Abstract No. 1673

Identifying Distinct Multi-ethnic Genome-wide Alterations in Breast Cancer Using Paraffin Embedded Samples

MILLER SCHOOL OF MEDICINE UNIVERSITY OF MIAMI

L. Baumbach[1], M.E. Ahearn[1], C. Gomez[1], M. Jorda[1], O. Silva[1] T.A. Halsey[2], J. Yan[2], K. Ellison[2], S. Gluck[1]

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

ALMAC Diagnostics

Abstract

It is recognized that ethnic-specific disparities in stage of presentation/survival rates exist in breast cancer (BC) patients. **These disparities remain an enigma**. To investigate possible genetic contributions to these disparities, we are extending our previous study of genomic changes in BC samples from African-American (AA) women to a multi-ethnic cohort consisting of 20 each AA, His (His) and non-Hispanic white (Cauc) women matched for age of diagnosis, cancer stage, and hormone receptor status. A main study goal is to identify differentially expressed genes between tumor and normal tissue that are common or unique among the three ethnic groups. Tissue samples are evaluated for gene expression differences, as well as DNA copy number (CNV)/ chromosome alterations by high-density SNP arrays. We present here our most recent results of investigations focusing on triple negative (ER-/PR-/Her2-) patients. Pathology specimens were freshly cut from FFPE blocks and marked by a pathologist as to normal vs. tumor tissue. Almac Diagnostics performed RNA isolation, labeled cDNA preparation, and hybridization of tumor and normal cDNAs to a breast cancer focused gene expression microarray (*Breast Cancer DSA Research Tool.* Each patient had self-matched gene expression studies (tumor vs. normal). Using the FFPE samples, to date, approximately 17516 transcripts were expressed on the *Breast Cancer DSA* with intensity significantly higher than background. For the normal tissue samples, 9399 transcripts were detected in all three ethnic groups, while in tumor tissue samples, 10,296 transcripts were detected. There were also selected transcripts (hundreds to a thousand) that were detected in one or two ethnic groups only. Using two-way ANOVA (disease state and ethnicity) and a p-value cutoff of 0.01, a subset of 6479 highly consistent/significant genes were selected and further used in data quality control. Data QC indicated patient samples clustered well with respect to both ethnicity and normal versus tumor tissue. Additional methods of analysis included K-mean 2-Dimensional clustering and Principal Component Analysis. From these analyses of this limited sample set, we have already identified ethnic-specific expression patterns in tumor specimens, as well as in matched normal tissue samples. We are completing these studies mapping clusters of differentially-expressed genes into pathway analysis, validation by real-time PCR, and genome-wide CNV studies. We will present our latest findings. The overall aim of the completed study is an increased understanding of the biological basis of ethnic-specific BC disparities, leading ultimately to individualized, ethnic-specific diagnostic and therapeutic approaches.

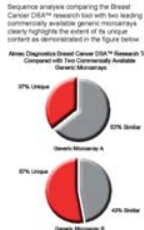
Almac Diagnostics Cancer DSA™ research tools

Standard microarrays provide large quantities of information, but are broadly representative of the human genome rather than a particular disease state.

The Breast Cancer DSA™ was developed by a process of high throughput sequencing, gene expression profiling and bioinformatics analysis, to fully characterize the transcriptome of disease and normal tissue. Therefore, it includes significant additional relevant data not available from other microarrays.

The Breast Cancer DSA™ research tool contains approximately 60,000 transcripts and is manufactured on Affymetrix GeneChip® technology.

Technical Specifications of the Breast Cancer DSA™ Research Tool	
Number of transcripts	~ 60,000
Feature Size	11 micron
Probe Length	25-mer
Probe Pairs/Probeset	11
Normalisation Ctrl Probesets	100
Hybridisation/Housekeeping Ctrl Probesets	62



Comparison of Breast Cancer DSA™ to the RefSeq database showed that 51% of the Breast Cancer DSA™ content is not present in the RefSeq database and 8% of the content represents antisense sequences to annotated transcripts.

Breast Cancer DSA™ and detection of differentially expressed transcripts

Technical assessment: Use with FFPE

The number of transcripts that were called present and were above the background in both RNAlater and FFPE samples was determined as shown in the table below.

	Transcripts detected in RNAlater	Transcripts detected in FFPE	Number of FFPE detected transcripts also detected in RNAlater
Breast Cancer DSA™ Research Tool	28,716	18,830	18,032
Affymetrix HG-U133 Plus 2	25,473	11,709	11,441

The high degree of data retention clearly demonstrates the power of the Breast Cancer DSA™ research tool when used in FFPE studies.

Use of the Breast Cancer DSA™ research tool in detection of differentially expressed transcripts

In most experimental studies, the end point of the analysis is the detection of differentially expressed transcripts. Analysis was carried out to determine the number of differentially expressed transcripts between normal and tumour tissue in our example study, defined as transcripts called present and above background with a fold change greater than 2 standard deviations of the mean with p-value <0.05. As can clearly be seen a large number of differentially expressed transcripts are detected in this experiment using both RNAlater and FFPE extracted samples. It is important to note that this experimental design is based on matched tumour and normal tissue from a single patient.

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 20 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**
- **Triple Negative hormone receptor status**

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

Ethnicity	Normal	Tumor
African-American women (AA)	10 samples	10 samples
Caucasian women (Cau)	8 samples	8 samples
Hispanic women (His)	10 samples	10 samples

Methods

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-.

