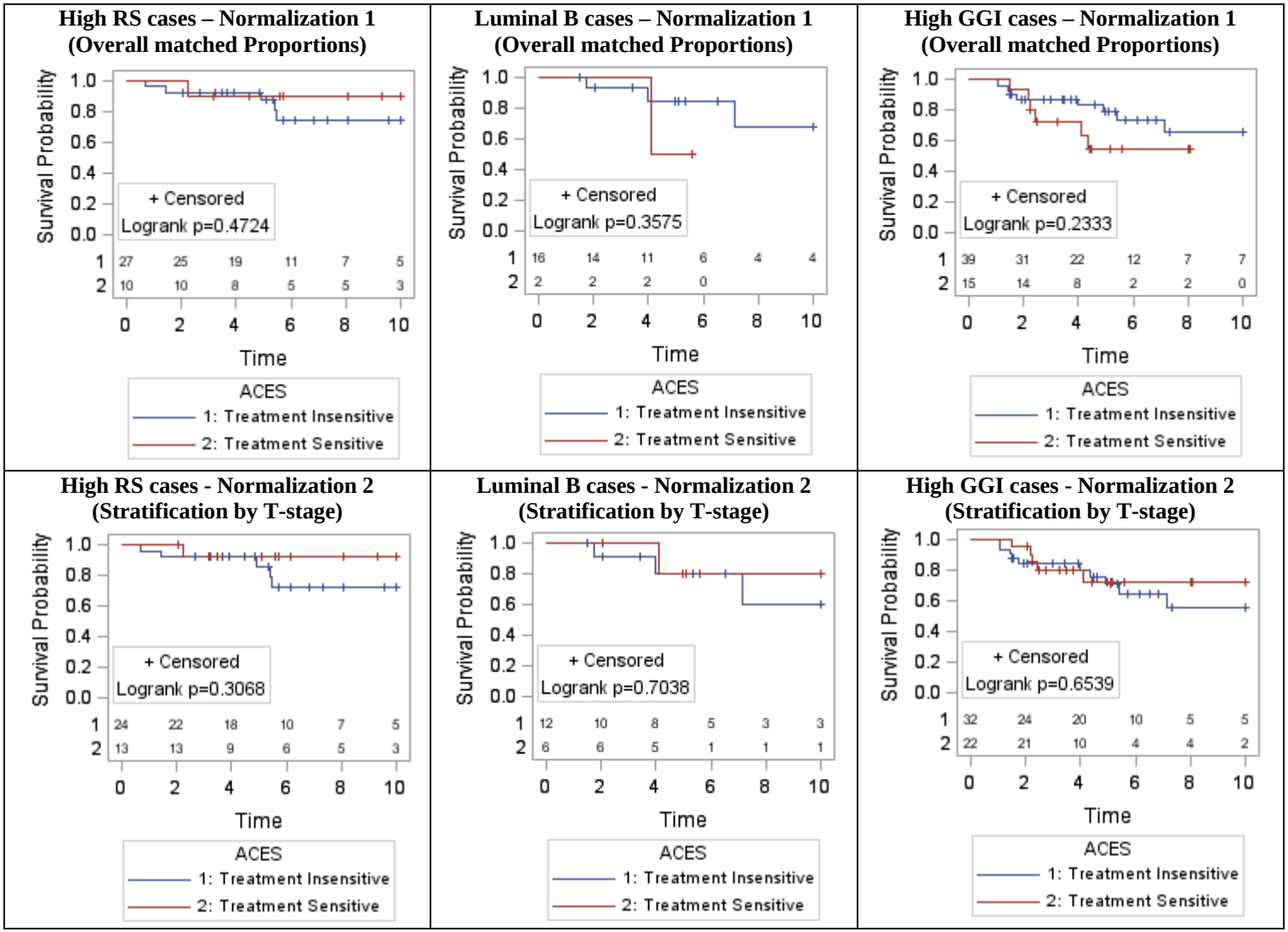


FIGURE A1 (continued). Kaplan-Meier Estimates of Relapse Free Survival Stratified by ACES in Validation Cohort 2

