AD_____

Award Number: W81XWH-09-1-0068

TITLE:
DEFINING GENOMIC CHANGES IN TRIPLE NEGATIVE BREAST CANCER IN WOMEN OF AFRICAN DESCENT

PRINCIPAL INVESTIGATOR:
Dr. Mark Pegram
Dr. Lisa Baumbach

CONTRACTING ORGANIZATION:
University of Miami
Miami, FL 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

□

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
01/June/2010	Annual	1 JUN 2009 - 31 MAY 2010

4. TITLE AND SUBTITLE
Defining Genomic Changes in Triple Negative Breast Cancer in Women of African Descent

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-09-1-0068

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Mark Pegram, MD
Email: mpegram@med.miami.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Miami

Miami, FL 33136

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
US Army Medical Research and Materiel Commmand
Ft. Detrick, MD 21702

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Background: Breast cancer (BC) is the second leading cause of cancer death among African-American (AA) women, with mortality 20% greater than that in Caucasians (Cauc). However, the basis for such disparity remains an enigma. Recent observations from our laboratory suggest the involvement of unidentified genes contributing to AA BC risk. Matched tumor and normal FFPE samples from Cauc and AA patients were obtained from the UM /Sylvester Breast Tissue Bank (UM/S BTB) under an IRB-approved protocol. Based on analysis of 22,000 transcripts, ethnic specific gene expression patterns were identified that may provide important new insights into molecular mechanisms of ethnic subtype differences in clinical outcomes. We propose to extend these preliminary findings to a large African tumor bank [available via collaboration between Drs. Peter A. Bird (Kijabe, Kenya) and Mark Pegram (UM Sylvester).] Additionally, we propose to analyze chromosomal alterations associated with gene expression differences utilizing array cGH (in collaboration with Alan Ashworth, England). 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. **Hypothesis**: ***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 decent and Kenya.*** **Aim I:** ***Analyze and compare genome-wide differences in gene expression in BC samples of AA ancestry vs. native African (Kijabe) samples (Drs. Pegram, Baumbach, Bird, Halsey). Aim II: Investigate possible chromosomal alterations associated with gene expression differences*** (***Drs. Pegram, Baumbach, Ashworth).*** **Aim III:** ***Analyze ancestry of each sample using a panel of ancestry-informative DNA markers (Drs. Kittles, Baumbach).*** **Synergy Statement:** The proposed investigations are highly synergistic. This study will also 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. We will correlate all experimental data with a spectrum of clinical data available on study subjects, and apply covariate modeling and logistic regression analysis to determine possible correlations between genomic signatures, genomic changes, clinical tumor characteristics and outcomes/ response measures among and across ethnic groups.

15. SUBJECT TERMS Triple negative breast cancer, Ethnic disparities, Breast cancer amongst African Americans and Africans, Gene expression profiling, Array CGH.

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	15	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

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

3) INTRODUCTION

The advent of microarray technology has enabled the robust, high throughput analysis of disease specific transcriptomes, including those in breast tumor specimens. Indeed, the molecular classification of breast cancer has been revolutionized by the advent of gene expression profiling. However, currently available commercial microarray design focuses on the most commonly known and characterized genes from all body tissues, therefore only a subset of genes on a generic microarray will yield informative results for any tissue-specific study. Moreover, since the transcriptome of a given tissue contains tissue/disease-specific splice variation as well as non-coding RNAs, many important transcripts solely expressed in the tissue of interest will not be represented. One innovative solution to this problem that we will utilize in this project is to exploit custom breast cancer-specific arrays developed by our collaborators at Almac Diagnostics. With tens of thousands of transcripts not found on generic arrays, specificity of differential gene expression patterns will be significantly enhanced. Furthermore, the use of expression array technology historically has been dependent upon the availability of intact RNA from fresh frozen tumor tissue for analysis, thus study of the many large retrospective cohorts with annotated clinical follow-up has not been possible. RNA extracted from FFPE samples tends to have shorter median length from 3' to 5' and the detection of these transcripts on generic array platforms is rarely successful. However, using an innovative approach we have recently successfully tested novel array probes specifically designed to detect partially degraded RNA from formalin-fixed, paraffin-embedded (FFPE) breast tumor material from samples at the University of Miami. The use of a probeset with extreme 3' sequence mitigates this previous technical limitation, and thus is considered highly innovative (Figure 1).

