HHS Public Access

Author manuscript

Breast Cancer Res Treat. Author manuscript; available in PMC 2016 July 01.

Published in final edited form as:

Breast Cancer Res Treat. 2015 July ; 152(2): 437–448. doi:10.1007/s10549-015-3474-4.

RACE-ASSOCIATED BIOLOGICAL DIFFERENCES AMONG LUMINAL A BREAST TUMORS

Monica D'Arcy[1,2], **Jodie Fleming**[3], **Whitney R. Robinson**[1,2], **Erin L. Kirk**[1], **Charles M. Perou**[2,4,5], and **Melissa A. Troester**[1,2,5]

[1]Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599

[2]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599

[3]Department of Biology, North Carolina Central University, Durham, NC

[4]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599

[5]Department of Pathology and Laboratory Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599

Abstract

Purpose—African American (AA) women have higher breast-cancer specific mortality rates. A higher prevalence of the worse outcome Basal-like breast cancer subtype contributes to this, but AA women also have higher mortality even within the more favorable outcomes Luminal A breast cancers. These differences may reflect treatment or health care access issues, inherent biological differences, or both.

Methods—To identify potential biological differences by race among Luminal A breast cancers, gene expression data from 108 CAU and 57 AA breast tumors were analyzed. Race-associated genes were next evaluated for associations with survival. Finally, expression of race- and survival-associated genes was evaluated in normal tissue of AA and CAU women.

Results—Six genes (ACOX2, MUC1, CRYBB2, PSPH, SQLE, TYMS) were differentially expressed by race among Luminal A breast cancers and were associated with survival (HR < 0.8, HR > 1.25). For all six genes, tumors in AA had higher expression of poor prognosis genes (CRYBB2, PSPH, SQLE, TYMS) and lower expression of good prognosis genes (ACOX2, MUC1). A score based on all six genes predicted survival in a large independent dataset (HR = 1.9 top vs. bottom quartile, 95% CI: 1.4 – 2.5). For four genes, normal tissue of AA and CAU women showed similar expression (ACOX2, MUC1, SQLE, TYMS), however, the poor outcome

Corresponding author: Dr. Melissa A. Troester, Department of Epidemiology, CB 7435, University of North Carolina at Chapel Hill, 135 Dauer Drive, Chapel Hill, NC 27599, USA. Phone: +1 919-966-7408; Fax: +1 919-966-2089; troester@unc.edu.

Ethical approval: "All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards."

Conflict of Interest: C.M.P is an equity stock holder, and Board of Director Member, of BioClassifier LLC and GeneCentric Diagnostics. C.M.P is also listed an inventor on a patent application for the PAM50 assay.

associated genes CRYBB2 and PSPH were more highly expressed in AA vs. CAU women's normal tissue.

Conclusions—This analysis identified gene expression differences that may contribute to mortality disparities and suggests that among Luminal A breast tumors there are biological differences between AA and CAU patients. Some of these differences (CRYBB2 and PSPH) may exist from the earliest stages of tumor development, or even precede malignancy.

INTRODUCTION

Compared to Caucasian (CAU) women, African American (AA) women have lower incidence, but higher breast cancer-specific mortality rates [1]. Higher prevalence of aggressive basal-like breast cancers in AA women [2] may explain some disparities, but even when AA women are diagnosed with less-aggressive Luminal A breast cancers, they fare worse than CAU women with the same subtype [3]. There are likely multiple factors contributing to the differences, including differential access to care [4] and lifestyle factors. There is some evidence that there may be biological differences in the tumors of AA versus CAU women, even within subtype. For instance, even after controlling for some socioeconomic status variables (SES) in a study where all women received the same treatment based on tumor characteristics, the Southwest Oncology Group [5] reported survival differences between CAU and AA women. Specifically, AA had a survival disadvantage compared with CAU women for ER+ premenopausal breast tumors [HR = 1.74, 95% CI = (1.11, 2.71)] and ER+ postmenopausal breast cancer [HR = 1.61, 95% CI = (1.35, 1.93)]. While many social variables are difficult to study and the role of SES cannot be fully ruled out in such studies, it is clear that both social and biological factors should be considered.

Only a few studies [6–9] have characterized molecular differences in breast tumors by race. Martin et al [8] hypothesized that the tumor microenvironment differed between AA and CAU. They reported that independent of ER status, 19 and eight genes were differentially expressed in the breast tumor stroma and epithelium, respectively, of 18 AA and 17 CAU women. Grunda et al [7] evaluated expression of 84 genes associated with breast cancer aggressiveness, prognosis and response to therapy, and found that 20 of these genes were differentially expressed in 12 AA and 12 CAU age- and stage-matched breast tumors. Field et al [6] identified genes that were differentially expressed in 26 AA and 26 CAU age, grade and ER-matched breast tumors. They found that a few genes, including CRYBB2, PSPHL and SOS1, were differentially expressed in both normal and tumor tissue. Most recently, Stewart et al [9] analyzed age- and stage-matched breast tumors from the Tumor Cancer Genome Atlas (TCGA) project and reported 674 unique genes or transcripts that were differentially expressed by race. Despite matching on clinical features in the TCGA analysis, AA had a significantly higher risk of mortality compared with CAU women (18.87% vs 10.28% - time period not given), and these investigators found gene expression differences among luminal A (46 genes), basal-like (15 genes) and HER2 (25 genes) among stage 1–3 tumors and increasing numbers of differentially expressed genes with increasing stage (from 26 in stage 1 to 223 in stage 3). The TCGA gene signatures were not evaluated for associations with survival nor tested in independent data.

Each of these previous studies evaluated molecular features that may contribute to mortality disparities between AA and CAU breast cancer cases, however, we propose that a disparity-associated gene should meet the following criteria: (1) the gene should be differentially expressed by race in the tumor, and this association should be not be driven solely by clinical features such as intrinsic subtype, ER status or patient age, (2) the differential expression of a candidate gene should be associated with a difference in breast cancer survival. If the gene is associated with race without consequences for survivorship, its utility in explaining mortality disparities is limited. We were also interested to know whether the gene was differentially expressed in normal because these gene expression differences are more likely to predate disease progression. Our goals were to extend previous studies by studying both expression and survival, to evaluate the joint effects of multiple disparity-associated genes on survival, and to evaluate how the disparity–associated genes are expressed in normal tissue.