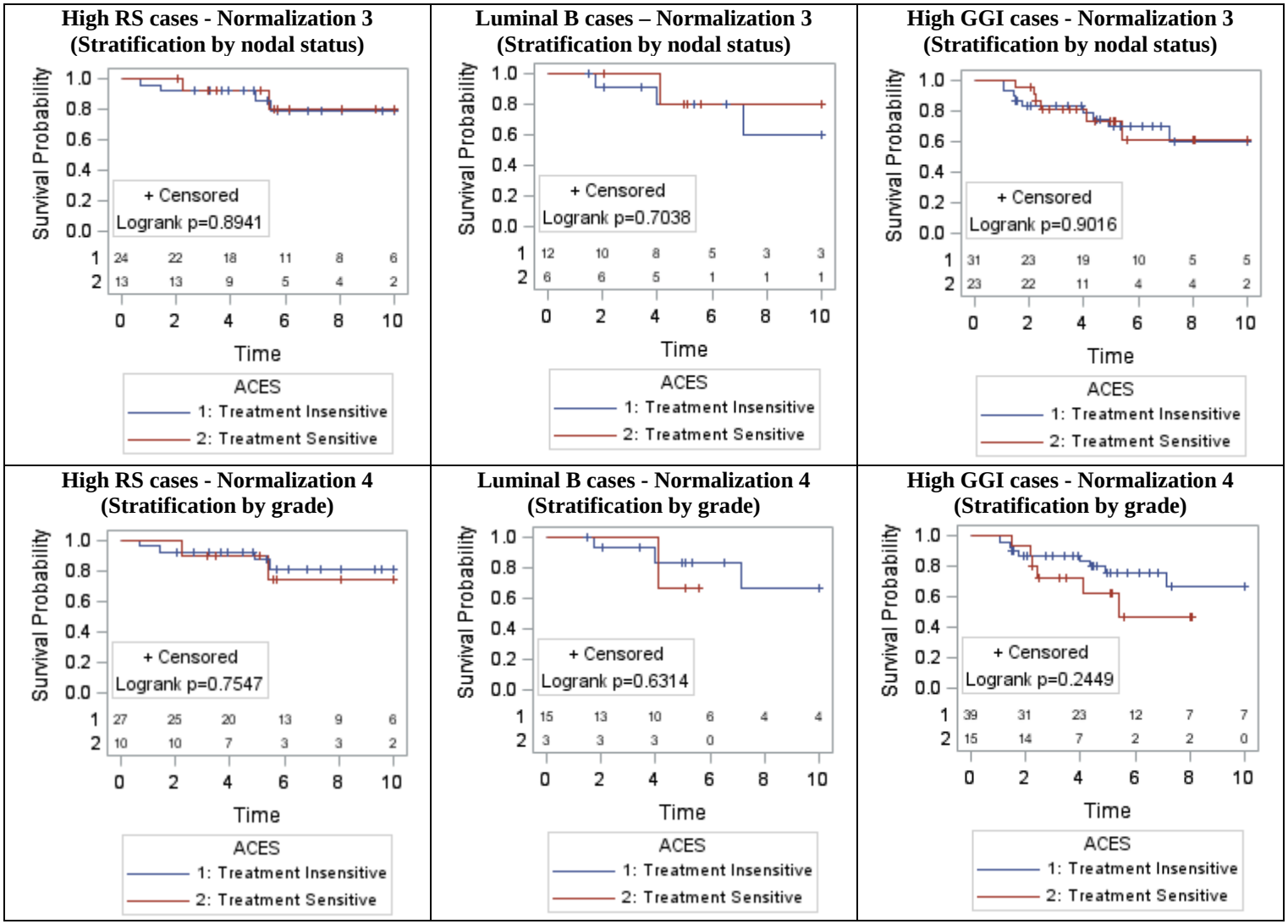


FIGURE A1 (continued). Kaplan-Meier Estimates of Relapse Free Survival Stratified by ACES in Validation Cohort 2

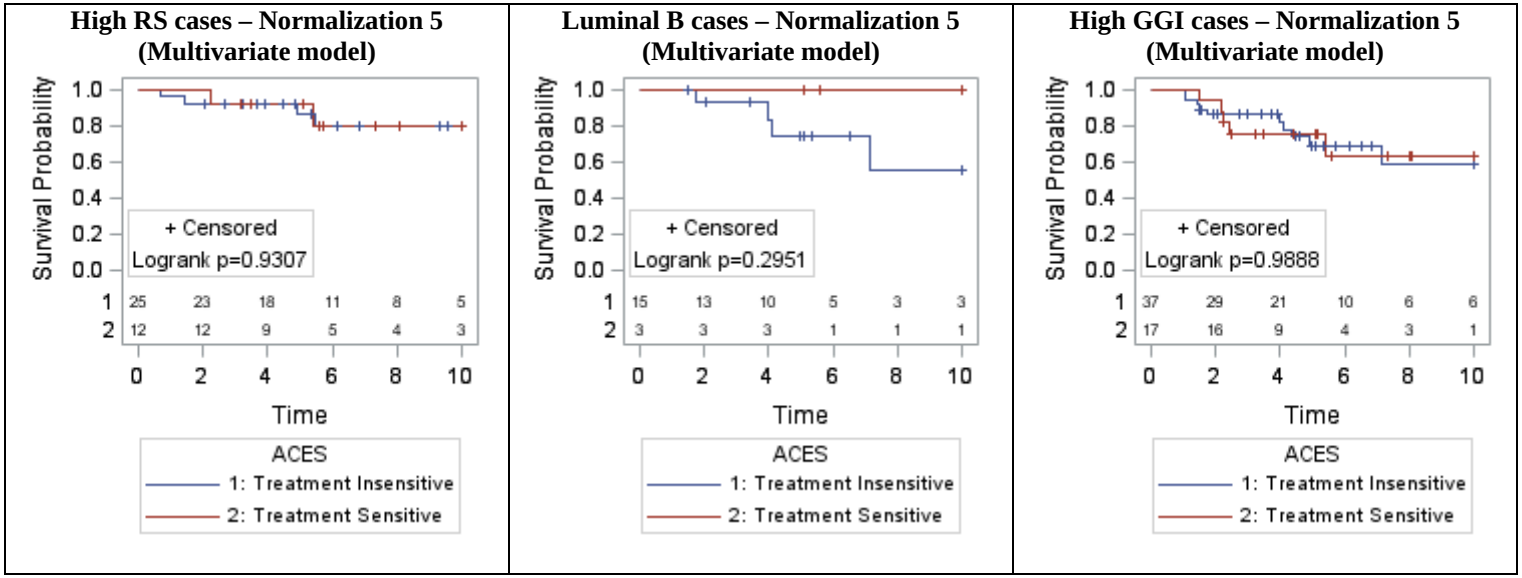


Figure A2. Kaplan Meier estimates of Relapse Free Survival in High RS cases stratified by ACES subcomponents

