RNA-seq Reveals Tumor MHC II Prognostic Signature in Triple Negative Breast Cancer

Katherine E. Varley$^{1,4}$, Andres Forero$^{2,6,7}$, Yufeng Li$^{2,6,7}$, Dongquan Chen$^{2,6,7}$, William E. Grizzle$^{2,6,8}$, Erinn Downs-Kelly$^{1,5}$, Todd C. Burwell$^{3}$, Christos Vaklavas$^{2,6,7}$, Donald J. Buchsbaum$^{2,6,10}$, Richard M. Myers$^{3,2,6,9}$, and Albert F. LoBuglio$^{2,6}$

$^1$University of Utah Huntsman Cancer Institute, $^4$Department of Oncological Sciences, $^5$Department of Pathology, Salt Lake City, UT; $^2$University of Alabama at Birmingham, $^6$Comprehensive Cancer Center, $^7$Department of Medicine, $^8$Department of Pathology, $^9$Department of Genetics, $^{10}$Department of Radiation Oncology, Birmingham, AL; $^3$HudsonAlpha Institute for Biotechnology, Huntsville, AL

RESEARCH SUPPORT
Funding for the work has been provided by the UAB SPORE in Breast Cancer (NCI P50 CA089019), the Breast Cancer Research Foundation of Alabama, the Cancer Center Core Support Grant (P30 CA013148), and the Susan G. Komen for the Cure Foundation (KG090969).

Corresponding Author:
Albert F. LoBuglio, M.D.
Director Emeritus and Distinguished Professor
Comprehensive Cancer Center
University of Alabama at Birmingham
1824 Sixth Avenue South, WTI 260H
Birmingham, AL 35294-3300
205-934-5077 (ph); 205-975-7428 (fax)
lobuglio@uab.edu

RUNNING HEAD – Tumor MHC II Prognostic Signature for Triple Negative Breast Cancer

This study has not been presented elsewhere.

DISCLAIMERS
Following the preparation of the manuscript the HudsonAlpha Institute for Biotechnology filed a provisional patent application describing this discovery. The manuscript authors (RMM, KEV, DJB, AF, AFL) are listed as inventors on the patent.
Abstract

**Purpose:** To identify genes in the tumor tissue from patients with triple negative breast cancer (TNBC) that are differentially expressed between patients who do or do not have disease relapse.

**Patients and Methods:** Forty-seven snap frozen primary TNBC tumor specimens were analyzed using RNA-seq to identify gene expression differences between 22 patients who relapsed and 25 who did not.

**Results:** Twenty-four genes exhibited significantly higher expression in tumor tissue from patients who did not relapse. Eleven of these genes were integral members of the MHC II antigen presentation pathway. The 24 gene signature was significantly associated with progression free survival (PFS) (HR=0.24; log-rank p=0.00016) and individually, CIITA and CD74 had HR values of 0.167 and 0.49; log-rank p=0.0002 and 0.0164. In a public gene expression database of laser capture micro-dissected breast tumors the MHC II genes HLA-DRB1 and HLA-DPA1 were expressed in the epithelial tumor and stromal components at comparable levels. Further, immunohistochemical analysis detected CD74 and HLA-DPB1 protein in TNBC tumor cells. A large meta-analysis of microarray data from 199 patients with TNBC validated that 10 of the 24 genes (including 6 MHC II genes) were prognostic for PFS with a HR=0.31 (0.19–0.51); log-rank p =9e-07. CD74 alone was similarly prognostic with an HR =0.31 (0.18–0.51); log-rank p=0.0000019.
**Conclusion:** A 24 gene TNBC prognostic signature includes MHC II genes whose individual expression was similarly prognostic. Tumor MHC II antigen presentation pathway is likely an important component of anti-tumor immunity associated with good prognosis in TNBC.
**Introduction**

Triple negative breast cancer (TNBC) is a term utilized clinically to denote women with invasive breast cancer whose tumors lack expression of estrogen receptor (ER-), progesterone receptor (PR-), or overexpression of HER2/Neu. They represent a breast cancer entity in which tumors behave aggressively and are not candidates for ER or HER2/Neu targeted therapy. Most patients receive adjuvant or neoadjuvant chemotherapy with or without local radiation treatment. Patient outcome is heterogeneous with some patients having rapid relapses with a median of two years from diagnosis, and other patients have a low relapse rate from years 5-10 (1). TNBC tumor types vary in their genomic makeup with the majority categorized as basal-like (BL) subtype. In general, BL and non-BL subtypes share similar aggressive biology (2).