METHODS

Datasets and data preprocessing

We used several datasets that included tumor, tumor-adjacent normal and reduction mammoplasty gene expression data. Most of these data sources are publically available. Data characteristics, including Gene Expression Omnibus (GEO) accession numbers, are listed in Table 1. We used the *UNC337* tumor gene expression dataset to evaluate race-associated tumor gene expression. *UNC337* is a racially diverse population (race information listed in Supplemental Table 3), while the *NKI295* public dataset is racially homogenous (predominantly Caucasian European). Evaluating survival in an independent and racially homogenous population allows us to make broader inferences about the importance of relative gene expression on survival. We compared gene expression in normal (*RM)*, cancer-adjacent normal and tumor (*UNC337* + *NKI295*) datasets. Reduction mammoplasty samples were from previous reports [10, 11] and from the Normal Breast Study, a study of patients undergoing surgery at UNC Hospitals [12]. All patients provided informed consent via a protocol approved by the Institutional Review Board of the University of North Carolina at Chapel Hill. All four of these datasets were on the same expression platform (Agilent), allowing us to compare gene expression across tissue type. We used the *METABRIC* [13] dataset as an independent test dataset to evaluate the tumor-based survival associations.

The isolation of RNA and methods of basic microarray processing are described in detail by Sun et al [10], Prat et al [14] and van de Vijver et al [15] for *RM, UNC337* and *NKI295* respectively. Array filtering and cleaning of the *RM* and *NKI295* datasets are described in Pirone et al [11]. From 149 microarrays, there were 130 unique *RM* samples, of which 100 self-reported AA or CAU race. There were 92 cancer-adjacent, histologically normal samples of self-described CAU or AA women (NBS). Genes that were present (above detection limit) in fewer than 15% of samples were excluded. From the *UNC337* data, we removed autopsies, samples without corresponding demographic and race information, and averaged all replicates by probe. Genes with more than 30% missing across all samples were excluded. We used k-nearest neighbors (KNN) (k=10) to impute missing data. Missing data

were imputed using k nearest neighbors (KNN) (k=10). Data for all the normal breast tissues (*RM_NBS*) were combined using distance-weighted discrimination (DWD) [16]. Data processing and analyses were completed using BioConductor and R Version 2.14.

Race-associated gene expression in tumor and normal tissue

Race-associated genes were identified in tumors overall and stratified by subtype. Tumors were classified into intrinsic subtype using the PAM50 [17]. Supervised analysis was performed on all subtypes in *UNC337* after selection criteria, as described above, using Linear models for Microarray Data (LIMMA) [18] and a False Discovery Rate (FDR) of 5%. Subtype-stratified supervised analyses were performed on N=68 Luminal A tumors and N=39 Basal-like tumors. Using Cluster 3.0 [19], we clustered the expression data by both gene and sample, and visualized the resulting cluster dendogram in Java Treeview [20]. For genes with multiple probes, we selected the probe with the highest standard deviation to display in the two dimensional cluster. We evaluated the statistical association between cluster and race using a Chi-square test. We performed sensitivity analyses wherein models were adjusted for tumor characteristics (grade, stage, node, age) to evaluate whether tumor characteristics confounded the association between race and gene expression.

LIMMA analyses were also used to identify genes associated with race in non-tumor tissue at an FDR = 5%. Since these samples included both cancer-adjacent normal and normal tissue (N=192) from two different populations, we statistically adjusted for data source in addition to performing DWD correction as described above.

NKI295 survival

We defined disparity-relevant gene expression as gene expression associated with race in the tumor and with survival in test data. The first criteria was met by identifying race-associated gene expression at an FDR = 5% in the *UNC337* (Luminal A or Basal-like tumors) data. These genes were mapped to the NKI dataset, and were then extracted and median centered. We then performed a survival analysis that compared individuals with above-median expression to those with below-median (referent) expression for each of the race-associated genes. Two race-associated genes (FAM177A1, GSTT2) were not available in the NKI dataset. For each gene, we plotted Kaplan-Meier curves and estimated hazard ratios (HR) using Cox Proportional hazards models. A HR < 1 signifies that higher expression confers a survival advantage, whereas a HR >1 suggests that higher expression confers a survival disadvantage.

There were six race-associated genes that showed a high magnitude association (HR <0.8 or HR >1.25) with survival in the NKI295 dataset. These genes were used to create a Multi-gene Race-associated Expression (MRE) score that varied between −6 and 6. A score of −6 should predict the best survival and that of +6, the worst survival. We generated this score for each patient by summing up the "deleterious" effects of each race- and survival-associated gene. The deleterious effect was −1 when the patient level expression was below the median expression for genes with a HR > 1 or above the median expression for genes with a HR < 1. Similarly, the deleterious effect was +1 when the patient level gene

expression was above the median expression for genes with a HR > 1 or above the median expression for genes with HR < 1.

The association between the MRE score and survival was evaluated using Cox Proportional Hazards in both the training (NKI295 + UNC337, N=450) and independent (METABRIC, N=1584) dataset. We tested the statistical association between mean MRE score and tumor subtype in both datasets, and between CAU and AA tumors overall and in Luminal A tumors using either ANOVA (N > 2 groups) or Student T's test (2 groups). We also assessed association between MRE score and survival among luminal tumors. Finally, we calculated tumor proliferation scores—a marker of tumor proliferation capability and defined as the sum of expression of the following genes: {CCNB1, UBE2C, BIRC5, KNTC2, CDC20, PTTG1, RRM2, MKI67, TYMS, CEP55, CDCA1}—for all tumors, and evaluated the association between this marker and the MRE score.

Evaluating gene expression changes by race and tissue type

Patterns of gene expression in normal and tumor tissue, stratified by race are informative for whether the differential expression is a disease feature or exists prior to carcinogenesis. Thus, we tested each of the tumor race- and survival-associated genes for their expression in reduction mammoplasty (N=100), cancer-adjacent normal (N=92) and tumor tissue (N=460). In datasets where there were multiple probes for a particular gene, we chose the probe that was differentially expressed in the *UNC337* dataset if available. Alternatively, any probe mapping to the same gene was used if the specific probe was unavailable. We median centered the dataset of all three, tissue types and then plotted average relative expression in boxplots, stratified by race and tissue type.

RESULTS

Training data identification of race- and survival associated genes

Compared with CAU tumors (Table 2), AA tumors were more likely to be node positive (60% vs 42%, p = 0.03), ER negative (53% versus 31%, p = 0.03) and less differentiated (61% versus 46%, p = 0.10). At an FDR = 5%, there were 40 probes, representing 38 distinct genes, that were differentially expressed by race across all tumors. A cluster of these 38 genes (rows) and race (columns) is shown in Fig. 1. The left cluster (Cluster 1, N=63) was primarily composed of AA samples (N = 39, 62%), and the right cluster (Cluster 2, N=102) was predominately CAU samples (N = 84, 82%), (chi-square = 31.8, df = 1, p < 1.7e-8).