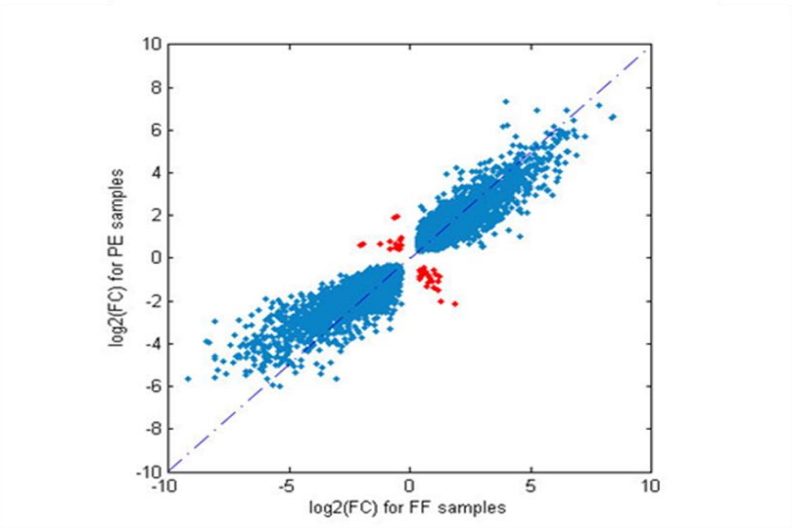

Figure 1. Retention of differentially expressed transcripts

Sign matrix for differentially expressed transcripts: 99.5% of transcripts retain sign. Spearman correlation coefficient for the fold changes in FFPE and FF is $\rho = 0.95$. The ability to extract useful information from FFPE samples up to 13 years of age has been demonstrated.

Another innovation in this study is the genomic analysis of a published East African breast cancer cohort, the largest of its kind from the region. Importantly, the integration of high density array cGH technology with the expression array data is highly innovative (to our knowledge, the first study of this kind in a native African, or even African American cohort). This approach will allow identification of ethnic specific copy number variation and loss of heterozygosity, and their relation to gene expression changes. Finally, the incorporation of an ancestry marker panel makes this a particularly novel study which is sure to produce data of interest to the community. Our eventual goal will be to develop further understanding of biology of disease, prognostic biomarkers, and eventually, the targets for therapeutics for ethnic-specific subgroups in breast cancer.

Defining Genomic Changes in Triple Negative Breast Cancer in Women of African Descent
4) BODY

Quality Assessment of the Gene Expression Data

The gene expression profiling data were generated from Almac Diagnostics Breast Cancer Disease Specific Array (DSA). Totally, 75 FFPE samples were profiled, including 72 samples detailed in table 1 and 3 additional Native African Triple Negative tumor samples from Kenya. The Breast Cancer DSA was designed and supplied by Almac Diagnostics.

Table 1: FFPE samples from Three Ethnic Groups.

Patients	Normal Biopsy	Patient-matched tumor cell	Patient-matched normal cell
African American	4	11	11
Caucasian American	3	9	9
Hispanic American	3	11	11

The Quality of DSA chip was assessed on the basis of parameters automatically extracted from GCOS report (RPT) files per chip using MATLAB script based web application developed in Almac Diagnostics. Data pre-processing used Resolver Error Model. All parameter including Raw Q, background, Scaling Factor and all the controls met the quality criteria set by Affymetrix and Almac Diagnostics SOPs. The present calls were assessed by both Rosetta Resolver system and MAS5. In present call calculation in Resolver, we define the present call based on the criteria that the intensity is above the average background plus 3 standard deviations of the chip background or outside the background distribution, and that the Resolver intensity p-value is 0.01 or smaller. The results were shown in figure 2. In general, both present calls correlate well, our present calls using Resolver system tends to be more conservative than MAS5. On average, these FFPE samples have a present call around 43% and more than 90% samples have present call above 25% which is a general threshold or guidance for FFPE samples.

Table 2: Comparison of Present call in both Resolver and MAS5
(the breast cancer DSA has ~60K probe sets in total).

Statistics	No of Detected Probe sets (Resolver)	Resolver Present Call (%)	MAS5 Present call (%)
Mean	25634	42.1	45.1
Stdev	6931	11.4	13.3
Median	26149	43.0	46.7
Mode	26149	43.0	43.7
Max	36560	60.1	65.4
Min	8044	13.2	15.1