Over the last 15 years, there has been a major research effort directed at using genomic techniques to analyze the biology of breast cancer and to establish genomic signatures to assess prognosis (3). These data have been primarily derived from microarray genomic platforms (Affymetrix, Illumina, etc.) with more recent studies using RNA-sequence (seq) technology (4). This has been most notably successful in ER+ breast cancer. Some of these genomic assays have received FDA approval and are used widely to assist therapy decision making in ER+ disease. Prognostic gene expression signatures are not as well developed for TNBC although several large multi-gene signatures have performed well in multivariate analysis (5-7).

Tumor infiltrating lymphocyte (TIL) assessment determined by tumor morphology, immunohistology, and genomic methodologies has also had positive prognostic outcomes in
TNBC (8-11). The conclusions from many of these TIL studies are that the patient’s immune response has a positive effect on PFS, therapy response, and survival, especially in TNBC (12).

In our study, we utilized RNA-seq technology which has multiple advantages over microarray genomic platforms (13, 14). We examined the issue of TNBC prognostic genomic analysis by determining which genes had significantly different expression between patients who relapsed versus those with no relapse. Analysis of the transcriptomes of these samples revealed that a specific immune system pathway, MHC Class II antigen presentation (MHC II), is expressed in tumor cells in TNBC patients with good outcomes. This study provides a means to assess prognosis in TNBC and may also provide a coherent mechanism for the generation of an anti-tumor immune response.

Methods

Patient Material - The Tumor Procurement Shared Facility of the UAB Comprehensive Cancer Center has an IRB approved protocol for collection of tumor and normal tissue samples for research purposes using de-identified clinical data and laboratory analysis. Forty-seven TNBC breast cancer tissues were selected for analysis on the basis that the tumors were ER and PR negative, HER2/Neu not over-expressed, snap frozen tissue available, adequate patient follow-up (>24 months), and the patient had received no anti-cancer therapy prior to tissue collection.

Tissue Processing - The tumor tissue underwent standard macro-dissection by a board certified pathologist (WEG) (see Supplemental Data) to enhance tumor cell content. The de-identified
tumor specimens had >50% tumor nuclei and were shipped on dry ice to HudsonAlpha Institute for Biotechnology (Huntsville, AL).

**RNA-seq** – The 47 tumor specimens were weighed and underwent RNA extraction (see Supplemental Data), quantified, RNA-seq libraries were constructed (16), and were quantified using the Qubit dsDNA High Sensitivity Assay Kit and the Qubit 2.0 fluorometer (Invitrogen). Three barcoded libraries were pooled in equimolar quantities per sequencing lane on an Illumina HiSeq 2000 sequencing machine. They were sequenced using paired-end 50 bp reads and a 6 bp index read to a depth of at least 50 million read pairs per library. The RNA-seq data are publicly available through GEO Accession GSE58135 ([http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE58135](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE58135)).

**RNA-seq Data Analysis** – Gene expression values (fragments per kilobase of transcript per million, FPKMs) were calculated using TopHat v 1.4.1 (16), GENCODE version 9 (17), BEDtools (18), and Cufflinks 1.3.0 with -u option (19) (see Supplemental Data). Identity of subclusters used Consensus Cluster Plus R package (20), and the SAM seq function was used to identify genes differentially expressed between tumors from patients who did or did not relapse with q values of <5% significant. Kaplan-Meier curves and survival analysis were performed using RNA-seq FPKM values and an R script (22) (see Supplemental Data).

**Public microarray data analysis** - Kaplan-Meier and survival analysis was performed on public microarray data using the Kaplan-Meier Plotter tool ([http://kmplot.com/analysis/index.php?p=service](http://kmplot.com/analysis/index.php?p=service)) (23). Patients were censored at the follow-
up threshold (8-10 years). Only JetSet best probe sets were used for each gene in the microarray data analysis (24). Analysis was restricted to the 199 patients whose tumors were ER-, PR-, and were classified as basal intrinsic breast cancer subtype (25).

**Tumor versus Stroma Gene Expression** – Five archived de-identified TNBC tumor specimens underwent standard immunohistochemical analysis with anti-CD74 (Leika/Novocastra) and anti HLA-DPB1 (Sigma-Aldrich). An anatomic pathologist estimated the fraction of antibody positive tumor cells and the localization of the staining (see Supplemental Data). To examine gene expression by epithelial tumor cells versus stroma, we utilized a public laser capture micro-dissection dataset (GEO-GSEJ847) (26). The raw dataset (.cel and matrix files) was uploaded to Partek Genomic Suite (PGS, St. Louis, MO) for data background subtraction, quality control, and RMA-normalization. We selected the 14 invasive TNBC patient samples to compare gene expression in stroma versus epithelial cells.