To identify genes that were differentially expressed by race among less aggressive tumors (Luminal A), or more aggressive tumors (Basal-like), we performed two supervised analyses at an FDR = 10%, restricted to Luminal A or basal-like tumors. There were 23 genes differentially expressed by race at a 10% FDR among Luminal A tumors, of which 10 genes (Table 3a) were significant given 5% FDR. There were only two genes differentially expressed given 5% or 10% FDR in Basal-like tumors (Table 3b). We also adjusted for age, grade, size and node status in multivariable analyses (Supplemental Tables 1a–1b); among Luminal A breast cancers, most genes were still differentially expressed by race after

statistical adjustment including: CRYBB2, PSPH, MUC1, HSDL1, GSTT2, CLEC2D, FAM177A1. AMFR and PSPH remained differentially expressed by race among Basal-like tumors in multivariable model.

Among these race-associated genes, six were also associated with survival in the NKI295 dataset. High expression of CRYBB2, PSPH, TYMS, and SQLE was associated with higher mortality, while low expression of MUC1 and ACOX2 predicted worse survival (Table 4). The CRYBB2 survival curves violated the proportional hazards assumption, with the crossover of the two curves occurring at ~8–10 years. This pattern of crossing hazards at ~8–10 years has been previously documented among ER positive breast cancers [11]. Four other race-associated genes (AMFR, CLEC2D, HSDL1, SLC9A3R2) were not associated with survival in the NKI dataset (Supplemental Fig. 1).

Gene expression in normal versus tumor

To elucidate patterns of expression for race- and survival-associated genes from normal to tumor tissue, we evaluated expression of these genes in reduction mammoplasty, cancer-adjacent normal and tumor. On average, expression was higher in tumors of AA compared to CAU for SQLE and TYMS, and lower in ACOX2 and MUC1 tumors comparing AA to CAU. However, the pattern of expression in the normal to tumor expression continuum was similar between CAU and AA (Figs. 2a–d). In contrast, both CRYBB and PSPH were differentially expressed by race in both normal and tumor tissue; higher gene expression among AA compared with CAU women (Figs. 2e–f). Higher expression of these genes by race in benign tissue was not substantially attenuated even after statistical adjustment for normal tissue type (normal versus adjacent-normal).

For all six MRE-associated genes, AA had higher expression of poor prognosis genes and lower expression of good prognosis genes (Table 4) compared with CAU. Higher relative expression of ACOX2 and MUC1 in tumors were each associated with a ~35% reduction in mortality (Table 4), and AA tended to have lower expression of these genes compared with CAU (Fig. 2a–b, Table 4). For CRYBB2, PSPH, TYMS and SQLE, higher relative gene expression was associated with increased mortality in the NKI295 dataset (Table 4, Figs. 2a–b, 2e–f). In both CAU and AA women, gene expression of TYMS and SQLE increases dramatically from normal and adjacent-normal tissue to tumor tissue; however, the tumor expression was highest among AA women (Table 4, Fig. 2c–d). For PSPH and CRYBB2, gene expression increased from normal to tumor tissue among CAU women, whereas relative expression was higher in both tumor and normal of AA women.

Multi-gene Race-associated Expression (MRE) score

Applying the information learned from training on NKI data to an independent dataset, we computed MRE scores for all individuals in our test data *(METABRIC*, N=1584) and training data (*NKI295+UNC337*, N=450), where −6 = best and +6 = worst risk. Each increasing MRE point was associated with a 6% increase in hazard, HR = 1.06, 95% CI = (1.04, 1.09), such that the HR comparing a 6 point individual to a −6 point individual was 2.03, 95% CI = (1.98, 2.08) in METABRIC. This result was attenuated when we adjusted for size, grade and node status (HR = 1.03, 95% CI = (1.00,1.06) for each increase in MRE

points), but remained statistically significant. We also observed strong associations for MRE score and survival when we restricted to N=401 Luminal tumors, with a HR = 1.76, 95% CI = (1.64, 1.89) for the comparison of an individual with a score of +6 to an individual with a score of −6. After adjustment for size, grade and node status, the association was: HR = 1.36, 95% CI = (1.26, 1.46). Interestingly, high "MRE" scores were associated with Basal-like tumors in both the training and test datasets (Fig. 3a–b). AA patients had a significantly higher ($p < 0.001$) MRE score (2.42) than Caucasians (−0.32) in our test dataset (NKI295 + UNC337) over all tumors (Fig. 3c), and specifically in Luminal A tumors (1.67 vs −2.43, $p < 0.001$) (Fig. 3d). Associations between MRE score and race could not be assessed in METABRIC due to the predominance of CAU patients in that dataset.

Although there was a monotonic increase in the hazard ratio with each increase in the MRE score, the largest increase in risk occurs with just a few gene expression changes; in our test dataset, patients with the middle 50% MRE scores (compared to those in bottom quartile) had a strong elevation in risk (HR = 1.7, 95% CI = (1.3, 2.1)), and those with the top quartile MRE scores had a HR = 1.9, 95% CI = (1.4, 2.5) compared to the bottom quartile (Fig. 4b). There associations were slightly attenuated results when we restricted to Luminal A/B tumors. Comparing those with the highest quartile MRE score and those with the middle 50% of MRE score to the referent, lowest quartile group, there was a 50% [95% CI = (1.0, 2.3)] and 60% [95% CI = (1.2, 2.1)] increase in hazard respectively (Fig. 4c). We were limited by a relatively small number of luminal A tumors with a high MRE score (Fig. 4d) to sufficiently examine the association between MRE score and survival among luminal A tumors. In sensitivity analyses (data not shown) we found that relative measures of association were much stronger in the METABRIC dataset when we restricted to women <= 60 years of age. The UNC337 and NKI295 datasets are comprised predominantly of younger women so the attenuation of effect in test data may be partially attributed to population differences between the METABRIC and UNC337+NKI297 datasets.

Tumor proliferation scores were correlated with MRE points in both the training ($\rho = .55$, $p < 0.001$) and test ($\rho = 0.59$, $p < 0.001$) datasets. Proliferation scores were significantly ($p < 0.001$) higher in AA women (1.63) than in CAU women (−0.73) over all tumors in the UNC337 dataset, and in luminal A tumors (−0.99 AA versus −3.90 CAU, $p < 0.001$). It is important to note that the disparities score was attenuated but remained significant after adjusting for standard clinical variables, but was not significant after adjusting for breast cancer subtype (Table 5).

DISCUSSION

It has been established that AA women suffer from worse breast cancer outcomes compared to CAU women. While aggressive forms of BC disproportionately affect AA women, this does not fully explain the disparities; even within subtype there are differences in survival by race. From previous studies we have learned that: 1) survival differences exist between AA and CAU despite equal treatment [5] or tumor subtype [3] and 2) genes are differentially expressed between AA and CAU tumors even when matching on clinical features [6, 7, 9]. Genes that are both differentially expressed by race in tumors, and confer a survival disadvantage could explain a portion of the observed racial survival disparity, although to

date no study had evaluated whether race-associated genes conferred a survival disadvantage. To elucidate biological factors that predispose AA women to worse mortality outcomes, the current study showed that race-associated genes affect survival across multiple datasets.