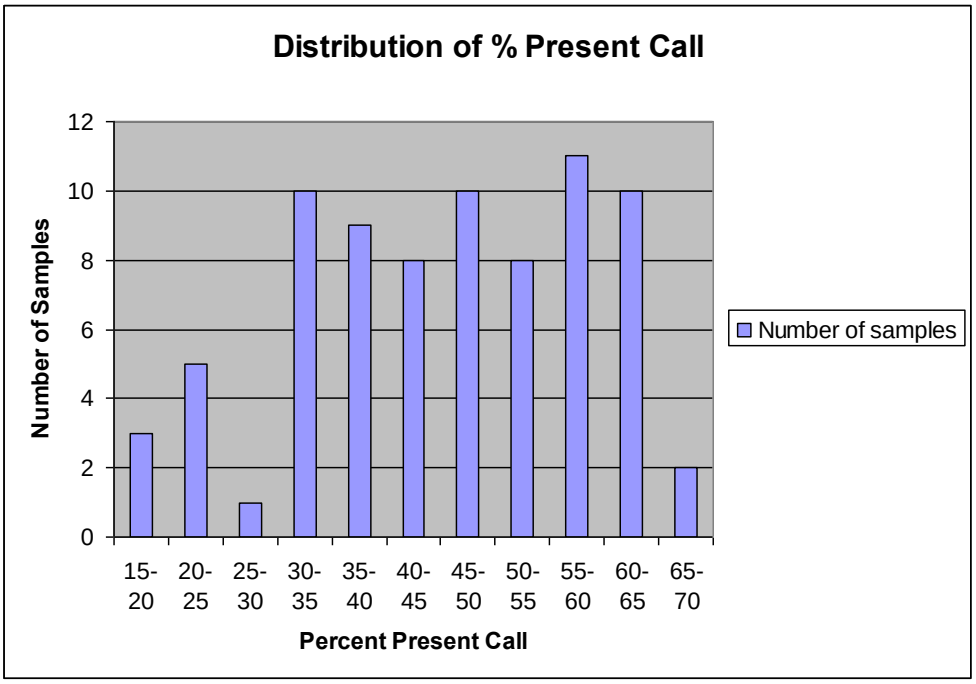

Figure 2A: Distribution of percent present call in MAS5 for FFPE samples

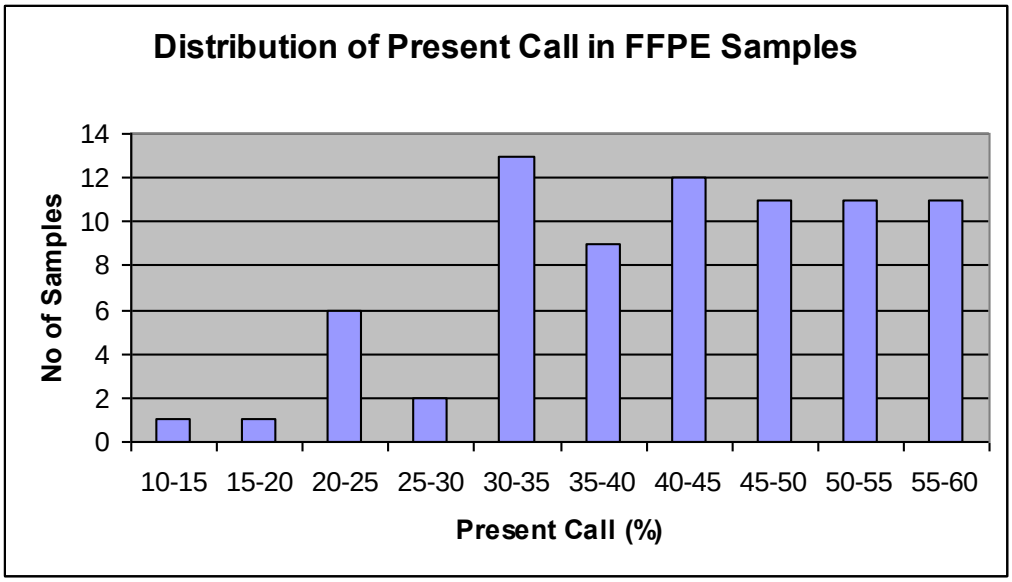

Figure 2B: Distribution of percent present call in Resolver for FFPE samples

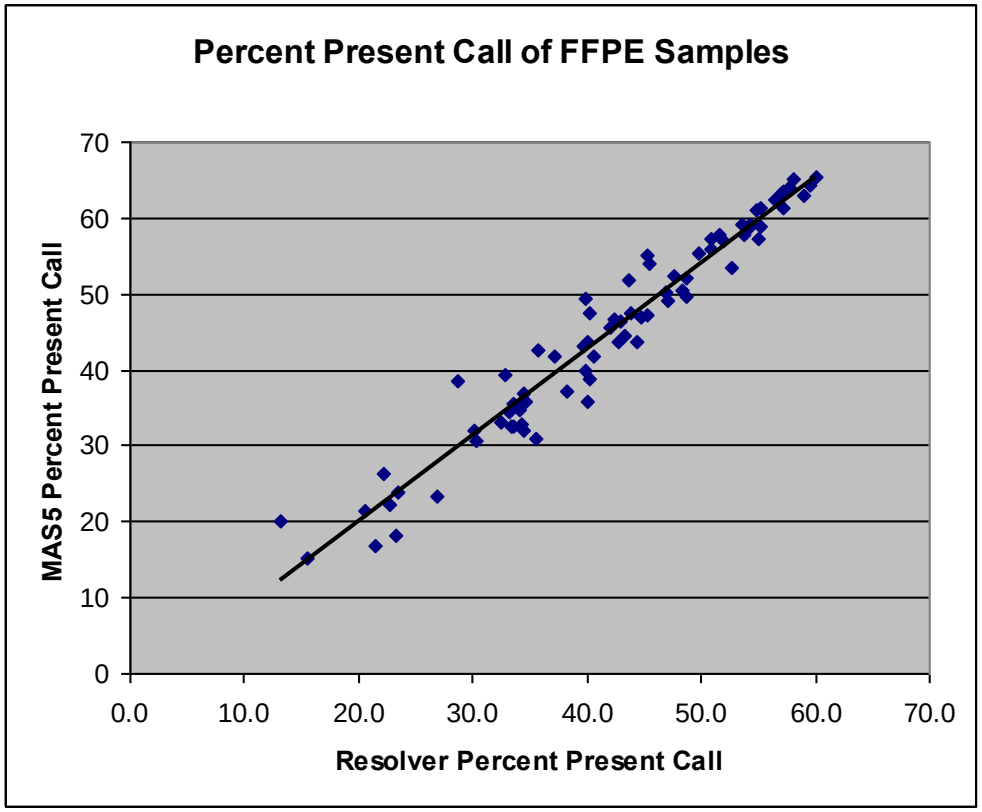

Figure 2C: Correlation of MAS5 present call and Resolver present call.

Gene Expression Sample integrity Assessment

Samples or chips were assessed using Principal component Analysis and Clustering Analysis in Rosetta Resolver Gene Expression Data Analysis System 7.1 to identify potential outliers or contamination or intratumor heterogeneity. As shown in Figure 3A (below), the PCA analysis identify 63b as an outlier, a tumor sample from native African triple negative cancer patients with low present call while normal cell sample 28 and 39 from patients could be contaminated by