Award Number: W81XWH-09-1-0069

TITLE: Defining Genomic Changes in Triple Negative Breast Cancer in Women of African Descent

PRINCIPAL INVESTIGATOR: Lisa Baumbach, PhD

CONTRACTING ORGANIZATION: University of Miami
Miami, Florida 33136

REPORT DATE: June 2010

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT:

☑ Approved for public release; distribution unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.
Breast cancer (BC) is the second leading cause of cancer death among African-American (AA) women, with a 20% greater death rate than in Caucasians (Cauc). However, the basis for such disparity remains an enigma. Recent observations from our laboratory suggest the involvement of novel genes contributing to AA BC risk. Tumor and normal breast tissue samples from Cauc and AA patients were obtained from the UM/Sylvester Breast Tissue Bank (UM/S BTB). Our main goal was to identify which genes were turned “on” (expressed) and which genes were turned “off” (using gene “chip” technology) between tumor and normal samples that were both common and unique among ethnic groups. Based on analysis of 22,000 genes, some common changes in breast cancer were identified, as well as ethnic specific gene expression patterns that may provide important new insights into ethnic differences in clinical outcomes. In this study, we proposed to extend these preliminary findings to a much larger native African tumor bank [available via collaboration between Drs. Peter A. Bird (a missionary surgeon in Kijabe, Kenya) and Mark Pegram (UM Sylvester).] Additionally, we propose to analyze structural chromosomal alterations associated with gene expression differences utilizing advanced genomic techniques (in collaboration with Alan Ashworth, England).

The underlying hypothesis of the overall project is that discrete genomic alterations and gene expression changes will be identified and shared between triple negative tumor specimens within an ethnic group, i.e., North Americans/African descent and Kenya. The method in which this is being investigated includes three approaches: gene expression studies in breast cancer samples of African-American ancestry vs. native African samples, investigation of possible chromosomal or copy number alterations in the same samples and an analysis of a panel of ancestry-informative DNA markers. The study cohort consists of 50 African-American breast cancer samples and 50 Kijabe (Kenyan) Native African breast cancer samples. This work will contribute to development of rationale designs of preventive, predictive and therapeutic measures for BC in different ethnicities, and thus, a significant reduction in current ethnic-specific disparities in BC incidence, morbidity and mortality.
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>4</td>
</tr>
<tr>
<td>Body</td>
<td>5</td>
</tr>
<tr>
<td>Key Research Accomplishments</td>
<td>10</td>
</tr>
<tr>
<td>Reportable Outcomes</td>
<td>10</td>
</tr>
<tr>
<td>Conclusion</td>
<td>11</td>
</tr>
<tr>
<td>References</td>
<td>11</td>
</tr>
<tr>
<td>Appendices</td>
<td>12</td>
</tr>
</tbody>
</table>
3) INTRODUCTION
Staggering statistics arise from analysis of cancer disparities. Breast Cancer (BC) is the second leading cause of cancer death among African-American (AA) women (1). Mortality is 20% greater than that in Caucasian (Cauc) women, and is partially attributed to more aggressive disease and poorer prognosis. In addition, AA women ≤ 50 years have the highest rate of new BC cases in the US (1,2). General consensus exists that AA women of all ages are more likely to have poorly differentiated BC, which is likely to occur at an earlier age, be ER and PR negative, and to have a higher proliferative fraction - all factors associated with more aggressive tumorigenicity (2). Therefore, the prognosis in AA patients is worse, even adjusted for stage of presentation. Ethnic-specific differences in response to adjuvant therapy have also been reported (3,4). Taken together, the cumulative data suggests that intrinsic, ethnic-specific, and biological/genetic differences contribute to disparities in BC morbidity and mortality. Very few studies have directly compared multi-ethnic differences in breast cancer incidence, staging, tumor characteristics, and mortality in the United States.