Six candidate genes (CRYBB2, PSPH, ACOX2, MUC1, SQLE, TYMS) emerged from our analysis as both race- and survival- associated. Some of these genes have known biological functions, while others do not. Although AMFR tumor expression was, and has been previously shown to be associated with race [8], its expression is not associated with a survival advantage, suggesting that its differential expression may not contribute to racial mortality disparities. In contrast, we replicated the association between CRYBB2 [6, 8, 9, 21, 22] and PSPH [21, 23] tumor expression and race, and also found that higher expression is associated with poorer survival. CRYBB2 encodes for the beta-crystallin B2 protein located at 22q11.23. Genetic variation in CRYBB2 is associated with macular degeneration [24], but the protein has no documented or hypothesized role in carcinogenesis. PSPH (Phosposerine-phosphotase, 7p11.2) is located near a region where gain of function is associated with advanced prostate tumor stage [25]. This gene has also been implicated in metabolism [26]. An emerging hallmark of cancer [27], the "Warburg effect", is the ability of cancer cells to thrive in an oxygenated environment through glycolysis. Based on candidate gene studies of genes involved in cellular metabolism, Kim et al [26] found that PSPH expression was higher in basal-like tumors than in Luminal A tumors, and that high expression was associated with poor survival, HR = 2.07, 95% CI = (1.10, 4.18), an effect size similar to what we found here.

There is limited literature on ACOX2 (acyl-CoA oxidase 2, branched chain, 3p14.3), but there may be a genetic variant that is a shared risk factor for preeclampsia and cardiovascular disease [28], and some evidence that a transcript is associated with hepatocellular carcinoma [29]. ACOX2 up-regulation in the tumor conferred the same protective effect on mortality as MUC1 (mucin 1, cell surface associated 1q21). Variants of MUC1 interact with estrogen [30], and higher expression is associated with late stage epithelial ovarian cancer [31, 32] and prostate cancer [33].

Although little has been documented about the role that SQLE (squalene epoxidase, 8q24.1) plays in breast cancer progression, one study found that SQLE mRNA expression was inversely associated with survival among ER+ stage 1 or 2 patients [34]; this parallels our results that increased expression in tumor tissue is associated with almost a 2-fold increase in mortality in the NKI295 dataset. High TYMS (thymidylate synthetase, 18p11.32) expression was the largest independent predictor of mortality in our analysis. TYMS is associated with tumor proliferation, and is one of the 50 genes whose expression is used to classify breast cancer into intrinsic subtype [17]. Genetic variants of TYMS or its expression predict sensitivity to 5-fluorouracil [35–37], are prognostic factors for lymph node infiltration in CRC patients [38], and lower expression of TYMS is a positive prognostic factor for non-small cell lung cancer [39].

Disparities in survival may result from the joint expression of multiple genes, rather than from a single gene. Our MRE score captured the cumulative effects of multiple genes and

showed that high MRE points were associated with worse outcomes in both our training and test datasets. This score was also positively correlated with tumor proliferation score, providing independent confirmation of biological relevance. However, this score should not be considered as a substitute for established prognostic markers such as intrinsic subtype. After controlling for intrinsic subtype, the MRE score was not statistically significantly associated with breast cancer survival in either the UNC/NKI dataset or the METABRIC data, and generally appears to be higher in in basal-like and luminal B tumors. A major limitation to our study is that we were not sufficiently powered to evaluate subtype specific survival advantages associated with the MRE score. For instance, there were only 7 METABRIC luminal A individuals with an MRE score higher than 3. The direction of effect is also unknown: these genes may increase probability of progressing to a more aggressive subtype, or more aggressive subtypes may have increased probability of upregulating these genes. These two possibilities cannot be evaluated in human tumor specimens that are sampled only at a single point in time. However, future research identifying the mechanism of action of the genes in the MRE score could help establish their biological relevance.

Although some disparities in luminal A breast cancer mortality may be attributable to treatment or access to care, the patterns of expression of these genes in the continuum of normal to malignancy suggest that intrinsic biological differences between at-risk AA and CAU women may also be operating, and these patterns guide our interpretations of the data. Furthermore, because CRYBB2 and PSPH expression were elevated in both normal and tumor tissue of AA compared with CAU, racial differences likely exist from the earliest stages of tumor development. Previous studies have suggested that PSPHL (a PSPH homolog) and CRYBB2 were differentially expressed in normal tissue of AA and CAU women who underwent reduction mammoplasty (N=6) or those without evidence of a malignancy (N=19) [6]. These investigators posited that SNP rs6700—located close to PSPHL—may explain the differences in expression, since the minor allele frequency of AA is higher compared with CAU. We note that future studies using RNAseq should evaluate the specific transcripts of PSPH and CRYBB2 with respect to race, because recent studies [40, 41] suggest that these genes have significant homology to pseudogenes that could produce signal on a microarray. Additionally, Sturtz et al [41] concluded that the PSPHL disparity signal observed in several studies may be due to population stratification.

Breast cancer mortality disparities are likely driven by a number of social and biological forces. Uncovering the factors that drive disparities is complicated and necessitates evaluating the problem from many different vantage points. Replication of findings across multiple study populations, and by investigators using different analytical and technical approaches, strengthens the evidence in support of these genes as possible targets. Continued evaluation of genes that differ by race in both tumor and normal, such as CRYBB2 and PSPH, as candidate markers of race-associated disparities should include larger population-based studies. Mechanistic studies are also needed, especially for CRYBB2, which has now been shown to associate with race and survival in multiple studies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We acknowledge Michael Grauer, MS for reading this manuscript for clarity, and Chris Fan, PhD, for obtaining and sending the METBARIC data to MD.

Funding: This study was funded by: the National Institutes of Health (NIH) Cancer Control and Education Program pre-doctoral fellowship (5 R25 CA057726-23), and by the following grants from the NIH: R21 CA175783, U54 CA156735, U01 CA 179715, U01 ES 019472, P50 CA058223, K01CA172717.

