AD_____

Award Number:
W81XWH-08-1-0437

TITLE: "Determination of Metastatic Potential in Breast Tumors by Global Molecular Characterization Using Multiple Modalities"

PRINCIPAL INVESTIGATOR: Richard J. Mural, PhD

CONTRACTING ORGANIZATION:
Windber Research Institute
Windber, PA 15963-1331

REPORT DATE: October 2010

TYPE OF REPORT: Final

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)*
* 29-OCT-2010	Final	30 JUN 2008 - 29 SEP 2010

4. TITLE AND SUBTITLE
Determination of Metastic Potential in Breast Tumors
By Global Molecular Characterization Using Multiple Modalities

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-08-1-0437

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Richard J. Mural, PhD

Email: r.mural@wriwindber.org

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Windber Research Institute
Windber, PA 15963

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
US Army Medical Research and Materiel Command
Fort Detrick, MD 21702-5014

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
Breast cancer is the most frequently diagnosed non-skin cancer in women and the second leading cause of cancer related deaths in women today. Early detection has been instrumental in improving outcomes for women with breast cancer. Primary tumors are rarely, if ever, the cause of cancer mortality, rather cancer deaths are the consequence of metastasis. The ability of a tumor to metastasize is a fundamental property which determines the severity of the disease. Better ways of predicting the likelihood of metastasis and finding markers that identify tumors that are likely to metastasize is of critical importance to the optimal management of breast cancer patients. This pilot project is designed to obtain comprehensive data on gene expression, epigenetic changes and genomic abnormalities from breast tumor samples in the Windber Research Institute tissue repository, collected as part of the Clinical Breast Care Project (CBCP), from patients stratified by lymph node status (diseased or free of disease). Two hundred twenty four patients enrolled in the CBCP have been found to meet the criteria for inclusion in this project. One hundred of these samples have been selected to begin this study. The HRPO determined that the proposal constitutes research not involving human subjects. Sample selection has been completed and samples are undergoing laser capture microdissection. DNA and RNA will be extracted from these samples for further analysis. Data are being generated and analyzed.

15. SUBJECT TERMS
Breast Cancer, metastasis, gene expression, epigenetic changes, genomic abnormalities, laser capture microdissection

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Richard J. Mural, PhD
U	U	U	UU	33	19b. TELEPHONE NUMBER *(include area code)*

Final Report for W81XWH-08-1-0437: "Determination of metastic potential in breast tumors by global molecular characterization using multiple modalities"

Table of Contents

INTRODUCTION:

Breast cancer is by far the most commonly diagnosed invasive cancer among women and is the second leading cause of cancer-related deaths in United States. In 2009, it is estimated that 192,370 women were diagnosed with breast cancer and 40,170 women died of the disease (1). Although the primary tumor often causes significant morbidity, metastasis to distant organs accounts for >90% of breast cancer–related mortality. The molecular mechanisms underlying the development and progression of breast cancer are far from being understood. Early detection, determining whether a tumor has metastasized, or is likely to metastasize is critical to disease prognosis and in determining the course of treatment for the patient (2) .

Breast cancer is a highly heterogeneous and complex disease with many risk factors ranging from genetic predisposition to life style factors. In addition, breast tissue is a difficult tissue to study as it is composed of several cell types and undergoes structural changes during the menstrual cycle, pregnancy and aging process. High-throughput genomic technology, developed in the last decade and enabling the simultaneous measurement of variation in thousands of DNA sequences, mRNA transcripts and peptides, has rapidly become a major tool for the study of breast cancer.

Microarray based gene expression profiling has been performed in multiple ways by various researchers to develop gene sets or gene signatures to improve the diagnosis and risk stratification for breast cancer. Perou et al (3;4) using unsupervised hierarchical clustering with an 'intrinsic gene list' identified 5 molecular groups based on their gene expression patterns. Gene expression based class prediction analyses have also been carried out to develop gene signatures to accurately predict the class membership of a new breast cancer sample on the basis of the expression levels of key genes. Using this supervised approach, several microarray based prognostic gene signatures have been developed (5-11). The first prognostic signature described was the 70-gene signature (Mammaprint) by the Amsterdam group to identify patients who develop metastasis within 5 years and those who do not (10). This gene signature is being prospectively tested in the MINDACT trial.

Specific DNA copy number alterations (CNA) are major genomic alterations that contribute to carcinogenesis and tumor progression (12-14). However to identify the specific causative gene CNAs responsible for gene expression regulation, which ultimately lead to malignant transformation and progression is a challenge. Introduction of high density single nucleotide polymorphism (SNP) genotyping arrays has helped not only for whole genome profiling but also for detecting copy number changes based on the measured intensities of both alleles of a SNP. Employing the SNP array technology limited studies have investigated the CNAs in relation to their prognostic significance (15-17).

Epigenetic malfunction also has been shown in the last decade to play a significant role comparable to genetics in cancer development. DNA methylation is an epigenetic event crucial in regulating the gene expression and genomic stability (18;19). Aberrant methylation of CpG island within a promoter causing silencing of tumors suppressor genes is a wide spread phenomenon in cancer cells (20) . Approximately 60% of protein-coding mammalian genes harbor CpG islands in their promoter region. Understanding aberrant DNA methylation patterns will help in developing markers for classification,, detection and risk assessment of breast cancer. Using microarray technology attempts have been made, mostly using cell lines, to identify genome-wide epigenetic alterations and to understand the functional consequences of these changes in the context of breast cancer metastasis (21;22).