A number of articles have been written concerning the incidence and types of breast cancer among native Africans in Nigeria and Kenya (5-10). All of these studies are in agreement that Nigerian women develop BC at an earlier age and suffer high mortality. An important recent study by Bird et al (11) focused on a cohort of BC patients from the Kijabe Hospital in Kenya and reported a very low frequency of hormone receptor expression: 24% ER-positive and 34% ER-or PR-positive tumors. Compared to BC in Western or Cauc populations, the Kijabe patients have a high proportion of poorly differentiated, advanced cancers and irrespective of disease stage, were much less likely to be hormone sensitive (ER and PR negative). Overall, the possibility of inherently more aggressive tumor biology, coupled with low hormone receptor sensitivity and lack of HER2, may represent manifestations of modified biology in African populations. Further research is urgently needed to fully characterize the tumors in the Kijabe clinical cohort, and to determine whether BC subtypes with a poor prognosis are overrepresented in the African race. Comparative molecular studies of BC in indigenous African populations, African Americans, and Caucasian Americans are needed to allow genetic and molecular detailing of these tumors that appear to have modified biology relative to breast cancers in other populations.

We have developed a collaboration with Almac Diagnostics (www.almacgroup.com). In an early pilot study, samples from three AA and three Cauc BC patients (matched for age; ER+/PR+/Her 2-) demonstrated feasibility of the Almac platform in formalin-fixed, paraffin-embedded (FFPE) samples obtained from UM Pathology laboratories. Next, a broader multi-ethnic project focusing on triple negative breast cancer cases in AA, His and Cau patients was initiated. An investigation of 28 cases (10 AA cases, 8 Cauc cases and 10 His cases). Analysis of the detected transcripts for each patient demonstrated low variability between samples and a high number of detected transcripts. Across all 56 samples (28 tumor and 28 matched normal), greater than 22,000 transcripts were detected significantly above background (data not shown). Further statistical analyses (including ANOVA and T-tests) resulted in the identification of 2,622 statistically significant, differentially expressed genes. The 2,622 genes were used as input sequences for the expression pattern identification across the six sample groups (tumor and normal in 3 ethnic groups). In the tumor samples, six ethnic-specific expression patterns were identified in the resultant heatmap. The subset of differentially expressed genes was analyzed using real-time PCR to validate gene expression differences. To summarize, even in a small pilot study, we have identified both common and ethnic-specific transcriptional changes between AA and Cauc breast cancer samples. If confirmed in larger data sets, these observations may well provide important new insights into molecular mechanisms of ethnic subtype differences in clinical outcomes.

The current project is a study of the potential differences or similarities between a cohort of African-American triple negative cases with a cohort of triple negative cases from Kijabe (Kenya) obtained through collaboration with Dr. Peter Bird (Nairobi, Kenya). It applies the gene expression array technology described above in addition to investigation of possible chromosomal or copy number alterations in the same samples.
(through collaboration with Dr. Alan Ashworh, England) and an analysis of a panel of ancestry-informative DNA markers (through collaboration with Dr. Rick Kittles, Chicago).

4) BODY

Task 1: Obtain UM IRB and USAMRMC HRPO Approvals. Status: Completed

Approval by the UM Institutional Review Board for Human Subjects research was obtained on May 6, 2009. It is UM protocol number 20081188. The title of the protocol is the same as the grant “Defining Genomic Changes in Triple Negative Breast Cancer in Women of African Descent” It was approved as Exempt status.

Task 2: Transport of Kijabe samples to the University of Miami. Status: Completed

We received the Kijabe native African breast cancer samples from Dr. Peter Bird, the last were received in December 2009. To date we have received 64 samples from Africa. Dr. Bird and Dr. Pegram will be meeting during the ASCO meeting in Chicago this June to discuss the progress on the project.


In the original proposal we indicated that we would need to determine the HER2 status of the Kijabe samples as it is not commonly performed during the cancer treatment for these individuals. ER and PR receptor status assays are available for most of the samples. We evaluated an initial subset of 10 Kijabe samples for ER, PR and Her2 status in the UM Pathology laboratory. Comparison with the available ER and PR status from Africa showed a couple of discrepancies in ER or PR status. These samples were repeated and the discrepancies remained, therefore, it was decided that we should redo ER and PR status testing for all of the Kijabe samples, as well as testing Her2 status. With the addition of this work we have not completely finished receptor status testing for all of the Kijabe samples, but anticipate completion in a few weeks. All of the UM African-American samples have ER/PR and Her2 status already available.