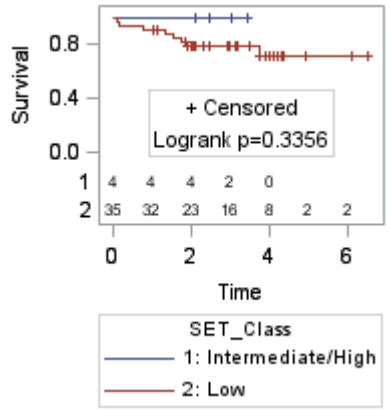
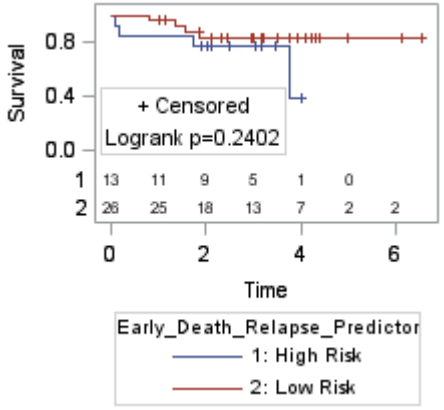
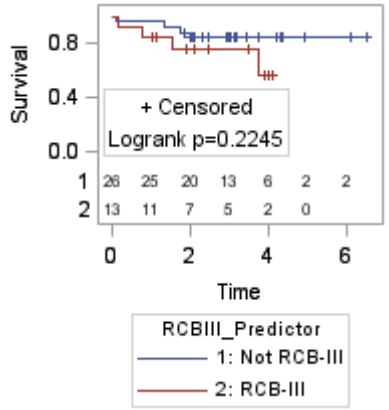
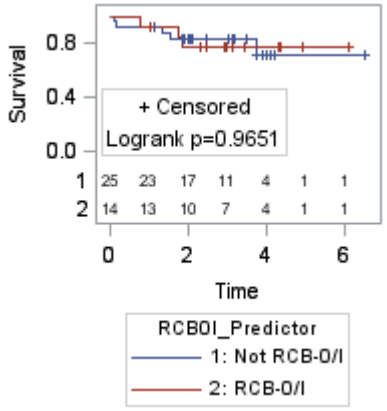
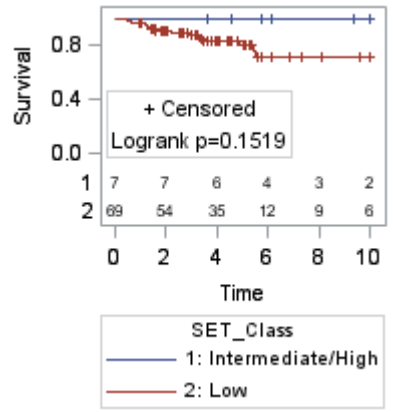
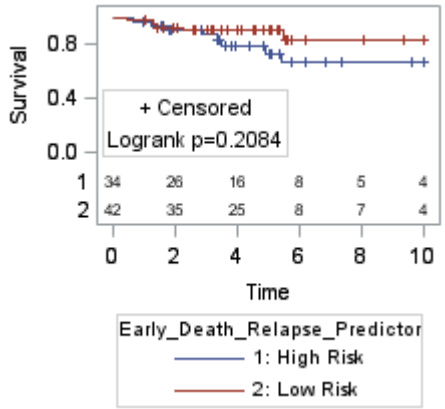
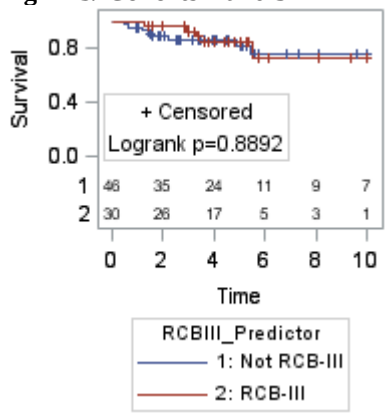
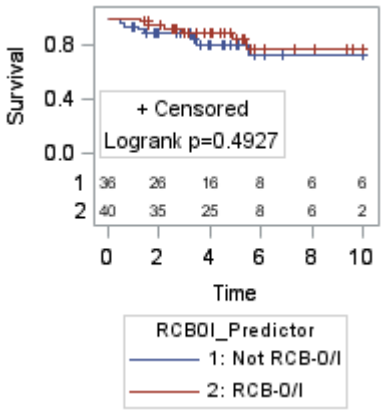
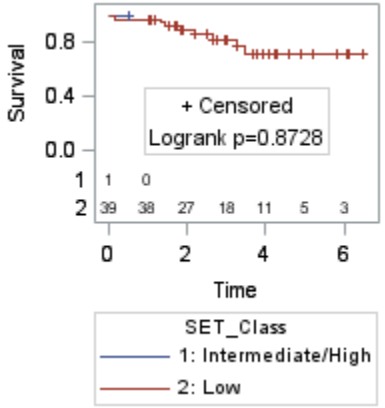
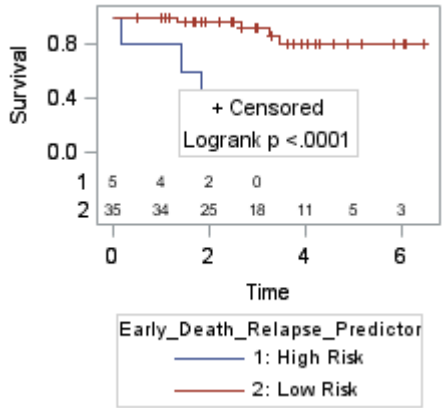
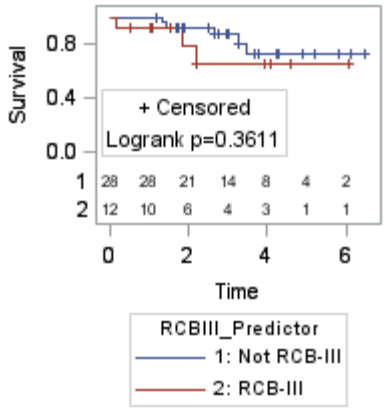
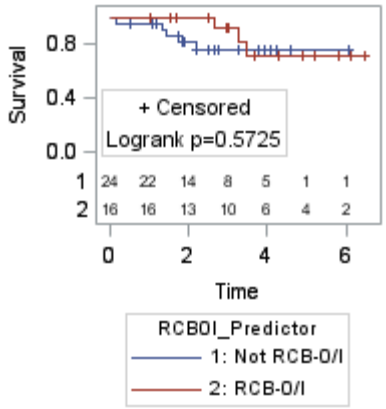
Stratified by SET Class	Stratified by Early Relapse/Death Predictor	Stratified by RCB-III Predictor	Stratified by RCB-0/I Predictor
High RS/ Cohort 1	**High RS/ Cohort 1**	**High RS/ Cohort 1**	**High RS/ Cohort 1**
			