**Statistics** - Descriptive analysis was provided for patients’ characteristics including student t test and chi-square statistics. The 24 individual genes expression were transformed to best fit a normal distribution using log 2 base (27, 28). High or low expression levels of individual genes were assessed around median value or by tertiles. PFS is defined as the time from diagnosis to the first documented disease progression or death due to any cause, whichever occurs first. Subjects without relapse were considered censored. The Kaplan-Meier method and log-rank test was used to assess the expression difference. The hazard ratio and its 95% confidence interval from the Cox model (29) with Efron's method were reported (30). Pearson correlation coefficient was estimated to examine the correlation among individual genes.
Results

Patients – A total of 47 women with TNBC represent the study population (Table 1). As expected, the relapsed patients had more advanced stage of disease than the no relapse group. Absence of nodal involvement was more prevalent in the no relapse group (60% versus 27%). Adjuvant therapy was similar in both groups with anthracycline combinations used in the majority of patients although five relapse and two no relapse patients received no adjuvant treatment. Similar numbers of patients had conservative surgical management and radiotherapy. Median time to relapse was 18.5 months (8 to 97 months), and the follow-up of non-relapse patients had a median of 96 months (25 to 137 months). Racial makeup of the two groups was similar, and overall 81% of the tumors were basal-like and similar in both groups.

Unsupervised Consensus Clustering Analysis - The analysis identified three main clusters (1, 2 and 3) composed of 20, 17, and 10 patients as illustrated in Figure 1. The cluster analysis did not simply reflect previously defined TNBC subtypes (31, 32). Figure 2 provides the PFS for the three cluster groups. Cluster 2 had improved PFS for the groups as a whole (p=0.023) and subdivided based on lymph node involvement (p=0.013). As seen in Table 2, patient tumors in Cluster 2 have higher expression of immunomodulatory genes than tumors in the other clusters. The rate of relapse in cluster 2 (18%) is significantly lower than in Clusters 1 and 3 (63%); p=.0067.

Gene Expression Analysis – We determined which genes in the transcriptome had significant expression differences between tumors from patients who relapsed compared to tumors from
patients who did not relapse. Table 3 provides the list of 24 genes identified with a false discovery rate (FDR) of 5% (q-value <.05). A heatmap of the 24 genes is provided in Figure 3. All 24 genes exhibited higher expression in tumors from patients who did not relapse. These genes include the major components of the MHC II antigen presentation pathway (Figure 4). There are strong correlations between the expression of the various members of the MHC II pathway across patient samples as demonstrated in Table 4.

Figure 5A depicts the PFS curves for patients whose mean expression of the 24 gene signature was in the top two tertiles compared to the lower tertile. The hazard ratio was 0.24 and a log-rank p-value of 0.00016. Given the strong correlation among overexpressed MHC II pathway genes, we examined the impact of high and low expression on a single gene basis. Figure 5B illustrates the impact of high expression of CIITA on PFS (log-rank p=0.0002; HR=0.167), and 5C illustrates the PFS effect of high expression of CD74 with log-rank p=0.0164 and HR=0.49. A summary of HR values for ten of the MHC II differentially expressed genes is provided in Table 4. Even more dramatic are the PFS curves derived from tertiles (high, intermediate, and low values) for CIITA and CD74 (Figure 6A and B). High values for CIITA are associated with only 2/16 relapses including one with relapse at >90 months. Low values for CD74 are associated with 12/16 relapses, all of which occur within 25 months. With median value as a cut point, CIITA is an independent predictor for PFS (p=0.008) by multivariable Cox regression analysis. When controlling for tumor stage, the HR for high versus low CIITA is 0.147 (CI 0.048 – 0.450). Similarly, CD74 is an independent predictor for PFS (p=0.0322) with a HR of 0.362 (0.143 – 0.917) after adjusting for tumor stage.
Cell Source of MHC II Gene Expression - Classically, MHC II antigen processing and display are attributed to dendritic cells, B cells, and macrophages which are found in tumor stroma. Figure 7 depicts an example of the immunohistochemistry analysis of CD74 (Figure 7A) and HLA-DPB1 (Figure 7B) protein expression in a triple negative breast tumor. In 4 out of 5 TNBC tumor specimens CD74 protein was expressed in >20% of tumor cells and 2 of 5 TNBC tumors had HLA-DPB1 protein expression in >20% of tumor cells. These results are consistent with previous observations of TNBC tumor cells expressing MHC II proteins (33-35). In addition, we examined the expression of MHC II genes in laser capture micro-dissected breast tumor tissues using a publicly available Affymetrix microarray dataset (26, GSE5847). In 14 patients with TNBC, HLA-DRB1 had expression of 11.77 ± 0.78 in stroma and 11.84 ± 0.74 in epithelial tumor; HLA-DPA1 had expression of 9.46 ± 1.08 in stroma and 9.50 ± 1.24 in epithelial tumor. T test and paired T analysis were not significant between stroma and epithelial expression of MHC II genes. These analyses further support the conclusion that TNBC epithelial tumor cells expresses MHC II genes.