Separate extraction procedures are done for the extraction of DNA or RNA from the FFPE tumor and normal tissue subsamples for each patient sample. The quality or purity of the extracted samples is evaluated looking at the overall concentration, the A260/280 ratio and the presence/quality of distinct 18S and 28S peaks using a bioanalyzer for RNA. The UM Pathology samples have shown a consistent high yield of good quality RNA/DNA. Extractions of the Kijabe African samples is not yet complete, but will be completed when the final ER/PR/Her2 status information is available, which will allow for preferential treatment of the triple negative samples.
Task 5: Aim I, Analyze and compare genome-wide transcript expression in BC samples of AA ancestry vs. native African (Kijabe) samples. Status: Underway

To date about half of the total African-American samples have been hybridized to the Almac Breast Cancer DSA chip and a selected few Kijabe samples have been hybridized. Numerous Caucasian samples have also been analyzed (see Pegram PI report for sample tables). The quality of resulting data was excellent. Detailed steps in quality control and data analyses are further discussed below.

A) Distribution Analysis: The distribution of the sample data (histogram of normalised intensity values) is assessed to determine what statistical tests, metrics; etc should be applied. Data showed a normal (Gaussian) distribution, therefore a Pearson correlation, which is based on the normal distribution, was chosen for clustering.

![Data Distribution for initial African-American samples on DSA arrays.](image)

**Figure 1.** Data Distribution for initial African-American samples on DSA arrays. Data distribution was examined using all transcripts on the breast cancer DSA chip. Each sample group was from combining all samples from an ethnic group with same condition (tumor or normal). The data distribution is close to normal, indicating it is suitable for the data analysis.

B) K-Mean Clustering: Groups were created which show the relationships among the expression levels of conditions (tumor vs. normal) or samples. This allowed identification any spurious samples. K-Mean was used because of prior knowledge of sample condition either from tumor or normal tissue. Samples falling into a group that does not match their biological condition would suggest potential abnormality or contamination of tumor and normal tissue.

Inclusion of one contaminated “normal” sample from an African-American patient in the two-dimentional K-mean clustering analysis showed that this “normal” sample exhibited tumor-like gene expression profile and fell into the tumor cluster (see Figure 2 below).
Figure 2. Identification of contaminated cells in an FFPE sample. On the far right, one dissected FFPE slice of normal cells exhibits similar gene expression pattern to tumor cells.

C) Principal Components Analysis: This data analysis technique is a decomposition technique that produces a set of expression patterns known as principal components. Linear combinations of these patterns can be assembled to represent the behavior of all of the genes in a given data set. PCA is not a clustering technique; however the aim of PCA is similar to that of clustering. It is a tool to characterize the most abundant themes or building blocks that reoccur in many genes in the experiment. See the Pegram PI report for illustration of the PCA results.

D) Native African samples yield good quality RNA and Data:
Illustrated below is a preliminary analysis of the first Kijabe samples. These samples have been through the data quality and analysis steps described above. The heatmap displayed here is a comparison of triple negative Kijabe, Native African samples to adjacent normal tissue samples from African-Americans. The Kijabe samples clearly cluster together separate from the African-American samples. Some differences can be seen within the African-American samples and within the Kijabe samples. These differences may be due to different nodal status but are being evaluated in the correlation with clinical parameters (Task 10).
Figure 3. Cluster diagram from analysis of the initial triple negative Native African samples. This comparison shows that high quality RNA can be extracted from the Kijabe samples.

Task 6: Aim II, Investigate possible chromosomal alterations associated with gene expression differences using array cGH. Status: Underway

DNA extracted from each tumor and normal sample are being hybridized to high density cGH arrays in the laboratory of Dr. Alan Ashworth at the Breakthrough Cancer Center in England. Data is assessed for quality and analyzed for copy number variation and possible chromosomal aberrations. The array being used is an aCGH 32K tiling path microarray, which has a complete coverage of the whole genome with a resolution of 50kb.

A) DNA labelling and array hybridization

Reference (pool normal DNA extracted from blood lymphocytes from 24 individuals) and tumor DNA samples (300 ng aliquots) are labeled with Cy3- or Cy5-conjugated dCTP respectively (Amersham Biosciences, Buckinghamshire, UK) using random primer BioPrime DNA labeling. Labeled reference and tumor DNA are co-ethanol precipitated with 100 µg human Cot-1 DNA (Invitrogen Life Technologies) and re-suspended in 45 µl hybridization buffer; denatured and incubated at 37 °C to allow blocking of repetitive sequences by human Cot-1 DNA. Denatured DNA samples are then hybridized to the microarray at 37 °C for 18 h. The arrays are then washed and dried by centrifugation. aCGH slides are scanned using an Axon 4000B scanner (Axon Instruments, Burlingame, CA, USA) and images will be processed using Genepix Pro 4.0 image analysis software (Axon Instruments). See Figure 4 below for an example of hybridization of a Kijabe Native African case.
B) Data analysis