High RS/ Cohorts 2 and 3	**High RS/ Cohorts 2 and 3**	**High RS/ Cohorts 2 and 3**	**High RS/ Cohorts 2 and 3**
			

Figure A2 (continued). Kaplan Meier estimates of Relapse Free Survival in Luminal B cases stratified by ACES subcomponents

Stratified by SET Class	**Stratified by Early Relapse/Death Predictor**	**Stratified by RCB-III Predictor**	**Stratified by RCB-0/I Predictor**
Luminal B/ Cohort 1	**Luminal B/ Cohort 1**	**Luminal B/ Cohort 1**	**Luminal B/ Cohort 1**
			
Luminal B/ Cohorts 2 and 3	**Luminal B/ Cohorts 2 and 3**	**Luminal B/ Cohorts 2 and 3**	**Luminal B/ Cohorts 2 and 3**

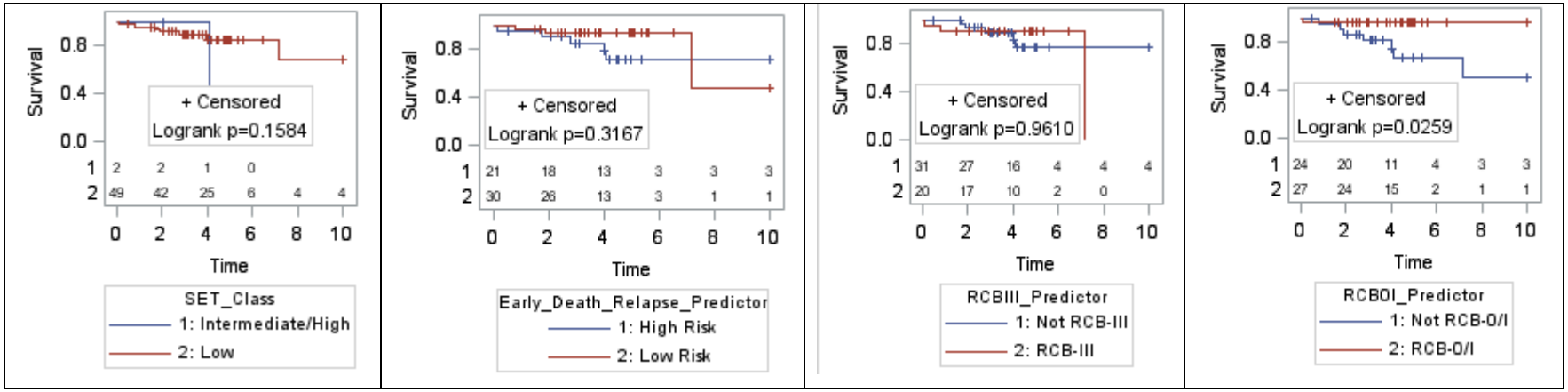
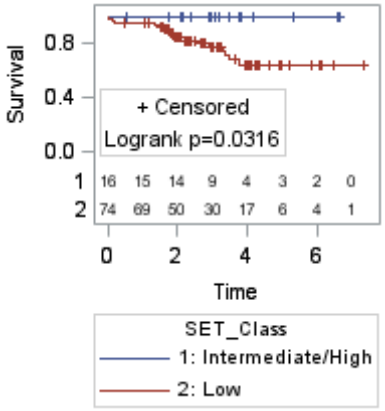
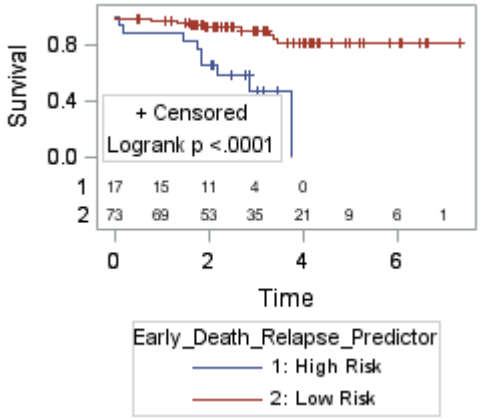
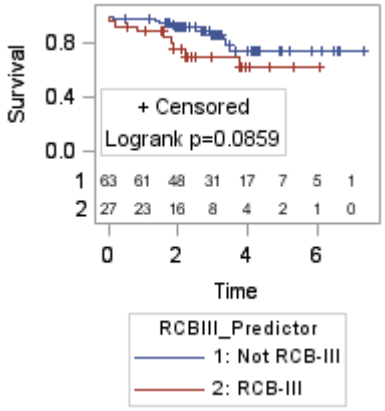
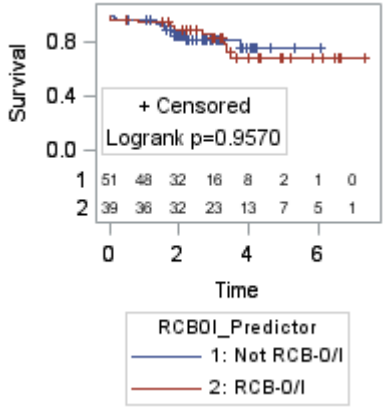
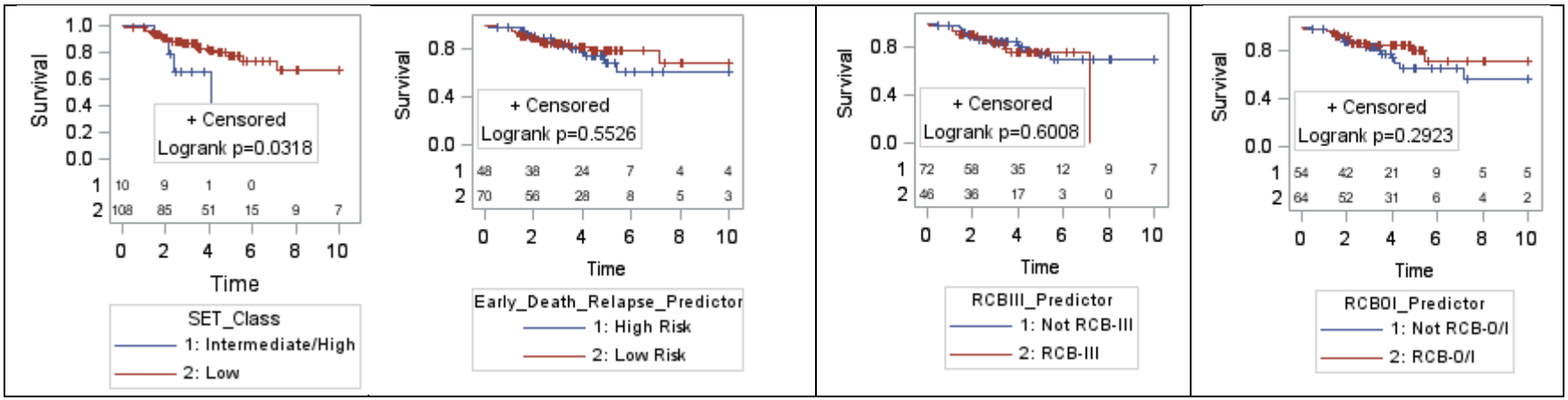


Figure A2 (continued). Kaplan Meier estimates of Relapse Free Survival in High GGI cases stratified by ACES subcomponents

Stratified by SET Class	**Stratified by Early Relapse/Death Predictor**	**Stratified by RCB-III Predictor**	**Stratified by RCB-0/I Predictor**
High GGI/ Cohort 1	**High GGI/ Cohort 1**	**High GGI/ Cohort 1**	**High GGI/ Cohort 1**
			