significant portion of tumor cells and tumor cell sample 41 may have a large portion of normal cells. These QC analyses demonstrated that the Breast Cancer DSA is able to separate tumor samples from normal in FFPE tissues and to identify samples of quality or integrity issues.

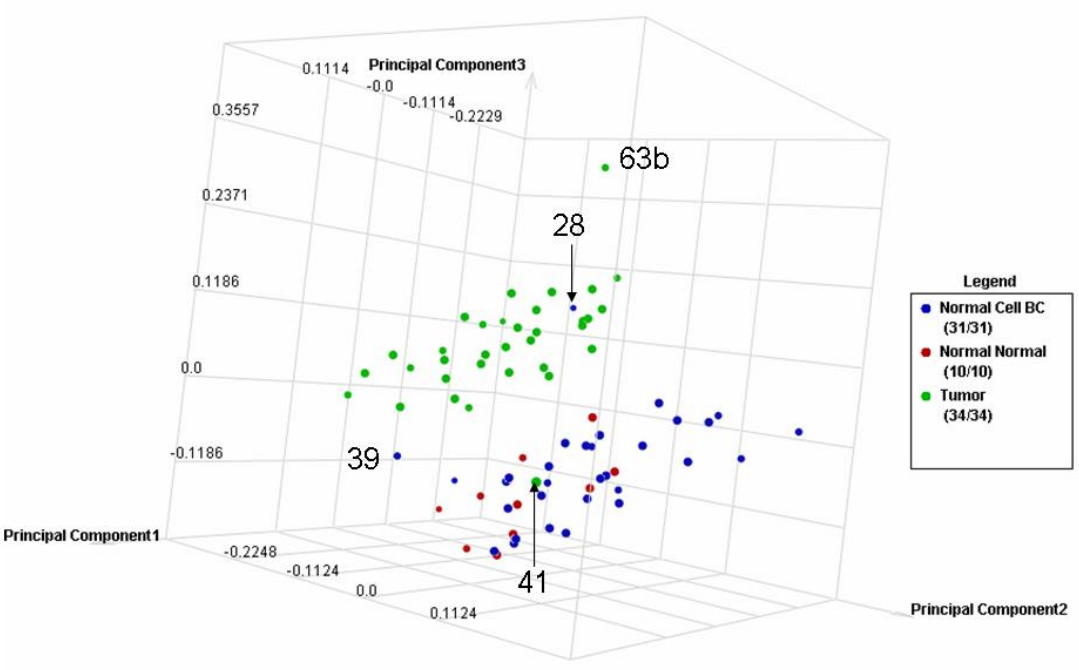

Figure 3A: PCA Analysis identified outlier or contamination.

After dropping these samples, PCA and Clustering analysis were performed again to confirm the sample are grouped in tumor and normal as expected (figure 3B and 3C). Although the "normal" cells from the cancer patient FFPE blocks and the normal biopsy tissue from health women fell into a larger group in PCA and the clustering, the normal cells from the cancer patients clearly have two distinguishable expression patterns after examining the K-mean clustering results (Figure 3C).