Cases with >10% of clones missing and clones for which data are not available in ≥10% of cases will be excluded. Log2 ratios will be normalized for spatial and intensity dependent biases using a two-dimensional loess regression followed by a BAC-dependent bias correction. The final dataset of BAC clones with unambiguous mapping information according to the build hg19 of the human genome (http://www.ensembl.org) are used for further analysis. Analysis of samples is currently underway in Dr. Ashworth’s lab. Our lab in Miami is in regular contact with Dr. Ashworth’s lab and Dr. Ashworth has visited Miami, while in the US. Drs. Baumbach and Pegram also meet with Dr. Ashworth at cancer meetings such as the AACR. Co-analysis of expression array and array cGH data to correlate changes in gene expression with gene copy number will be completed in Year 2 of the project.

Task 7: Aim III, Analyze ancestry of each sample using a panel of ancestry-informative DNA markers (AIMs). Status: Started—Scheduled to start beginning of Year 2

A sample of the DNA extracted from the North American African samples is being used by the laboratory of Dr. Rick Kittles to survey a set of 100 ancestry informative markers. These studies will allow for accurate estimates of European ancestries in the AA subjects, and allow use of individual ancestry estimates as additional covariates in overall experimental analyses. We have sent an initial set of samples to Dr. Kittles to begin this aim of the project, data from the samples is not yet available.

Task 8: Aim I, Independent validation of gene expression differences using quantitative real-time PCR. Status: Beginning—continues in Year 2

Quantitative real-time PCR of a subset of statistically significant genes is being used as an independent validation of gene expression differences. Additionally a set of genes will be investigated to validate the combined data from the gene expression arrays and cGH arrays. As data analysis for both gene expression and array CGH is ongoing, a final list of genes to be evaluated by real-time PCR is not yet available. We have selected a subset of genes to be tested, such as the Estrogen Receptor gene, to investigate congruence with the gene expression data and the ER negative status of the samples. The total number of genes analyzed
will a minimum be 10 over expressed genes and 10 under expressed genes, but the actual percentage of differentially expressed genes is to be determined after the total number of differentially expressed genes is known.

5) KEY RESEARCH ACCOMPLISHMENTS

- **Task 1**: Obtain UM IRB and USAMRMC HRPO Approvals. Completed.
- **Task 2**: Transport of Kijabe samples to the University of Miami. Completed.
- **Task 3**: Determination of HER2 (and ER/PR) status in the Kijabe clinical cohort using fluorescence in situ hybridization. Underway.
- **Task 5**: Aim I, Analyze and compare genome-wide transcript expression in BC samples of AA ancestry vs. native African (Kijabe) samples Underway.
- **Task 6**: Aim II, Investigate possible chromosomal alterations associated with gene expression differences using array cGH. Underway.
- **Task 7**: Aim III, Analyze ancestry of each sample using a panel of ancestry-informative DNA markers (AIMs). Started—scheduled for Year 2.
- **Task 8**: Aim I, Independent validation of gene expression differences using quantitative real-time PCR—Beginning—continues in Year 2.

6) REPORTABLE OUTCOMES


7) CONCLUSION

We have demonstrated the utility and reliability of using archived FFPE tissues (including those obtained from native Africa) to complete the investigations proposed in this award – namely gene expression profiling and CGH array analysis. Prior to these investigations, there has not been a published study that simultaneously evaluates genome-wide gene expression differences between tumor and self-matched normal breast tissue in a series of African/AA and Cauc patients, and correlates these results to potential chromosome/DNA alterations in those same tumor specimens. This study will allow for the first direct comparison of gene expression/genomic copy number data in triple negative tumor specimens across Americans of African descent and Kenyan East Africans. These data sets will be posted to publically-accessible online microarray databases for immediate availability to the research community. Importantly, we have engaged an advocate who has participated in group meeting and will continue to help guide us in dissemination of our findings to the African American community. We will correlate all experimental data with a spectrum of clinical data available on study subjects, and determine possible correlations between genomic signatures, genomic changes, clinical tumor characteristics and outcomes/ response measures among and across ethnic groups. We are well on our way to completing the proposed tasks in the proposed timeframe.