High GGI/ Cohorts 1 and 2	**High GGI/ Cohorts 1 and 2**	**High GGI/ Cohorts 1 and 2**	**High GGI/ Cohorts 1 and 2**



Abbreviations: SET, Sensitivity to Endocrine Therapy; RCB-0/I or RCB-III, Minimal or Extensive Residual Cancer Burden; RS, Recurrence Score; GGI, Genomic Grade Index

FIGURE A3. Early relapse predictor genes for Estrogen Receptor positive tumors in ACES algorithm

Probe Set	Symbol	Description	GeneID	Chromosome	Cytoband
212174_at	AK2	adenylate kinase 2	204	1	1p34
215407_s_at	ASTN2	astrotactin 2	23245	9	9q33.1
205626_s_at	CALB1	calbindin 1, 28kDa	793	8	8q21.3-q22.1
212816_s_at	CBS	cystathionine-beta-synthase	875	21	21q22.3
216923_at	CDLK5	cyclin-dependent kinase-like 5	6792	X	Xp22.13
205471_s_at	DACH1	dachshund homolog 1 (Drosophila)	1602	13	13q22
221681_s_at	DSPP	dentin sialophosphoprotein	1834	4	4q21.3
201539_s_at	FHL1	four and a half LIM domains 1	2273	X	Xq26
215744_at	FUS	fusion (involved in t(12;16) in malignant liposarcoma)	2521	16	16p11.2
209604_s_at	GATA3	GATA binding protein 3	2625	10	10p15
209602_s_at	GATA3	GATA binding protein 3	2625	10	10p15
209603_at	GATA3	GATA binding protein 3	2625	10	10p15
203821_at	HBEGF	heparin-binding EGF-like growth factor	1839	5	5q23
219976_at	HOOK1	hook homolog 1 (Drosophila)	51361	1	1p32.1
212531_at	LCN2	lipocalin 2	3934	9	9q34
220906_at	LDB2	LIM domain binding 2	9079	4	p15.32
217506_at	LOC339290	hypothetical LOC339290	339290	18	18p11.31
204058_at	ME1	malic enzyme 1, NADP(+)-dependent, cytosolic	4199	6	6q12
200899_s_at	MGEA5	meningioma expressed antigen 5 (hyaluronidase)	10724	10	10q24.1-q24.3
203419_at	MLL4	myeloid/lymphoid or mixed-lineage leukemia 4	9757	19	19q13.1
211874_s_at	MYST4	MYST histone acetyltransferase (monocytic leukemia) 4	23522	10	10q22.2
40569_at	MZF1	myeloid zinc finger 1	7593	19	19q13.4
203621_at	NDUFB5	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 5, 16kDa	4711	3	3q26.33
202886_s_at	PPP2R1B	protein phosphatase 2 (formerly 2A), regulatory subunit A, beta isoform	5519	11	11q23.2
201834_at	PRKAB1	protein kinase, AMP-activated, beta 1 non-catalytic subunit	5564	12	12q24.1
212743_at	RCHY1	ring finger and CHY zinc finger domain containing 1	25898	4	4q21.1
219869_s_at	SLC39A8	solute carrier family 39 (zinc transporter), member 8	64116	4	4q22-q24
210692_s_at	SLC43A3	solute carrier family 43, member 3	29015	11	11q11
213103_at	STARD13	StAR-related lipid transfer (START) domain containing 13	90627	13	13q12-q13
202342_s_at	TRIM2	tripartite motif-containing 2	23321	4	4q31.3
212534_at	ZNF24	zinc finger protein 24	7572	18	18q12
219635_at	ZNF606	zinc finger protein 606	80095	19	19q13.4
214202_at	---	---	---	5	5q22.3

FIGURE A3 (continued). Excellent pathologic response predictor genes for Estrogen Receptor positive tumors in ACES algorithm