References

1. American Cancer Society. Breast Cancer Facts and Figures 2013–2014. ACS; 2013.
2. Carey LA, Perou CM, Livasy CA, et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. JAMA. 2006; 295:2492–2502.10.1001/jama.295.21.2492 [PubMed: 16757721]
3. O'Brien KM, Cole SR, Tse CK, et al. Intrinsic breast tumor subtypes, race, and long-term survival in the Carolina Breast Cancer Study. Clin Cancer Res. 2010; 16:6100–6110.10.1158/1078-0432.CCR-10-1533 [PubMed: 21169259]
4. Reeder-Hayes KE, Meyer AMB, Dusetzina S, et al. Racial disparities in initiation of adjuvant endocrine therapy of early breast cancer. Breast Cancer Res Treat. 2014; 145:743–751.10.1007/s10549-014-2957-z [PubMed: 24789443]
5. Albain KS, Unger JM, Crowley JJ, et al. Racial disparities in cancer survival among randomized clinical trials patients of the Southwest Oncology Group. J Natl Cancer Inst. 2009; 101:984–992.10.1093/jnci/djp175 [PubMed: 19584328]
6. Field LA, Love B, Deyarmin B, et al. Identification of differentially expressed genes in breast tumors from African American compared with Caucasian women. Cancer. 2012; 118:1334–1344.10.1002/cncr.26405 [PubMed: 21800289]
7. Grunda JM, Steg AD, He Q, et al. Differential expression of breast cancer-associated genes between stage- and age-matched tumor specimens from African- and Caucasian-American Women diagnosed with breast cancer. BMC Res Notes. 2012; 5:248–0500-5-248.10.1186/1756-0500-5-248 [PubMed: 22616718]
8. Martin DN, Boersma BJ, Yi M, et al. Differences in the tumor microenvironment between African-American and European-American breast cancer patients. PLoS One. 2009; 4:e4531.10.1371/journal.pone.0004531 [PubMed: 19225562]
9. Stewart PA, Luks J, Roycik MD, et al. Differentially expressed transcripts and dysregulated signaling pathways and networks in African American breast cancer. PLoS One. 2013; 8:e82460.10.1371/journal.pone.0082460 [PubMed: 24324792]
10. Sun X, Casbas-Hernandez P, Bigelow C, et al. Normal breast tissue of obese women is enriched for macrophage markers and macrophage-associated gene expression. Breast Cancer Res Treat. 2012; 131:1003–1012.10.1007/s10549-011-1789-3 [PubMed: 22002519]
11. Pirone JR, D'Arcy M, Stewart DA, et al. Age-associated gene expression in normal breast tissue mirrors qualitative age-at-incidence patterns for breast cancer. Cancer Epidemiol Biomarkers Prev. 2012; 21:1735–1744.10.1158/1055-9965.EPI-12-0451 [PubMed: 22859400]
12. Casbas-Hernandez P, Sun X, Roman-Perez E, et al. Tumor Intrinsic Subtype is Reflected in Cancer-Adjacent Tissue. Cancer Epidemiol Biomarkers Prev. 2014 cebp.0934.2014 [pii].
13. Curtis C, Shah SP, Chin SF, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature. 2012; 486:346–352.10.1038/nature10983 [PubMed: 22522925]
14. Prat A, Parker JS, Karginova O, et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Res. 2010; 12:R68.10.1186/bcr2635 [PubMed: 20813035]

15. van de Vijver MJ, He YD, van't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med. 2002; 347:1999–2009.10.1056/NEJMoa021967 [PubMed: 12490681]
16. Benito M, Parker J, Du Q, et al. Adjustment of systematic microarray data biases. Bioinformatics. 2004; 20:105–114. [PubMed: 14693816]
17. Parker JS, Mullins M, Cheang MC, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol. 2009; 27:1160–1167.10.1200/JCO.2008.18.1370 [PubMed: 19204204]
18. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol. 2004; 3:Article3.10.2202/1544-6115.1027 [PubMed: 16646809]
19. de Hoon MJ, Imoto S, Nolan J, et al. Open source clustering software. Bioinformatics. 2004; 20:1453–1454.10.1093/bioinformatics/bth078 [PubMed: 14871861]
20. Saldanha AJ. Java Treeview--extensible visualization of microarray data. Bioinformatics. 2004; 20:3246–3248.10.1093/bioinformatics/bth349 [PubMed: 15180930]
21. Jovov B, Araujo-Perez F, Sigel CS, et al. Differential gene expression between African American and European American colorectal cancer patients. PLoS One. 2012; 7:e30168.10.1371/journal.pone.0030168 [PubMed: 22276153]
22. Wallace TA, Prueitt RL, Yi M, et al. Tumor immunobiological differences in prostate cancer between African-American and European-American men. Cancer Res. 2008; 68:927–936.10.1158/0008-5472.CAN-07-2608 [PubMed: 18245496]
23. Ferguson SE, Olshen AB, Levine DA, et al. Molecular profiling of endometrial cancers from African-American and Caucasian women. Gynecol Oncol. 2006; 101:209–213.10.1016/j.ygyno.2005.11.028 [PubMed: 16380157]
24. Sturgill GM, Bala E, Yaniglos SS, et al. Mutation screen of beta-crystallin genes in 274 patients with age-related macular degeneration. Ophthalmic Genet. 2010; 31:129–134.10.3109/13816810.2010.486774 [PubMed: 20565250]
25. Alers JC, Rochat J, Krijtenburg PJ, et al. Identification of genetic markers for prostatic cancer progression. Lab Invest. 2000; 80:931–942. [PubMed: 10879743]
26. Kim SK, Jung WH, Koo JS. Differential expression of enzymes associated with serine/glycine metabolism in different breast cancer subtypes. PLoS One. 2014; 9:e101004.10.1371/journal.pone.0101004 [PubMed: 24979213]
27. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011; 144:646–674.10.1016/j.cell.2011.02.013 [PubMed: 21376230]
28. Johansson A, Curran JE, Johnson MP, et al. Identification of ACOX2 as a shared genetic risk factor for preeclampsia and cardiovascular disease. Eur J Hum Genet. 2011; 19:796–800.10.1038/ejhg.2011.19 [PubMed: 21343950]
29. Hodo Y, Hashimoto S, Honda M, et al. Comprehensive gene expression analysis of 5′-end of mRNA identified novel intronic transcripts associated with hepatocellular carcinoma. Genomics. 2010; 95:217–223.10.1016/j.ygeno.2010.01.004 [PubMed: 20096344]
30. Zaretsky JZ, Barnea I, Aylon Y, et al. MUC1 gene overexpressed in breast cancer: structure and transcriptional activity of the MUC1 promoter and role of estrogen receptor alpha (ERalpha) in regulation of the MUC1 gene expression. Mol Cancer. 2006; 5:57.10.1186/1476-4598-5-57 [PubMed: 17083744]
31. Deng J, Wang L, Chen H, et al. The role of tumour-associated MUC1 in epithelial ovarian cancer metastasis and progression. Cancer Metastasis Rev. 2013; 32:535–551.10.1007/s10555-013-9423-y [PubMed: 23609751]
32. Wang L, Ma J, Liu F, et al. Expression of MUC1 in primary and metastatic human epithelial ovarian cancer and its therapeutic significance. Gynecol Oncol. 2007; 105:695–702.10.1016/j.ygyno.2007.02.004 [PubMed: 17368732]
33. Li Y, Cozzi PJ. MUC1 is a promising therapeutic target for prostate cancer therapy. Curr Cancer Drug Targets. 2007; 7:259–271. [PubMed: 17504123]