Validation of MHC II Expression as a Good Prognosis Signature – In publicly available gene expression repositories there are too few experiments of RNA-seq performed on TNBC tumors with adequate clinical follow-up to be used for validation. As an alternative, we carried out validation on a large meta-analysis of Affymetrix microarray data that was assembled to encompass gene expression profiles from all available breast cancer studies that had adequate clinical follow-up (25). We analyzed 199 patients in this meta-analysis data set with ER-, PR-, basal intrinsic subtype tumors and examined the expression levels of our 24 gene MHC II signature. Seven of the 24 genes were not represented by unambiguous JetSet (24) microarray
probes, and 7 additional genes did not generate survival curves with significant log-rank p-values when analyzed individually. The remaining 10 genes from the 24 gene signature identified in our cohort were adequately represented and each produced PFS curves with significant log-rank test p-values (<0.05). The 10 validated genes are CD74, CTSH, HLA-DPA1, HLA-DPB1, HLA-DRB1, HLA-DRB6, MBNL, CD1E, PTGDS, and TOX. Figure 8A provides the PFS curves of patients whose mean expression of the 10 gene signature was in the top two tertiles compared to the lower tertile. The high expression group has a much better PFS than the low expression group with a HR=0.31 (0.19 – 0.51) with a log-rank of p=9e-07. Further, we examined the ability of CD74 (invariant chain) single gene expression to delineate prognosis. As seen in Figure 8B, CD74 expression delineated patients in the top two tertiles that had a good prognosis with similar predictive power as the 10 gene signature. The HR was 0.31 (0.18 – 0.51) with a log-rank of p=0.0000019. Despite the variation in gene expression measurement technology, multiple institutions and studies included in the public meta-dataset, MHC II expression was validated as being strongly associated with TNBC prognosis.

**Discussion**

This study of 47 TNBC patient tissues used unsupervised consensus cluster analysis to identify 3 clusters (Figure 1) and Cluster 2 had moderate prognostic significance, reduced frequency of patients with relapsed disease (Figure 2 and Table 2), and a large number of immunoregulatory genes (Table 2) similar to prior gene signatures (5-7). Analysis of the differentially expressed genes between relapsed and no relapse patients identified 24 genes (Table 3) including 11 major genes of the MHC II antigen presentation pathway (Figure 4). The 24 gene signature
was strongly prognostic (HR 0.24; log-rank p=0.00016; Figure 5A). The MHC II genes had strong interaction with Pearson correlative coefficients (Table 4) of 0.68 to 0.92 (p=0.0001). Expression levels above or below the median of CIITA or CD74 were strongly prognostic (Figure 5B and 5C) as individual genes. The tertile expression PFS curves for CIITA and CD74 revealed that the high level of CIITA had only 2 relapsed patients, and 1 of them relapsed at 90+ months (log-rank p=0.0054) while low levels of CD74 had 12/16 relapses all within 25 months, reflecting the aggressive early relapse rate in TNBC (Figure 6A and 6B).

Validation of the role of MHC II gene expression in prognosis utilized a public database from a meta-analysis of microarray gene expression from multiple different institutions and studies (23) which included 199 ER/PR-, basal like tumors. The 10 gene MHC II enriched signature produced a highly significant difference in PFS curves (p=9e-07) with a HR=0.31 (0.19-0.51; Figure 8A) as did single gene CD74 (Figure 8B) with log-rank p=1.9e-6 and HR=0.31 (0.18-0.51).