Probe Set	Symbol	Description	GeneID	Chromosome	Cytoband
204332_s_at	AGA	aspartylglucosaminidase	175	4	4q32-q33
36865_at	ANGEL1	angel homolog 1 (Drosophila)	23357	14	14q24.3
219437_s_at	ANKRD11	ankyrin repeat domain 11	29123	16	16q24.3
205865_at	ARID3A	AT rich interactive domain 3A (BRIGHT-like)	1820	19	19p13.3
215407_s_at	ASTN2	astrotactin 2	23245	9	9q33.1
204493_at	BID	BH3 interacting domain death agonist	637	22	22q11.1
205557_at	BPI	bactericidal/permeability-increasing protein	671	20	20q11.23-q12
42361_g_at	CCHCR1	coiled-coil alpha-helical rod protein 1	54535	6	6p21.3
205937_at	CGREF1	cell growth regulator with EF-hand domain 1	10669	2	2p23.3
208817_at	COMT	catechol-O-methyltransferase	1312	22	22q11.21
202250_s_at	DCAF8	DDB1 and CUL4 associated factor 8	50717	1	1q22-q23
202570_s_at	DLGAP4	discs, large (Drosophila) homolog-associated protein 4	22839	20	20q11.23
218103_at	FTSJ3	FtsJ homolog 3 (E. coli)	117246	17	17q23.3
216651_s_at	GAD2	glutamate decarboxylase 2 (pancreatic islets and brain, 65kDa)	2572	10	10p11.23
205505_at	GCNT1	glucosaminyl (N-acetyl) transferase 1, core 2 (beta-1,6-N-acetylglucosaminyltransferase)	2650	9	9q13
213020_at	GOSR1	golgi SNAP receptor complex member 1	9527	17	17q11
212597_s_at	HMGXB4	HMG box domain containing 4	10042	22	22q13.1
212898_at	KIAA0406	KIAA0406	9675	20	20q11.23
220652_at	KIF24	kinesin family member 24	347240	9	9p13.3
218486_at	KLF11	Kruppel-like factor 11	8462	2	2p25
202057_at	KPNA1	karyopherin alpha 1 (importin alpha 5)	3836	3	3q21
209204_at	LMO4	LIM domain only 4	8543	1	1p22.3
201818_at	LPCAT1	lysophosphatidylcholine acyltransferase 1	79888	5	5p15.33
208328_s_at	MEF2A	myocyte enhancer factor 2A	4205	15	15q26
215491_at	MYCL1	v-myc myelocytomatosis viral oncogene homolog 1, lung carcinoma derived (avian)	4610	1	1p34.2
202944_at	NAGA	N-acetylgalactosaminidase, alpha-	4668	22	22q11
218886_at	PAK1IP1	PAK1 interacting protein 1	55003	6	6p24.2
207081_s_at	PI4KA	phosphatidylinositol 4-kinase, catalytic, alpha	5297	22	22q11.21
210771_at	PPARA	peroxisome proliferator-activated receptor alpha	5465	22	22q12-q13.1
203096_s_at	RAPGEF2	Rap guanine nucleotide exchange factor (GEF) 2	9693	4	4q32.1
218593_at	RBM28	RNA binding motif protein 28	55131	7	7q32.1
211678_s_at	RNF114	ring finger protein 114	55905	20	20q13.13
202762_at	ROCK2	Rho-associated, coiled-coil containing protein kinase 2	9475	2	2p24
206239_s_at	SPINK1	serine peptidase inhibitor, Kazal type 1	6690	5	5q32
221276_s_at	SYNC	syncoilin, intermediate filament protein	81493	1	1p34.3-p33
213155_at	WSCD1	WSC domain containing 1	23302	17	17p13.2
37117_at	PRR5	proline rich 5 (renal)	55615	22	22q13
220855_at	AC091271.1	no-protein transcript	---	17	17q23.2
222275_at	---	---	---	5	5p12

FIGURE A3 (continued). Extensive residual disease predictor genes for Estrogen Receptor positive tumors in ACES algorithm

Probe Set	Symbol	Description	GeneID	Chromosome	Cytoband
200045_at	ABCF1	ATP-binding cassette, sub-family F (GCN20), member 1	23	6	6p21.33
218868_at	ACTR3B	ARP3 actin-related protein 3 homolog B (yeast)	57180	7	7q36.1
213532_at	ADAM17	ADAM metallopeptidase domain 17	6868	2	2p25
217090_at	ADAM3A	ADAM metallopeptidase domain 3A (cyritestin 1)	1587	8	8p11.23
205013_s_at	ADORA2A	adenosine A2a receptor	135	22	22q11.23
208042_at	AGGF1	angiogenic factor with G patch and FHA domains 1	55109	5	5q13.3
215789_s_at	AJAP1	adherens junctions associated protein 1	55966	1	1p36.32
221825_at	ANGEL2	angel homolog 2 (Drosophila)	90806	1	1q32.3
202631_s_at	APPBP2	amyloid beta precursor protein (cytoplasmic tail) binding protein 2	10513	17	17q21-q23
200011_s_at	ARF3	ADP-ribosylation factor 3	377	12	12q13
202492_at	ATG9A	ATG9 autophagy related 9 homolog A (S. cerevisiae)	79065	2	2q35
212930_at	ATP2B1	ATPase, Ca++ transporting, plasma membrane 1	490	12	12q21.3
218789_s_at	C11orf71	chromosome 11 open reading frame 71	54494	11	11q14.2-q14.3
219022_at	C12orf43	chromosome 12 open reading frame 43	64897	12	12q
214322_at	CAMK2G	calcium/calmodulin-dependent protein kinase II gamma	818	10	10q22
218384_at	CARHSP1	calcium regulated heat stable protein 1, 24kDa	23589	16	16p13.2
212586_at	CAST	calpastatin	831	5	5q15
218592_s_at	CECR5	cat eye syndrome chromosome region, candidate 5	27440	22	
218439_s_at	COMMD10	COMM domain containing 10	51397	5	5q23.1
211808_s_at	CREBBP	CREB binding protein	1387	16	16p13.3
209164_s_at	CYB561	cytochrome b-561	1534	17	17q11-qter
203979_at	CYP27A1	cytochrome P450, family 27, subfamily A, polypeptide 1	1593	2	2q33-qter
216874_at	DKFZp686O1327	hypothetical gene supported by BC043549; BX648102	401014	2	2q22.3
204797_s_at	EML1	echinoderm microtubule associated protein like 1	2009	14	14q32
218692_at	GOLSYN	Golgi-localized protein	55638	8	8q23.2
202453_s_at	GTF2H1	general transcription factor IIH, polypeptide 1, 62kDa	2965	11	11p15.1-p14
221046_s_at	GTPBP8	GTP-binding protein 8 (putative)	29083	3	3q13.2
208886_at	H1F0	H1 histone family, member 0	3005	22	22q13.1