34. Helms MW, Kemming D, Pospisil H, et al. Squalene epoxidase, located on chromosome 8q24.1, is upregulated in 8q+ breast cancer and indicates poor clinical outcome in stage I and II disease. Br J Cancer. 2008; 99:774–780.10.1038/sj.bjc.6604556 [PubMed: 18728668]

35. Martinez-Balibrea E, Abad A, Martinez-Cardus A, et al. UGT1A and TYMS genetic variants predict toxicity and response of colorectal cancer patients treated with first-line irinotecan and fluorouracil combination therapy. Br J Cancer. 2010; 103:581–589.10.1038/sj.bjc.6605776 [PubMed: 20628391]

36. Donada M, Bonin S, Barbazza R, et al. Management of stage II colon cancer - the use of molecular biomarkers for adjuvant therapy decision. BMC Gastroenterol. 2013; 13 36-230X-13-36. 10.1186/1471-230X-13-36

37. Loganayagam A, Arenas Hernandez M, Corrigan A, et al. Pharmacogenetic variants in the DPYD, TYMS, CDA and MTHFR genes are clinically significant predictors of fluoropyrimidine toxicity. Br J Cancer. 2013; 108:2505–2515.10.1038/bjc.2013.262 [PubMed: 23736036]

38. Lu Y, Zhuo C, Cui B, et al. TYMS serves as a prognostic indicator to predict the lymph node metastasis in Chinese patients with colorectal cancer. Clin Biochem. 2013; 46:1478–1483.10.1016/j.clinbiochem.2013.06.017 [PubMed: 23810585]

39. Yokomise H, Liu D, Chang S, et al. Biomarkers as prognostic factors for cN2 or 3 non-small cell lung cancer treated by induction chemoradiotherapy and surgery. Anticancer Res. 2013; 33:1107–1115. [PubMed: 23482788]

40. Rummel S, Penatzer CE, Shriver CD, et al. PSPHL and breast cancer in African American women: causative gene or population stratification? BMC Genet. 2014; 15:38-2156-15-38.10.1186/1471-2156-15-38 [PubMed: 24650299]

41. Sturtz LA, Melley J, Mamula K, et al. Outcome disparities in African American women with triple negative breast cancer: a comparison of epidemiological and molecular factors between African American and Caucasian women with triple negative breast cancer. BMC Cancer. 2014; 14:62-2407-14-62.10.1186/1471-2407-14-62 [PubMed: 24495414]

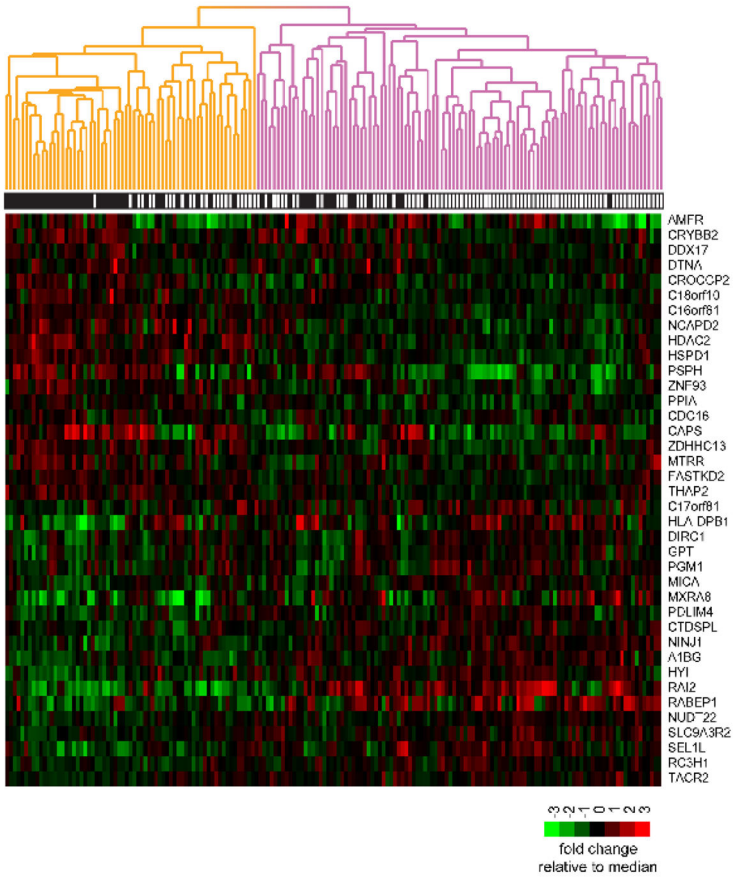

Fig. 1. Two-dimensional cluster of race-associated gene expression

All genes (rows) were median-centered across the samples (columns). AA women are represented by black boxes immediately above the heatmap and CAU women with white boxes. There are two distinct gene clusters with the orange cluster including primarily AA and the purple cluster including primarily CAU.

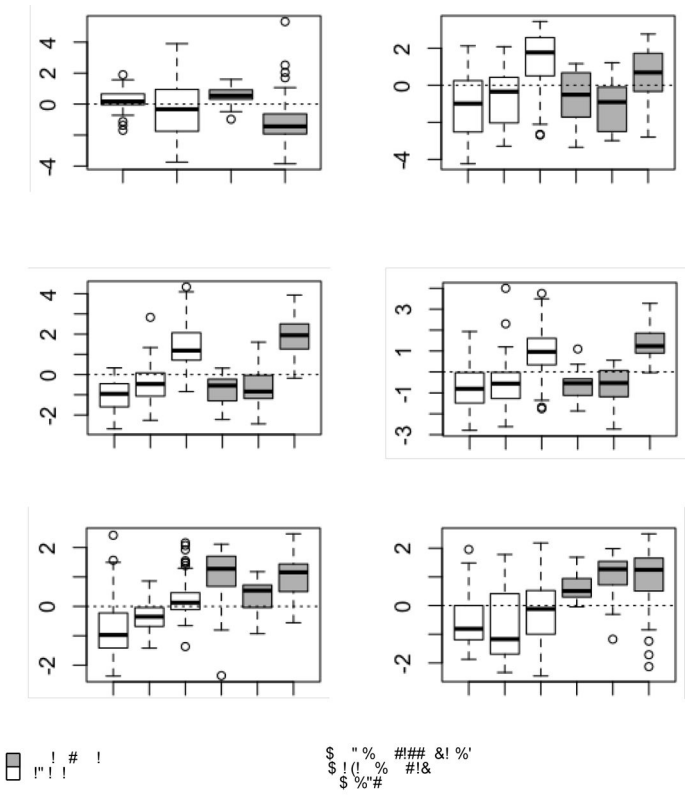

Fig. 2. Median-centered gene expression across samples, stratified by race and tissue type (normal, adjacent normal, tumor) for race- and survival-associated (disparity score) genes

There are two distinct patterns of expression. In figures a–d, expression is most distinct by race among tumors, whereas levels are similar by race in normal tissue; In figures e and f, however, CRYBB2 (e) and PSPH (f) levels are higher even in normal tissue of AA women. The increased expression persists in tumor..