Having validated these MHC II signatures and genes as prognostic, we examined the role of tumor cell or stromal cell expression of these genes. Immunohistology demonstrated CD74 and HLA-DPB1 protein in TNBC tumor cells (Figure 7) similar to other reports (33-35). In addition, some basal-like breast cancer cell lines express MHC II genes or can be induced to express them by gamma interferon or other drugs (36). Further, we examined gene expression of TNBC samples from a public database of laser capture micro-dissected paired epithelial and stroma breast tissues (26). The MHC II proteins HLA-DRB1 and HLA-DPA1 were expressed in the tumor tissue and stroma at comparable levels.
The effects of tumor cell MHC II expression has been explored in the elegant animal model studies of Accolla, et al, (37-40). Using the murine TS/A non-immunogenic syngeneic breast cancer model, he demonstrated that stable transfection of murine CIITA into the tumor cell line induced MHC II expression. Upon injection of the transfected tumor cells into syngeneic animals, the tumors were rejected on days 14-17 rather than proceeding with metastatic disease and death. Further he demonstrated the following: (1) Antitumor immunity and tumor rejection was directly related to the amount of CIITA expression. (2) Tumor rejection required both CD4 and CD8 cells. (3) Post-rejection, animals had lifelong immunity to non-transfected tumor cell challenge. (4) The sequence of tumor rejection was initial CD4 T cell infiltration followed by cytolytic CD8 T cells which rapidly became the predominant cell type (37). Other investigators have made similar observations via transfection of CIITA or MHC II genes (41, 42). Thus, the concept that breast tumor expression of MHC II can induce potent antitumor CD4 and CD8 T-cell immune effects has been well demonstrated in animal models.

The RNA-seq studies described in this report add clarity to the presence and role of MHC II display on the TNBC tumor surface. The strong prognostic influence of the pathway and individual MHC II pathway genes along with correlation of pathway members with each other emphasizes the important role of MHC II pathway on the patient’s anti-tumor immune efficacy. Tumor cells expressing MHC II have been shown to function as antigen presenting cells (APCs) (43, 44). They not only process exogenous proteins/antigens but also have access to endogenous proteins/antigens via autophagosome/lysosome fusion (45, 46). These studies support the hypothesis that MHC II positive tumor cells may function as APCs and have access
to the myriad of patient unique protein mutations (47, 48) for presentation to CD4 T helper cells. Activation of CD4 T helper cells enhance and expand CD8 cytolytic T cells as well as enhance B cell antibody maturation (39, 50). The success of this immune response is reflected in the excellent prognosis of MHC II positive TNBC patients. The importance of CD4 antitumor immune response in anti-tumor immunity has been emphasized in recent reviews (41, 50). The relationship of MHC II tumor cell expression and tumor infiltrating lymphocytes will require ongoing additional studies. It does appear that the use of RNA-seq technology has contributed insight into anti-tumor immunity in TNBC patients.
References


Figure Legends

**TO BE PUBLISHED ONLINE ONLY - Figure 1.** Consensus clustering of gene expression values across all genes identified 3 main groups of TNBC tumors. The heatmap shows the relative similarity of gene expression values in each sample compared to all other samples (darker blue indicates higher similarity). The dendogram at the top of the heatmap shows the pairwise similarity between samples and their assignment into three consensus clusters (Cluster 1 is a light blue bar, Cluster 2 is darker blue, and Cluster 3 is green).

**Figure 2.** Kaplan-Meier PFS curves for patients in Clusters 1-3. (A) PFS curves for Cluster 1 (C-1, black line), 2 (C-2, red line) and 3 (C-3, green line); log-rank p = 0.023. (B) Same as A except patients with lymph node tumor involvement (+) are dashed lines and lymph node negative (-) are solid lines; log-rank p = 0.013.

**TO BE PUBLISHED ONLINE ONLY - Figure 3.** Heatmap of the normalized gene expression values of each of the 24 prognostic genes in each of the 47 patient’s tumors. Red indicates higher expression and green indicates lower expression. Patients who relapsed are grouped on the left, and patients who did not relapse are grouped on the right. Gene identifiers on the right include the common gene symbol followed by the unique gene identifier used by Ensembl.

**TO BE PUBLISHED ONLINE ONLY - Figure 4.** Illustration of TNBC prognostic gene products in the MHC II pathway.
**Figure 5.** Kaplan-Meier PFS curves of patients with the TNBC prognostic genes. (A) The high expression (upper two tertiles) of the 24 gene signature is depicted in red while the lowest tertile is depicted in black; log-rank $p = 0.00016$. (B) The high expression (above the median) of CIITA is depicted in red and low expression (below the median) is depicted in blue; log-rank $p = 0.0002$. (C) The high expression (above the median) of CD74 is depicted in red and low expression (below the median) is depicted in blue; log-rank $p = 0.0164$.