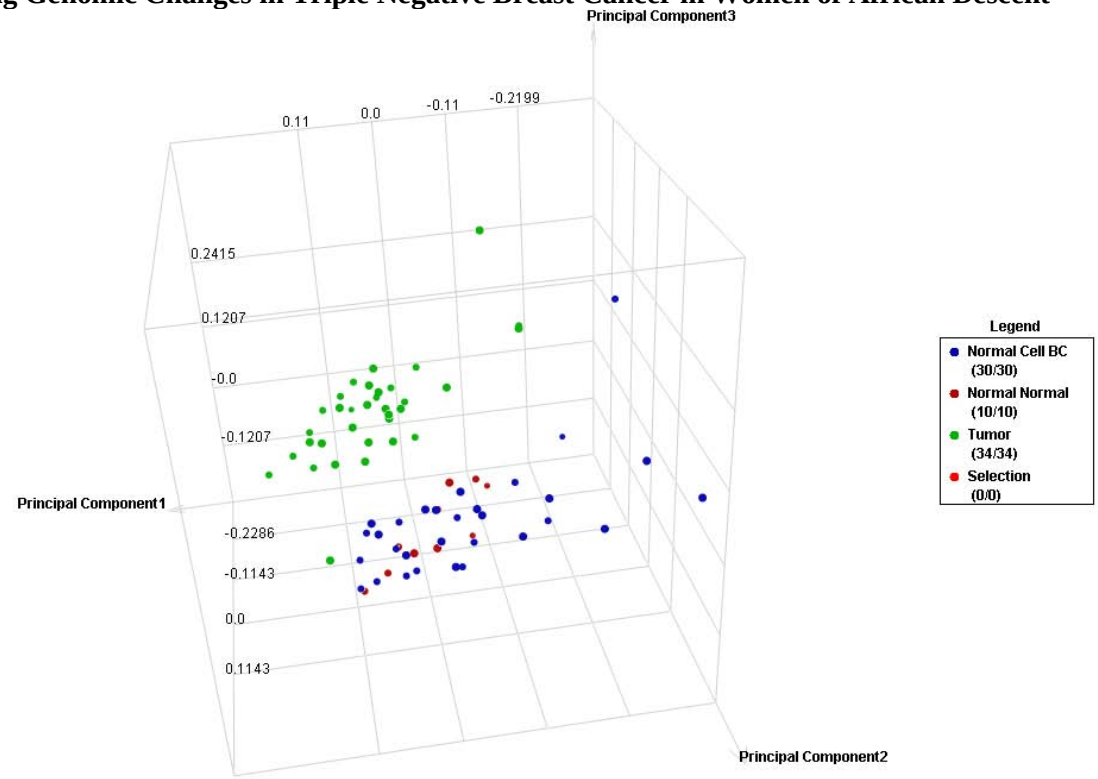

Figure 3B: Sample grouping by PCA after removing outlier and contaminated samples

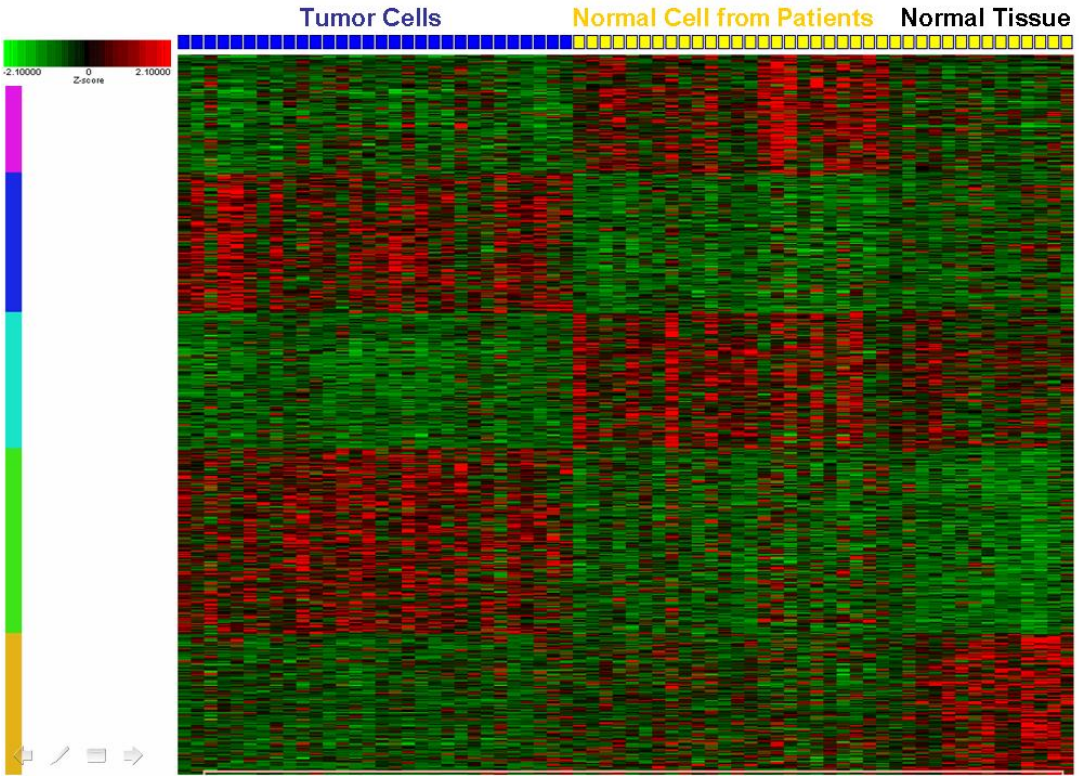

Figure 3C: Sample grouping and expression pattern in K-mean clustering after removing outlier and contaminated samples. Using probe sets from 2-way ANOVA, k-mean clustering was used to group samples. This heatmap shows that dissected tumor cells, "normal cells" adjacent to triple negative breast cancer and the normal breast biopsy from normal persons have distinct gene expression patterns.