8) REFERENCES


9) **APPENDICES** – sample poster presentation; American Association for Cancer Research, Special Symposium on Breast Cancer, San Diego, CA 2009; *Manuscript in preparation available on request.*
Defining multi-ethnic genome-wide transcriptional signatures in normal and cancerous breast tissue using paraffin embedded samples

Lisa L. Baumbach1, Mary Ellen Ahearn1, Carmen Gomez2, Aldo Mejias1, Merce Jorda1, Tom Halsey2, Jim Yan2, Kevin Ellison2, Karl Mulligan2, Stefan Gluck1 and Mark Pegram2

1) Miller School of Medicine, University of Miami, Miami, FL; 2) Almac Diagnostics, Durham, North Carolina

Overall Study Design
The overall goal of the project is to investigate possible ethnic differences in gene expression in breast cancer when patient samples are matched for age, stage of disease and hormone receptor status. For each sample normal tissue from the same woman is used as a control to evaluated gene expression from the tumor tissue. The final study will include 10 each African-American, Hispanic white and non-Hispanic white (Caucasian) women.

The study was originally designed to use fresh tissue samples but with the advent of the Breast Cancer DSA™ Research Tool has changed to using Formalin Fixed Paraffin Embedded (FFPE) samples.

Patient Study Criteria:
Age 60 years or less
No exposure to chemotherapy
Triplet Non-her hormone receptor status

Methods for RNA Extraction and Hybridization
Patient samples were obtained from the University of Miami Medical School Pathology Department, under IRB approval, as anonymous samples. All samples were from women less than 60 years of age and were known to be ER+/PR+/Her2-.

For each patient, pathologists cut new sections from paraffin-embedded sample blocks. Patient samples were sent to Almac Diagnostics for processing and hybridization to the Breast Cancer DSA™ Research Tool.

The following steps were performed by Almac Diagnostics:
1. Isolation of RNA from each sample
2. RNA Amplification using the NuGEN FFPE System
3. Generation of First Strand cDNA from 10-80 ng of total RNA
4. Generation of a DNA/RNA Heteroduplex Double Strand cDNA and amplification
5. cDNA Fragmentation and Labelling using a NuGEN System
6. Affymetrix hybridization washing, staining and scanning protocol applied to the Breast Cancer DSA™ Research Tool.

Cluster Analysis of Individual Tumor and Matched Normal Samples from Patients of Three Ethnicities

Current ER/PR/Her2- Breast Cancer Patients—FFPE Samples

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>Normal</th>
<th>Tumor</th>
</tr>
</thead>
<tbody>
<tr>
<td>African-American</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Caucasian</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Hispanic</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

Methods for RNA Extraction and Hybridization

Cluster Analysis of Patient Samples from Three Ethnicities

Three main dendrogram groupings of transcripts are seen: on the left are the clustered tumor samples, second (the middle group) are the self matched normal cells adjacent to the patient tumor cells (termed “tumor normal”), lastly on the right is a cluster of “normal normal” samples obtained from non-cancer patients undergoing reduction mammoplasty. (up-regulation of genes red and down-regulation green).

Normal Normal samples included 4 African-American, 3 Hispanic and 3 Caucasian samples.

The 2-dimensional clustering of intensity data included 1442 probe sets selected based on their intensity (Intensity>Background+3stdev, p-value <0.01, 2-way ANOVA p <0.01 (With MTC by Benjamini-Hochberg FDR).

Summary and Conclusions
Gene expression differences have been demonstrated to exist between BC “tumor normal” and “normal normal” tissue; suggesting that the tumor microenvironment has a strong influence upon surrounding “normal” cells.

Current Status of Project:
1. We are selecting approximately 10 differentially expressed genes for validation by qRT-PCR, and we are extracting DNA from a subset of these specimens for hybridization to high-density SNP arrays, to assess possible DNA copy number variations (CNVs) and/or LOH in tumor samples.

2. We are currently adding triple negative breast cancer samples from native Africans (Kijabe, Kenya) to the study. Initial studies show that high quality RNA can be obtained from the samples, and show gene expression differences from African-American samples.

The ability to obtain high quality RNA expression data from FFPE samples, illustrated here, offers new possibilities for genetic studies. Completion of this study should result in significant findings regarding genome-wide alterations associated with BC in several ethnic/racial groups, and increase understanding of the biological basis of ethnic-specific disparities in BC occurrence, mortality and therapeutic response.