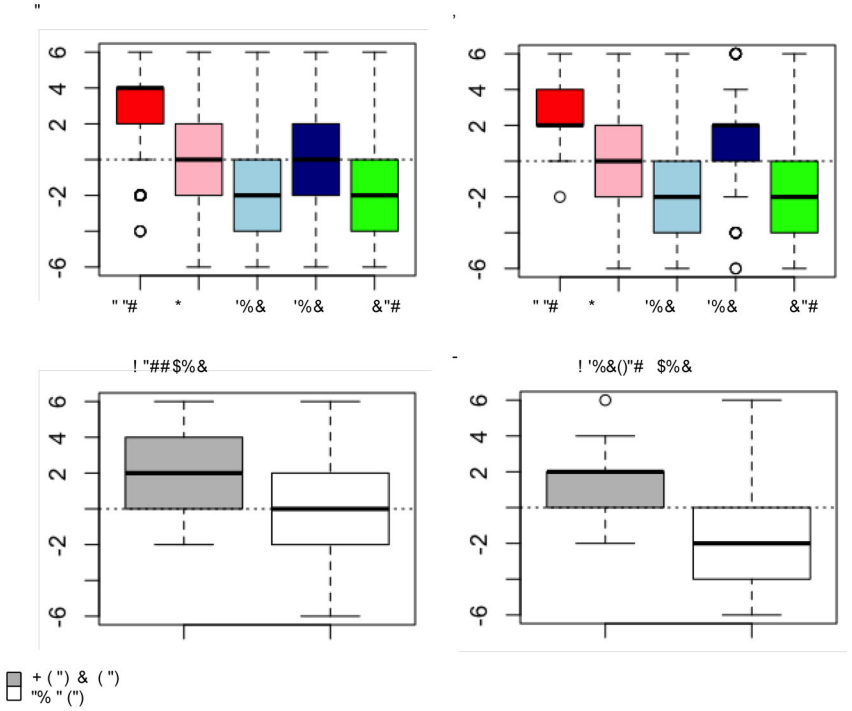

Fig. 3. Top: Boxplots showing the distribution of disparity score by subtype and by race

Disparity scores are highest for basal-like, HER2, and Luminal B tumors in both test (METABRIC, a) and training (UNC337 + NKI295, b) data. Across all tumors (c), disparity scores are higher in African Americans, but these differences are not solely driven by subtype because expression is higher even among Luminal A tumors only (d).

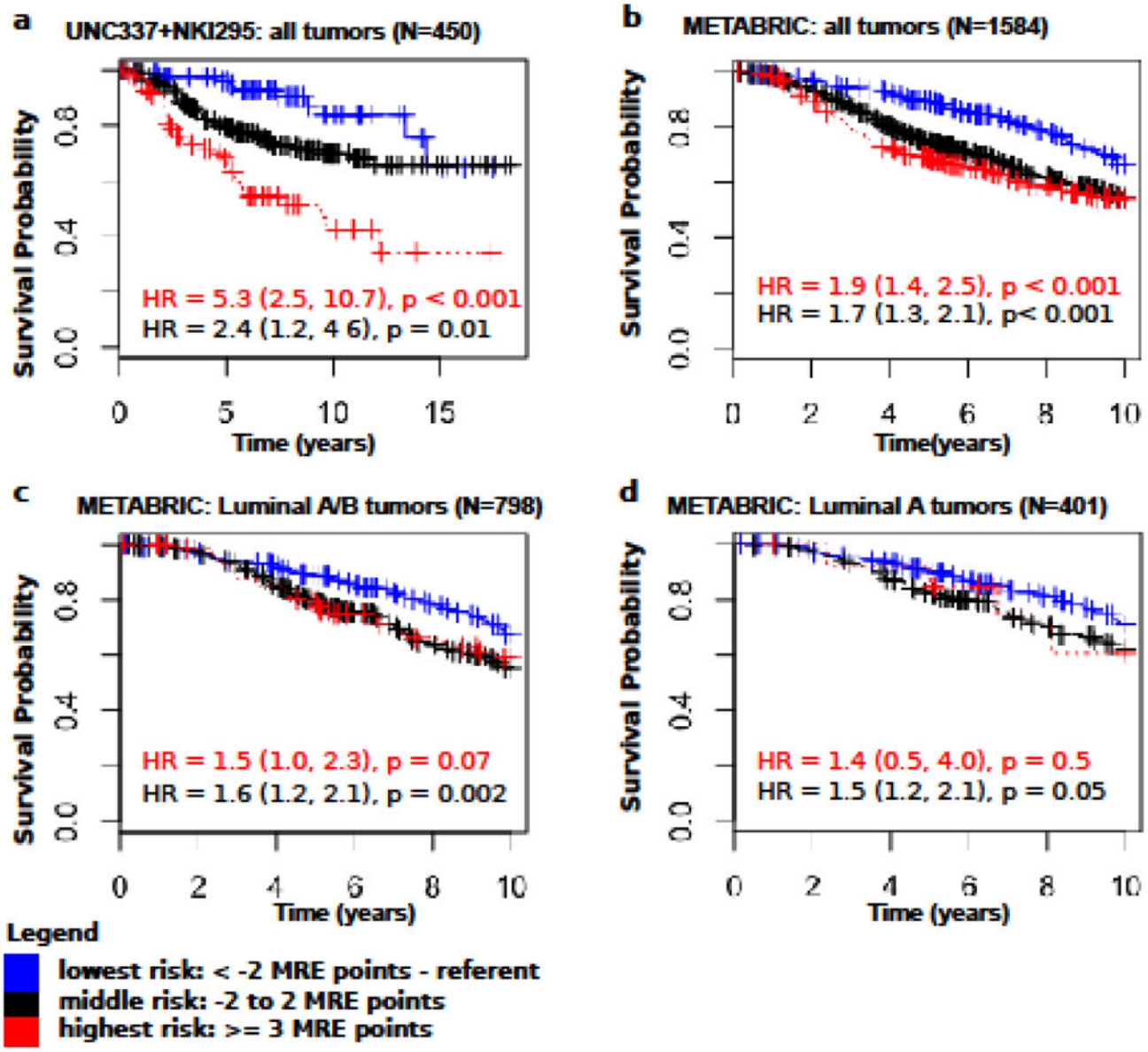

Fig. 4. Survival curves and corresponding Hazard Ratios (HR) by disparity score
(bottom 25%, middle 50%, top 25%) in the training data (a) and test data: overall (b), Luminal A/B (c) only, and Luminal A only (d).