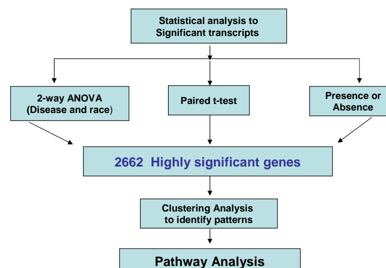
For each patient, pathologists cut new sections from the paraffin-embedded sample blocks for both normal and tumor tissue. Samples were sent to Almac Diagnostics for processing and hybridization to the Breast Cancer DSA™ Research Tool.

The following steps were performed by Almac Diagnostics:

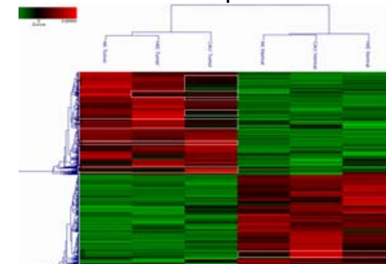
1. Isolation of RNA from each sample
2. RNA Amplification using the NuGEN FFPE System
3. Generation of First Strand cDNA from 10-80 ng of total RNA
4. Generation of a DNA/RNA Heteroduplex Double Strand cDNA and amplification
5. cDNA Fragmentation and Labelling using the NuGEN System
6. Affymetrix Hybridisation Washing, Staining and Scanning Protocol applied to the Breast Cancer DSA™ Research Tool.

ACKNOWLEDGEMENTS: This project has been generously supported by a grant from the Susan G. Komen Breast Cancer Foundation, Grant No. POP0601150 and the Women's Cancer Association, University of Miami Miller School of Medicine. We also thank the clinicians who contributed to this project.

Data Analysis Strategy



Ethnic Patterns of Gene Expression in Breast Cancer



Dendrogram from the cluster analysis of 28 patient samples from three ethnicities.

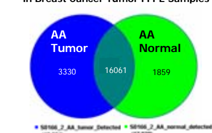
- ➢Differentially expressed genes between tumor and normal tissue were seen in each group. 1350 in AA, 1220 genes in CAU and 1226 genes in HIS. Combination of the groups resulted in a union set of 2662 differentially expressed genes.
- ➢The 2662 transcript set was applied to 2-D clustering analysis across the six sample groups.
- ➢There are two main dendrograms of transcripts: one with up-regulation (red) in tumor and down in normal (green) seen in the top half of the heat map, and the second with down-regulation (green) in tumor and up-regulation in normal tissues (red) seen in the bottom half of the heat map.
- ➢8 ethnic specific gene expression patterns were identified as shown by the white boxes in the heatmap. These 8 expression patterns demonstrated ethnic specific differentials.

Detection of Expressed Genes in Normal and Cancer FFPE Breast Samples by BC DSA

Number of Transcripts Detected by Ethnic Group

	AA	CAU	HIS
Detected in Tumor	3330	4164	4727
Detected in Normal	1859	1255	841
Detected in Both	16061	14504	10182

Numbers of Expressed Genes Detected in Breast Cancer Tumor FFPE Samples



21250 transcripts are expressed on the Breast Cancer DSA with intensity significantly above background.
Data from self-matched tumor and normal breast tissue was compared for the three ethnic groups
There are some genes (several hundreds to over a thousand) are detected in one or two ethnic groups.

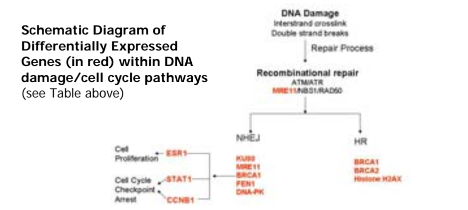
Biological Pathway Analysis

Differential Expression of DNA Repair Pathway Genes by Ethnicity.

Gene Symbol	Protein	African-American Fold Change	Caucasian Fold Change	Hispanic Fold Change
BRCA1	Breast cancer type 1 susceptibility protein	1.25	2.02	1.11
BRCA2	Breast cancer type 2 susceptibility protein	2.46	5.53	2.69
FEN1	Flap endonuclease 1	2.75	1.33	4.18
MRE11A	Double-strand break repair MRE11A	2.44	1.31	1.94
PRKDC	DNA-dependent protein kinase catalytic subunit	1.78	1.72	2.19
XRCC5/Ku80	ATP-dependent DNA helicase 2 subunit 2	2.12	1.00	2.75
H2AFX	Histone H2A.x	2.04	1.30	3.06
CCNB1	G2/mitotic-specific cyclin-B1	2.91	2.12	2.97
ESR1	Estrogen receptor	-5.61	-2.29	-6.54
STAT1	Signal transducer activator of transcription 1-alpha/beta	1.51	3.13	2.58

Transcripts from the ethnic specific expression patterns were analyzed with MetaCore
Note: Red numerals indicate increase in gene expression in tumor vs. normal tissue. Blue numerals indicate decrease in gene expression in tumor vs. normal.

Schematic Diagram of Differentially Expressed Genes (in red) within DNA damage/cell cycle pathways (see Table above)



Current Status of Project and Concluding Remarks

- ➢**We are selecting approximately 10 differentially expressed genes for validation of gene expression analysis by qRT-PCR.**
- ➢**We are also extracting DNA from a subset of these specimens for hybridization to high-density SNP arrays, to assay possible DNA copy number variations (CNV's) and/or LOH in tumor samples.**
- ➢**Initial pathway analysis using MetaCore shows that a number of genes related to the DNA repair pathway are differentially expressed in the samples. Further analysis of this data is ongoing.**

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