**TO BE PUBLISHED ONLINE ONLY - Figure 6.** Kaplan-Meier PFS curves of patients with expression levels (tertiles) of CIITA and CD74. (A) High (blue solid) intermediate (red), and low (blue interrupted) levels of CIITA; log-rank $p = 0.0054$. (B) High (blue solid line), intermediate (red), and low (blue interrupted) levels of CD74; log-rank $p = 0.0047$.

**TO BE PUBLISHED ONLINE ONLY - Figure 7.** Immunohistochemistry detection of CD74 and HLA-DPB1 protein expression in TNBC tumor tissue. (A) IHC detection of CD74 protein in TNBC tumor tissue shows staining in 20% of invasive tumor cells. Localization is primarily membranous (90%) with some granular cytoplasmic staining (large image is 10x magnification, inset is 20x magnification). (B) IHC detection of HLA-DPB1 protein in TNBC tumor tissue shows staining in 20% of invasive tumor cells. Localization is primarily membranous (90%) with some granular cytoplasmic staining (large image is 10x magnification, inset is 20x magnification).
**Figure 8.** Kaplan-Meier PFS curves of patients based on mean expression levels of the TNBC prognostic genes in the public microarray data set. (A) High expression (upper two tertiles) of the 10 gene signature (red) versus the lowest tertile (black); log-rank p = 9e – 07. (B) High expression (upper two tertiles of CD74 (red) versus lowest tertile (black); log-rank p = 1.9e – 06.
Figure 1
Figure 3
Figure 5A

HR = 0.24 (0.11 - 0.53)
logrank P = 0.00016

Expression
- low
- high

Number at risk
- low: 16, 5, 3, 2, 2, 1, 0, 0
- high: 31, 27, 23, 18, 14, 10, 8, 0

Time (months)
Figure 5B

Product-Limit Survival Estimates

HR (95% CI) 0.167 (0.056-0.496)
Logrank p=0.0002

Survival Probability

Time to Relapse or Last Follow Up (months)

MEDLOG2_CIITA
<=MEDIAN | >MEDIAN

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Event</th>
<th>Censored</th>
<th>Median Survival</th>
<th>95% CL</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;=MEDIAN</td>
<td>24</td>
<td>18</td>
<td>6</td>
<td>19</td>
</tr>
<tr>
<td>&gt;MEDIAN</td>
<td>23</td>
<td>4</td>
<td>19</td>
<td>92</td>
</tr>
</tbody>
</table>
Figure 5C

Product-Limit Survival Estimates

HR (95% CI)
0.349 (0.141-0.865)

+ Censored
Logrank p=0.0164
Figure 6A

Product-Limit Survival Estimates

+ Censored
Logrank p=0.0054

Survival Probability

Time to Relapse or Last Follow Up (months)

TLOG2_CIITA

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Event</th>
<th>Censored</th>
<th>Survival</th>
<th>95% CL</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIGH</td>
<td>16</td>
<td>2</td>
<td>14</td>
<td>92</td>
</tr>
<tr>
<td>INTERMEDIATE</td>
<td>15</td>
<td>8</td>
<td>7</td>
<td>65</td>
</tr>
<tr>
<td>LOW</td>
<td>16</td>
<td>12</td>
<td>4</td>
<td>19</td>
</tr>
</tbody>
</table>

Median
Figure 6B

Product-Limit Survival Estimates

+ Censored
Logrank p=0.0047

Survival Probability

Time to Relapse or Last Follow Up (months)

TLOG2_CD74
HIGH
INTERMEDIATE
LOW

Median
Subjects Event Censored Survival 95% CL
HIGH 16 4 12 . 92
INTERMEDIATE 15 6 9 . 13
LOW 16 12 4 19 12 26
Figure 7B
Figure 8A