Table 1

Data characteristics and GEO accession numbers

Data source	GEO accession	N	CAU	AA	Tissue	Ref	Purpose
UNC337	GSE18229	165 [a]	108	57	Tumor	[14]	Identify race associated genes
NKI295	NA	295	295	NA	Tumor	[15]	Evaluate survival benefit for genes identified in UNC337
NBS	GSE50939	92	65	27	Adjacent [b]	[12]	Expression comparison to tumor
RM	GSE43973	100	89	11	Normal [b]	[11]	Expression comparison to tumor
METABRIC	NA	1584	NA	NA	Tumor	[13]	Independent test set: association between disparity points and survival and tumor characteristics
RM_NBS	NA	192	165	44	Normal+Adjacent [b]	Derived, [11, 12]	Evaluate race/survival associated genes in normal tissue
UNC337+NKI295	NA	460	403	57	Tumor	[14, 15]	Expression comparison to normal/adjacent normal

[a] N=155 with survival data

[b] Normal from reduction mammoplasty, Adjacent from cancer adjacent normal

Table 2

Demographic and tumor characteristics of UNC337

	Caucasian N=108 (%)	African American N=57 (%)	P value
Age (years)			
< 40	12 (11)	8 (14)	
40–49	27 (25)	14 (25)	
>=50	69 (64)	35 (61)	
Missing	0	0	
			p = 0.86
Tumor Size			
<2 cm	33 (31)	11 (19)	
>= 2 cm	71 (66)	42 (74)	
Missing	4 (4)	4 (7)	
			p = 0.21
Tumor grade			
well (1)	12 (11)	2 (4)	
moderate (2)	39 (36)	17 (30)	
poor (3)	50 (46)	35 (61)	
Missing	7 (6)	3 (5)	
			p = 0.10
ER status			
Positive	70 (65)	26 (46)	
Negative	34 (31)	30 (53)	
Missing	4 (4)	1 (2)	
			p = 0.02
PR status			
Positive	48 (44)	20 (35)	
Negative	43 (40)	27 (47)	
Missing	17 (16)	10 (18)	
			p = 0.34
Node status			
Negative	59 (55)	21 (37)	
Positive	45 (42)	34 (60)	
Missing	4 (4)	2 (4)	
			p = 0.03
Subtype			
Basal	21 (19)	18 (32)	
Her2	13 (12)	6 (11)	
Lum A	49 (45)	19 (33)	
Lum B	17 (16)	11 (19)	
Normal	8 (7)	3 (5)	
Missing	0	0	

	Caucasian N=108 (%)	African American N=57 (%)	P value
			P = 0.38

Table 3a

Luminal A race-associated genes at FDR = 10%

Agilent Probe	Symbol	Entrez ID	MedianLog2rg_All	MedianLog2rg_AA	MedianLog2rg_C
A_23_P10182	ACOX2	8309	0.403	−0.046	1.058
A_23_P109427	GSTT2	2953	0.522	0.997	0.4765
A_23_P137856	MUC1	4582	2.306	1.437	2.5165
A_23_P146284	SQLE	6713	−1.37	−0.84	−1.596
A_23_P14986	HSD11B2	3291	0.536	0.536	0.524
A_23_P155989	CENPK	64105	−2.647	−2.521	−2.8315
A_23_P169137	NINJ1	4814	0.815	0.473	1.02
A_23_P19084	HNRNPAB	3182	−0.94	−0.77	−1.0135
A_23_P204689	CLEC2D	29121	0.721	0.912	0.618
A_23_P206324	HSDL1	83693	−0.879	−0.742	−0.9555
A_23_P208143	ZNF397	84307	0.161	0.398	0.066
A_23_P214037	NPM1	4869	−1.557	−1.358	−1.6025
A_23_P251984	PSPH	5723	−0.864	0.036	−1.3275
A_23_P26534	HCFC1R1	54985	0.371	0.201	0.5305
A_23_P2935	FAM177A1	283635	0.034	0.152	−0.0135
A_23_P40574	CRYBB2	1415	−0.011	0.794	−0.237
A_23_P50096	TYMS	7298	−2.67	−2.57	−2.845
A_23_P50108	NDC80	10403	−1.992	−1.858	−2.122
A_23_P52031	PGM1	5236	−0.558	−0.82	−0.3495
A_23_P87769	C12orf48	55010	−1.764	−1.61	−1.8355
A_23_P92441	MAD2L1	4085	−2.62	−2.34	−2.81
A_23_P96209	REEP4	80346	−0.642	−0.518	−0.76
A_24_P364838	SLC9A3R2	9351	0.557	0.269	0.729

Table 3b

Basal-like race-associated genes at FDR = 10%

Agilent Probe	Symbol	Entrez ID	MedianLog2rg_All	MedianLog2rg_AA	MedianLog2rg_C
A_23_P141005	AMFR	267	0.574	0.942	0.2355
A_23_P251984	PSPH	5723	−0.864	0.036	−1.3275

Table 4

Hazard ratios comparing in NKI295 and average gene expression in UNC337 tumors.

Gene	HR (95% CI)	Median expression overall (UNC337)	Median expression: AA	Median expression: CAU
ACOX2	0.65 (0.42, 1.02)	0.403	−0.046	1.058
CRYBB2	1.36 (0.87, 2.13)	−0.011	0.794	−0.237
MUC1	0.65 (0.42, 1.02)	2.306	1.437	2.5165
PSPH	1.65 (1.05, 2.58)	−0.864	0.036	−1.3275
SQLE	1.98 (1.25, 3.14)	−1.37	−0.84	−1.596
TYMS	2.67 (1.65, 4.32)	−2.67	−2.57	−2.845

Table 5

Multivariable Hazard Ratios for MRE Score

	NKI + UNC337		METABRIC	
Model	**HR (95% CI)**	**p-value**	**HR (95% CI)**	**p-value**
MRE Score	1.19 (1.12, 1.27)	2.4E-08	1.06 (1.04, 1.09)	3.7E-06
MRE Score + clinical	1.11 (1.03, 1.19)	0.0040	1.05 (1.02, 1.08)	0.0004
MRE Score + clinical + subtype	1.07 (0.99, 1.16)	0.08	0.99 (0.96, 1.02)	0.46