FIGURE A3 (continued). Extensive residual disease predictor genes for Estrogen Receptor positive tumors in ACES algorithm

205426_s_at	HIP1	huntingtin interacting protein 1	3092	7	7q11.23
202983_at	HLTF	helicase-like transcription factor	6596	3	3q25.1-q26.1
217145_at	IGKC	immunoglobulin kappa constant	3514	2	2p12
204863_s_at	IL6ST	interleukin 6 signal transducer (gp130, oncostatin M receptor)	3572	5	5q11
211817_s_at	KCNJ5	potassium inwardly-rectifying channel, subfamily J, member 5	3762	11	11q24
201776_s_at	KIAA0494	KIAA0494	9813	1	1pter-p22.1
209212_s_at	KLF5	Kruppel-like factor 5 (intestinal)	688	13	13q22.1
212271_at	MAPK1	mitogen-activated protein kinase 1	5594	22	22q11.2
206904_at	MATN1	matrilin 1, cartilage matrix protein	4146	1	1p35
206961_s_at	MED20	mediator complex subunit 20	9477	6	6p21.1
213403_at	MFSD9	major facilitator superfamily domain containing 9	84804	2	2q12.1
209733_at	MID2	midline 2	11043	X	Xq22.3
218205_s_at	MKNK2	MAP kinase interacting serine/threonine kinase 2	2872	19	19p13.3
209973_at	NFKBIL1	nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor-like 1	4795	6	6p21.3
217963_s_at	NGFRAP1	nerve growth factor receptor (TNFRSF16) associated protein 1	27018	X	Xq22.2
207400_at	NPY5R	neuropeptide Y receptor Y5	4889	4	4q31-q32
202097_at	NUP153	nucleoporin 153kDa	9972	6	6p22.3
220631_at	OSGEPL1	O-sialoglycoprotein endopeptidase-like 1	64172	2	2q32.2
205077_s_at	PIGF	phosphatidylinositol glycan anchor biosynthesis, class F	5281	2	2p21-p16
220811_at	PRG3	proteoglycan 3	10394	11	11q12
208733_at	RAB2A	RAB2A, member RAS oncogene family	5862	8	8q12.1
206066_s_at	RAD51C	RAD51 homolog C (S. cerevisiae)	5889	17	17q22-q23
206290_s_at	RGS7	regulator of G-protein signaling 7	6000	1	1q23.1
214519_s_at	RLN2	relaxin 2	6019	9	9p24.1
206805_at	SEMA3A	sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3A	10371	7	7p12.1
208941_s_at	SEPHS1	selenophosphate synthetase 1	22929	10	10p14
213755_s_at	SKI	v-ski sarcoma viral oncogene homolog (avian)	6497	1	1q22-q24
202667_s_at	SLC39A7	solute carrier family 39 (zinc transporter), member 7	7922	6	6p21.3

FIGURE A3 (continued). Extensive residual disease predictor genes for Estrogen Receptor positive tumors, in ACES algorithm

216611_s_at	SLC6A2	solute carrier family 6 (neurotransmitter transporter, noradrenalin), member 2	6530	16	16q12.2
211805_s_at	SLC8A1	solute carrier family 8 (sodium/calcium exchanger), member 1	6546	2	2p23-p22
205596_s_at	SMURF2	SMAD specific E3 ubiquitin protein ligase 2	64750	17	17q22-q23
203054_s_at	TCTA	T-cell leukemia translocation altered gene	6988	3	3p21
218099_at	TEX2	testis expressed 2	55852	17	17q23.3
217121_at	TNKS	tankyrase, TRF1-interacting ankyrin-related ADP-ribose polymerase	8658	8	8p23.1
220415_at	TNNI3K	TNNI3 interacting kinase	51086	1	1p31.1
209593_s_at	TOR1B	torsin family 1, member B (torsin B)	27348	9	9q34
215796_at	TRD@	T cell receptor delta locus	6964	14	14q11.2
210541_s_at	TRIM27	tripartite motif-containing 27	5987	6	6p22
213563_s_at	TUBGCP2	tubulin, gamma complex associated protein 2	10844	10	10q26.3
221839_s_at	UBAP2	ubiquitin associated protein 2	55833	9	9p13.3
213822_s_at	UBE3B	ubiquitin protein ligase E3B	89910	12	12q24.11
221746_at	UBL4A	ubiquitin-like 4A	8266	X	Xq28
219740_at	VASH2	vasohibin 2	79805	1	1q32.3
205877_s_at	ZC3H7B	zinc finger CCCH-type containing 7B	23264	22	22q13.2
218413_s_at	ZNF639	zinc finger protein 639	51193	3	3q26.33

Figures A2 from Hatzis et al, 2011