HR = 0.31 (0.19 - 0.51)
logrank P = 9e-07
Figure 8B

HR = 0.31 (0.18 - 0.51)
logrank P = 1.9e-06

Expression
- low
- high

Number at risk
- low: 81, 56, 34, 21, 9
- high: 118, 103, 72, 39, 26

Time (months)
Table 1. Patient Demographics

<table>
<thead>
<tr>
<th></th>
<th>No Relapse (25)</th>
<th>Relapse (22)</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (Years) (median)</strong></td>
<td>54 (37-67)</td>
<td>53 (22-76)</td>
<td>NS</td>
</tr>
<tr>
<td><strong>Stage</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>8 (32%)</td>
<td>2 (9%)</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>15 (60%)</td>
<td>12 (55%)</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>2 (8%)</td>
<td>8 (36%)</td>
<td>0.0295</td>
</tr>
<tr>
<td><strong>Tumor Size</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>11 (44%)</td>
<td>3 (14%)</td>
<td></td>
</tr>
<tr>
<td>T2</td>
<td>12 (48%)</td>
<td>10 (45%)</td>
<td></td>
</tr>
<tr>
<td>T3</td>
<td>1 (4%)</td>
<td>3 (14%)</td>
<td></td>
</tr>
<tr>
<td>T4</td>
<td>1 (4%)</td>
<td>5 (23%)</td>
<td></td>
</tr>
<tr>
<td>TX</td>
<td>0</td>
<td>1 (4%)</td>
<td>0.0029</td>
</tr>
<tr>
<td><strong>Nodes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>10 (40%) (N1:4, N2:2, N3:4)</td>
<td>16 (73%) (N1:11, N2:3, N3:2)</td>
<td>0.0394</td>
</tr>
<tr>
<td>No</td>
<td>15 (60%)</td>
<td>6 (27%)</td>
<td></td>
</tr>
<tr>
<td><strong>Grade</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>5 (20%)</td>
<td>5 (23%)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>20 (80%)</td>
<td>17 (77%)</td>
<td>NS</td>
</tr>
<tr>
<td><strong>Adjuvant Chemotherapy</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anthracycline Based Chemotherapy</td>
<td>19 (76%)</td>
<td>13 (59%)</td>
<td></td>
</tr>
<tr>
<td>Other Chemotherapy</td>
<td>4 (16%) (2 CBP based)</td>
<td>4 (18%) (1 CBP based)</td>
<td></td>
</tr>
<tr>
<td>No Adjuvant Chemotherapy</td>
<td>2 (8%)</td>
<td>5 (23%)</td>
<td>NS</td>
</tr>
<tr>
<td><strong>Surgery</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conservative</td>
<td>13 (52%)</td>
<td>7 (32%)</td>
<td></td>
</tr>
<tr>
<td>Mastectomy</td>
<td>12 (48%)</td>
<td>15 (68%)</td>
<td>NS</td>
</tr>
<tr>
<td><strong>Radiotherapy</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>13 (52%)</td>
<td>11 (50%)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>12 (48%)</td>
<td>11 (50%)</td>
<td>NS</td>
</tr>
<tr>
<td><strong>Time to Relapse (Months) (Median)</strong></td>
<td>N/A</td>
<td>18.5 months (8 to 97)</td>
<td></td>
</tr>
<tr>
<td><strong>Disease Free Survival (Months) (Median)</strong></td>
<td>96 months (25 to 137)</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>14 (56%)</td>
<td>15 (68%)</td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>10 (40%)</td>
<td>6 (28%)</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>1 (4%)</td>
<td>1 (4%)</td>
<td>NS</td>
</tr>
<tr>
<td><strong>Breast Cancer Subtype</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basal Like</td>
<td>19 (76%)</td>
<td>19 (86%)</td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>6 (24%)</td>
<td>3 (14%)</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td>Relapse</td>
<td>Basal-like Subtype</td>
<td>Number of Genes that are Specifically Upregulated (FDR &lt;0.01)</td>
</tr>
<tr>
<td>--------</td>
<td>----------</td>
<td>--------------------</td>
<td>---------------------------------------------------------------</td>
</tr>
<tr>
<td>Cluster 1</td>
<td>12/20 (60%)</td>
<td>(18/20 (90%))</td>
<td>806</td>
</tr>
<tr>
<td>Cluster 2</td>
<td>3/17 (18%)</td>
<td>13/17 (76%)</td>
<td>839</td>
</tr>
<tr>
<td>Cluster 3</td>
<td>7/10 (70%)</td>
<td>7/10 (70%)</td>
<td>71</td>
</tr>
</tbody>
</table>
Table 3. Genes (24) Elevated in No Relapse Patients

I. MHC II Pathway
1. CIITA (activator of MHC II)
2. CD74 (invariant chain; chaperone for all MHC II)
3. HLA-DPA1 (peptide presentation to T cells)
4. HLA-DPB1 (peptide presentation to T cells)
5. HLA-DPB2 (peptide presentation to T cells)
6. HLA-DQA1 (peptide presentation to T cells)
7. HLA-DRB1 (peptide presentation to T cells)
8. HLA-DRB5 (peptide presentation to T cells)
9. HLA-DRB6 (peptide presentation to T cells)
10. CTSH (cathepsin H; endosomal protease)
11. NCOA1 (MHC II nuclear co-activator)