To test if the expression profiling can identify potential sample contamination or mis-classification at microscopic level, we include one contaminated normal sample from an African American triple negative breast cancer patient in the two-dimentional K-mean clustering analysis. As figure 4 shown (below), this "normal" sample exihibited tumor-like gene expression profile and fell into the tumor cluster.

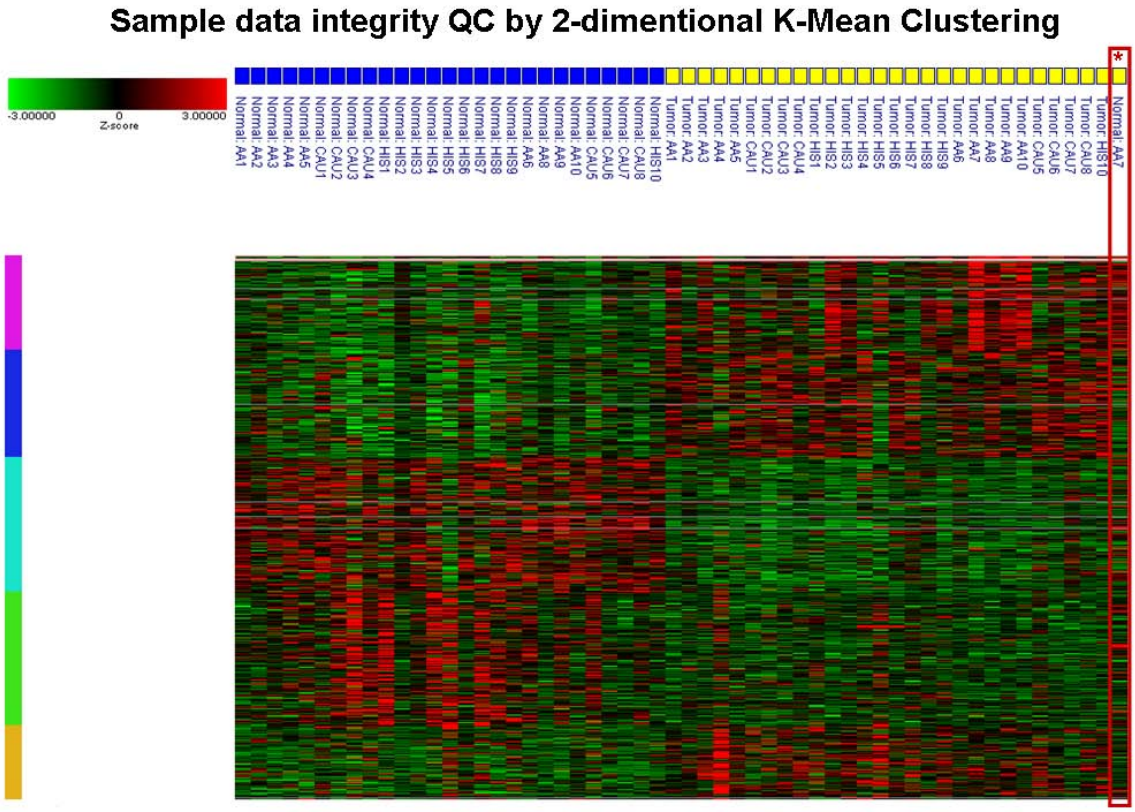

Figure 4. Identification of contaminated cells during dissection of FFPE slices. On the far right, one dissected FFPE slice of normal cells exhibits similar gene expression pattern to tumor cells.

Detection of Expressed Transcripts

The detection of expressed transcripts was defined by the probeset intensity that is beyond the background intensity by 3 standard deviations of the background. Venn Diagram analyses were performed between the tumor and normal cells for the African American, Caucasian American and Hispanic American group respectively to identify the unique and common transcripts detected. The results are summarized in the table 3 in numbers of transcripts. Apparently most of transcripts were detected in both the normal and tumor cells, but there are small fraction of transcripts from 841 to 4727 transcripts that were detected in only one side of samples.

Table 3: The overlap and unique transcripts detected in normal and tumor cells for each ethnic group.

Detection \ Group	AA	CAU	HIS
Detected in tumor only	3330	4164	4727
Detected in normal only	1859	1255	841
Detected in both	16061	14504	10182

Statistical Analysis and Ethnic Specific Expression Patterns

2-way ANOVA and paired t-test were used in gene selection process from those detected transcripts with multi-test correction with Benjamini-Hochberg false discovery. Also a fold change filter was applied to the tumor to normal comparison in each ethnic group to derive the differentially expressed transcripts that meet the criteria of the ANOVA p*-value of 0.01 or less, the t-test p*-value of 0.01 or less and fold change of 2 or up. In total, we found 1350 differentially expressed genes from the AA group, 1220 genes from the CAU group and 1226 genes from the HIS groups. These three transcripts were combined, resulting in a union set of 2662 genes. Two-dimensional hierarchical clustering analysis was performed to examine the gene expression patterns across samples groups at intensity level. The result was shown in figure 5 in heatmap. From the heatmap, there tumor and normal are clearly different in expression pattern. More interestingly, we observed certain subtle ethnic specific expression pattern across the three ethnic groups. After combining the transcripts as shown in ethnic specific visual pattern in the heatmap, we came up with 597 transcripts. These transcripts were subjected to pathway analysis in Metacore the results are highlighted in table 4. As the pathway list showed, some common tumorigenesis pathways revealed for DNA damage and repair, cell cycle control and cell death, cell adhesion etc.

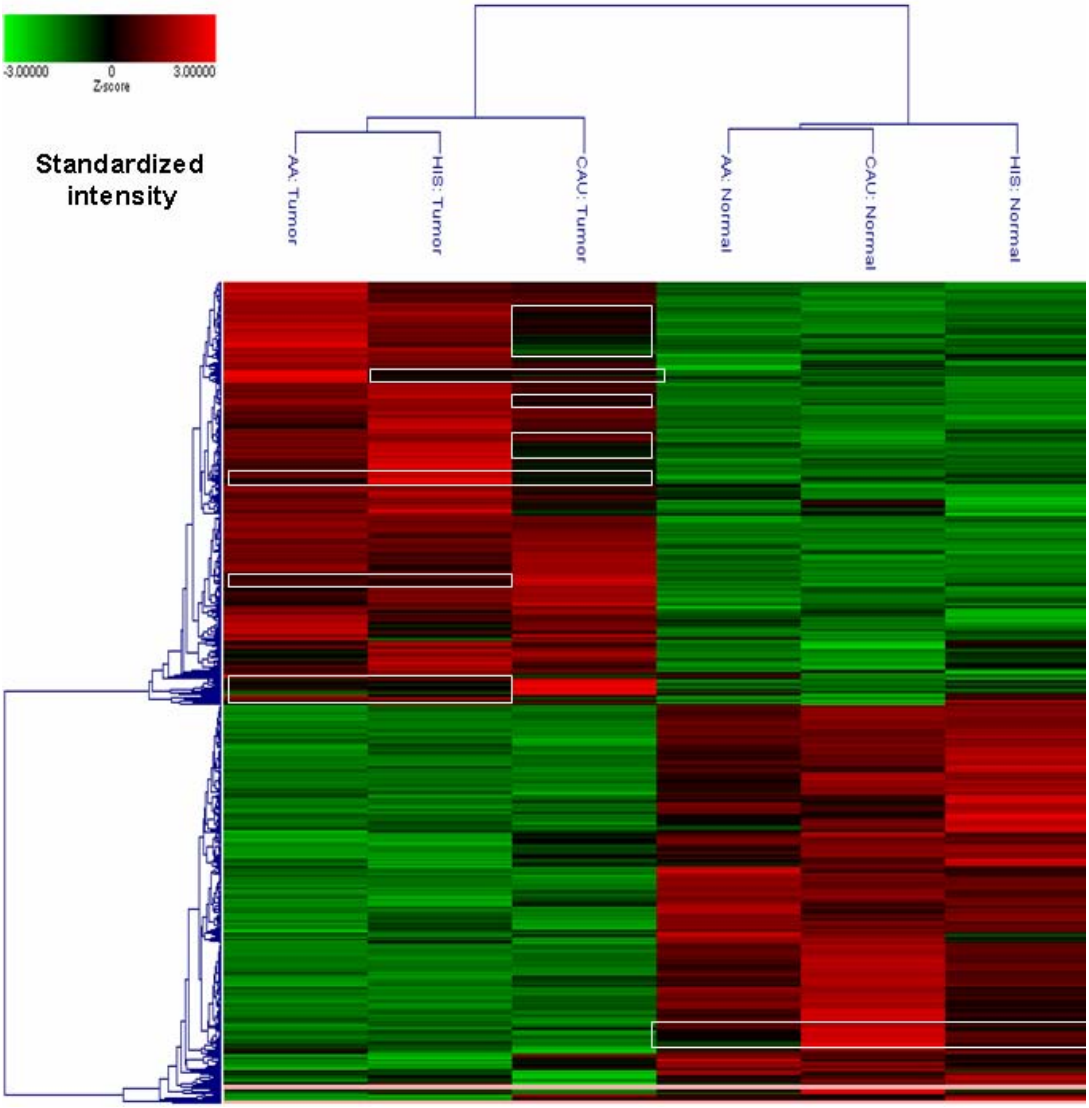

Figure 4: 2-D Clustering of 2662 transcripts across six sample groups.