II. MHC I Pathway
12. CD1E (MHC I-like; lipid presentation to T cells)
13. FCGRT (MHC I-like: Fc receptor transporter)

III. Possible MHC-related
14. KRT14 (keratin 14 – epithelial cytoskeleton)
15. LPAR5 (membrane protein involved in endocytosis)
16. FGD3 (regulates actin cytoskeleton and cell shape)
17. VAMP2 (vesicle associated membrane protein)

IV. Other
18. LRRK2 (leucine rich repeat kinase – mitochondria)
19. MBNL1 (regulator of splicing specific pre-RNA targets)
20. NTRK3 (neurotrophic tyrosine receptor kinase)
22. PTGD5 (prostaglandin D2 synthase – neuromodulator)
23. SH3BGRL – uncertain
24. TOX – DNA binding protein
<table>
<thead>
<tr>
<th>Genes</th>
<th>CIITA</th>
<th>CD74</th>
<th>CTSH</th>
<th>DPA1</th>
<th>DPB1</th>
<th>DPB2</th>
<th>DQA1</th>
<th>DRB1</th>
<th>DRB5</th>
<th>DRB6</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIITA</td>
<td>1</td>
<td>0.84</td>
<td>0.68</td>
<td>0.84</td>
<td>0.92</td>
<td>0.86</td>
<td>0.83</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
</tr>
<tr>
<td>CD74</td>
<td>0.83</td>
<td>1</td>
<td>0.86</td>
<td>0.98</td>
<td>0.86</td>
<td>0.79</td>
<td>0.92</td>
<td>0.92</td>
<td>0.87</td>
<td>0.87</td>
</tr>
<tr>
<td>CTSH</td>
<td>0.68</td>
<td>0.86</td>
<td>1</td>
<td>0.85</td>
<td>0.75</td>
<td>0.72</td>
<td>0.82</td>
<td>0.74</td>
<td>0.74</td>
<td>0.74</td>
</tr>
<tr>
<td>DPA1</td>
<td>0.84</td>
<td>0.98</td>
<td>0.85</td>
<td>1</td>
<td>0.88</td>
<td>0.78</td>
<td>0.93</td>
<td>0.89</td>
<td>0.82</td>
<td>0.82</td>
</tr>
<tr>
<td>DPB1</td>
<td>0.92</td>
<td>0.86</td>
<td>0.75</td>
<td>0.86</td>
<td>1</td>
<td>0.92</td>
<td>0.86</td>
<td>0.75</td>
<td>0.78</td>
<td>0.78</td>
</tr>
<tr>
<td>DPB2</td>
<td>0.86</td>
<td>0.77</td>
<td>0.72</td>
<td>0.78</td>
<td>0.92</td>
<td>1</td>
<td>0.76</td>
<td>0.65</td>
<td>0.70</td>
<td>0.70</td>
</tr>
<tr>
<td>DQA1</td>
<td>0.83</td>
<td>0.92</td>
<td>0.82</td>
<td>0.93</td>
<td>0.86</td>
<td>0.76</td>
<td>1</td>
<td>0.88</td>
<td>0.84</td>
<td>0.84</td>
</tr>
<tr>
<td>DRB1</td>
<td>0.71</td>
<td>0.92</td>
<td>0.74</td>
<td>0.89</td>
<td>0.75</td>
<td>0.65</td>
<td>0.88</td>
<td>1</td>
<td>0.74</td>
<td>0.94</td>
</tr>
<tr>
<td>DRB5</td>
<td>0.71</td>
<td>0.87</td>
<td>0.74</td>
<td>0.82</td>
<td>0.78</td>
<td>0.70</td>
<td>0.84</td>
<td>0.94</td>
<td>1</td>
<td>0.85</td>
</tr>
<tr>
<td>FGD3</td>
<td>0.84</td>
<td>0.85</td>
<td>0.85</td>
<td>0.83</td>
<td>0.86</td>
<td>0.80</td>
<td>0.82</td>
<td>0.68</td>
<td>0.73</td>
<td>1</td>
</tr>
</tbody>
</table>

**Progression Free Survival Hazard Ratio (> vs < median value)**

| HR (95% CI) | 0.167 (0.056-0.496) | 0.349 (0.141-0.865) | 0.225 (0.083-0.611) | 0.292 (0.114-0.751) | 0.234 (0.086-0.638) | 0.241 (0.088-0.654) | 0.280 (0.109-0.719) | 0.202 (0.074-0.552) | 0.190 (0.069-0.523) | 0.263 (0.101-0.684) |
| p Value**   | 0.0013 | 0.0230 | 0.0035 | 0.0106 | 0.0045 | 0.0052 | 0.0081 | 0.0018 | 0.0013 | 0.0062 |

* Univariate Cox Regression Model