The 2662 transcript set was applied to 2-D clustering analysis across the six sample groups. There are clearly two main dendrograms of transcripts, one with up-regulation (red) in tumor and down in normal (green), and the other with down-regulation (green) in tumor and up-regulation in normal tissues (red). In careful examination of the heatmap, 8 ethnic specific gene expression patterns were identified as shown in the labels by numbers in the heatmap. These 8 expression patterns demonstrated ethnic specific differentials.

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

Table 4: Top pathways from genes in the ethnic specific gene expression patterns

Pathway Name	Cell process	p-Value	# of genes in exp	Total # of genes
* Cell cycle_NHEJ mechanisms of DSBs repair	cell cycle	5.17E-05	7	53
* Cell cycle_The metaphase checkpoint	cell cycle	1.05E-04	9	101
* Cell cycle_Spindle assembly and chromosome separation	cell cycle	2.23E-04	8	88
* Cell cycle_Transition and termination of DNA replication	cell cycle	2.18E-03	6	72
* Cell cycle_Role of Nek in cell cycle regulation	protein kinase cascade, cell cycle	4.72E-03	6	84
* Cell cycle_Role of NFBD1 in DNA damage response	cell cycle	5.27E-03	4	38
* Cell cycle_Role APC in cell cycle regulation	cell cycle	6.25E-03	6	89
* Cell cycle_Role of Brca1 and Brca2 in DNA repair	cell cycle	6.60E-03	6	90
* Immune response _Oncostatin M signaling via MAPK in human cells	cytokine and chemokine mediated signaling pathway	2.87E-04	8	111
* Cell cycle_Role of Nek in cell cycle regulation	protein kinase cascade, cell cycle	1.77E-03	6	84
* Development_PDGF signaling via STATs and NF-kB	intracellular receptor-mediated signaling pathway, response to extracellular stimulus	2.11E-03	6	87
* Immune response _Oncostatin M signaling via JAK-Stat in human cells	cytokine and chemokine mediated signaling pathway	2.18E-03	5	60
* Development_Membrane-bound ESR1: interaction with growth factors signaling	transcription, response to hormone stimulus, response to extracellular stimulus	3.85E-03	6	98

5) KEY RESEARCH ACCOMPLISHMENTS

- Task 1: Determination of HER2 status in the Kijabe clinical cohort using fluorescence in situ hybridization.
- Task 2: Extraction and preparation of DNA and RNA from FFPE tumor samples from North American African, African African, and Caucasian cohorts.
- Task 3: Aim I, Analyze and compare genome-wide transcript expression in BC samples of AA ancestry vs. native African (Kijabe) samples.
- Task 4: Aim II, Investigate possible chromosomal alterations associated with gene expression differences using array cGH.

6) REPORTABLE OUTCOMES

- Baumbach, et al., Proceedings of the San Antonio Breast Cancer Symposium (2009).
- Baumbach, et al., Proceedings of the American Association for Cancer Research, Special Symposium on Breast Cancer, San Diego, CA (2009).
- Hurley, et al., Proceedings of the American Society of Clinical Oncology, Orlando, FL (2009).
- Baumbach, et al., Proceedings of the Miami Winter Symposium (2009).
- Baumbach, et al., Gene Expression Profiling of Formalin-fixed, Pariffin-embedded (FFPE) Tissues from Triple-negative Breast Cancer Patients (2010, manuscript in preparation).

7) CONCLUSION

RNA and DNA extracted from these samples are usually degraded, contaminated and of low quality in general. Despite the large banks of FFPE samples available for retrospective studies that include follow-up analysis of patient outcome, most of these studies currently focus on frozen samples because of the limited options available for paraffin samples. Additionally, FFPE processing holds advantages for tissue storage during prospective studies, in which many biopsies are collected but only a fraction of them are applied to downstream assays with selection based on clinical outcome. Because of the difficulty and time required to obtain fresh frozen tumor samples from the triple negative breast cancer patients with matched clinical criteria and curation, this study explored the possibility to profile both gene expression and genotype from FFPE tumor tissues. This study attempts to test and establish the feasibility and outline guidelines for selection of technology platforms and QC criteria for FFPE samples. FFPE RNA and DNA that are applied to the Almac Diagnostic Breast Cancer DSA arrays may still vary in quality and therefore require careful and rigorous QC to select samples that meet the quality standard including chip CQ and sample integrity check at profiling level. In SNP data, although the QC performance of FFPE sample are not comparable to fresh frozen sample, with careful QC and data analysis, valuable information such as LOH, and copy number assessment can still be obtained. The power comes when combine both the gene expression data with the genetic variation results; we could identify tumor suppressor genes that showed in both chromosomal aberration and transcriptional changes. These results outline guidelines for the application of FFPE samples to the same genome-wide platform already available to high-quality DNA samples, thus enabling widespread retrospective and prospective analysis of tumor samples in their most common form of storage.

8) REFERENCE

Baumbach, et al., Gene Expression Profiling of Formalin-fixed, Pariffin-embedded (FFPE) Tissues from Triple-negative Breast Cancer Patients (2010, manuscript in preparation).

9) APPENDICES – none