As described above, availability of high resolution platforms for profiling of genetic, epigenetic and gene expression changes allowed us to develop the current pilot project wherein we proposed to obtain comprehensive data on gene expression, epigenetic changes and genomic abnormalities from breast tumor samples obtained from breast cancer patients stratified by lymph node status (diseased or free of disease). Whole genome profiling using multiple platforms enables us to begin to understand the process of metastasis in breast cancer from a systems biology prospective. We used Affymetrix gene expression (U133 plus), promoter (1.0R) and SNP/CNV (SNP 6.0) microarray platforms to obtain high-resolution whole-genome gene expression, epigenetic and copy number/SNP data from lymph node positive and lymph node negative, post and premenopausal breast cancer patients. To date such global cancer related changes using multiple platforms has been reported only in osteosarcoma (23) and very recently in breast cancer cell lines (24).

One of the unique aspects of our project is the samples used for the study. The breast tumor samples used for this study are from the Windber Research Institute (WRI) tissue repository, collected as a part of the Clinical Breast Care Project (CBCP). CBCP is congressionally mandated and funded military-civilian collaboration between Walter Reed Army Medical Center (WRAMC) in Washington, D.C, Windber Medical Center in Windber, PA, and the Henry Jackson foundation, MD. It is an ongoing project with subjects being recruited in breast clinics at WRAMC, Anne Arundel Medical Center in Annapolis, Maryland, and Windber Medical Center in Windber, PA. HIPPA complaint IRB approved protocols are used to gather samples of breast and metastatic tissues and blood for future use in research studies from fully consented patients. All the sites follow efficient recruiting practices, high quality tissue collection and preservation protocols, and use the same questionnaires for data collection. Finally all the data is deposited and maintained in a central, querriable data warehouse. The extensive amount of clinical and epidemiological information collected with high quality research grade tissue specimens allowed us to carefully stratify the samples we included in this study.

BODY:

STUDY SUBJECTS AND STUDY DESIGN:

The subjects for this study were CBCP participants enrolled from the Walter Reed Army Medical Center. This project employed a 2x2 factorial design with 20 patients in each group, for a total of 80 patients. The two factors used were (1) pre- vs. post-menopausal onset of disease, and (2) lymph node status, tumor found in lymph nodes (positive) vs. nodes free of tumor (negative). In addition, lymph node positive and negative patients within each group, pre- vs. post-menopausal, were matched as closely as possible for age and BMI (Body Mass Index). All of the tumors were estrogen receptor positive (ER+). Because of availability, the subjects for this study are all Caucasian.

Though the original proposal envisioned the use of 25 patients for each group, however due to availability of the tissues from patients meeting the inclusion criteria and difficulties in amplifying the RNA from some breast tissues samples especially in the premenopausal groups, upon completion of the project, the number of subjects in each group were as follows.

Postmenopausal node positive=20

Postmenopausal node positive=20

Premenopausal node negative=18

Premenopausal node positive=19

RNA and DNA isolated from the breast tissues of these subjects were used for multiple analytical modalities including, gene expression, SNP and CNV analysis, and whole genome DNA methylation profiling. Gene expression data was successfully generated for all subjects of the above groups. However, due to time constraints and funding issues, SNP data and whole genome DNA methylation profiling was generated for subset of subjects. The details of the number of subjects for which the gene expression data, SNP data and methylation data is available is given in Appendix 1.

METHODOLOGY:

Tissue Samples: Invasive breast cancer (ER+) specimens were obtained from the subjects enrolled in the Clinical Breast Care Project (CBCP) at Walter Reed Army Medical Center (WRAMC). All tissue samples were collected with approval from the WRAMC Human Use Committee and Institutional Review Board. All subjects enrolled in the CBCP voluntarily agreed to participate and gave written informed consent.

For these invasive tissues, a single tumor sample is assayed. Laser capture microdissection was used to separate tumor tissue from surrounding stroma. Sections were cut for both RNA and DNA isolation at the same time to minimize the variations in the tumor cell population being used for various modalities.

Gene Expression:

RNA Isolation: RNA was isolated from frozen breast specimen sections using RNeasy Micro Kit (Qiagen Inc.) protocol. The concentration of each sample was determined with the NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA) while the RNA integrity was monitored with the Bioanalyzer (Agilent Technologies, Inc., Santa Clara, CA).

RNA Amplification. 10 ng of total RNA from each sample was subjected to two rounds of amplification using the MessageAmp II aRNA Amplification kit (Applied Biosystems Inc., Foster City, CA). GeneChip Eukaryotic Poly-A Controls (Affymetrix, Inc., Santa Clara, CA) were spiked into the total RNA samples prior to serve as a quality control measure for the amplification process in the first round of amplification. The first and second strand cDNA synthesis and cDNA purification as well as the in vitro transcription (IVT) reaction and aRNA purification for both rounds of amplification were performed according to the amplification kit instruction manual. The aRNA was labeled during the IVT reaction of the second round of amplification with 10 mM biotin-11-UTP (PerkinElmer, Inc., Waltham, MA). The concentration and size distribution of the biotin-labeled aRNA was determined on the NanoDrop ND-1000 and Bioanalyzer, respectively.

Fragmentation, Hybridization, Staining, and Scanning. Biotin-labeled aRNA (20 μg) was fragmented by incubation with 5X Fragmentation buffer (Affymetrix, Inc.) for 35 min. at 94° C. 15 μg of each fragmented sample was prepared for hybridization as specified in the Affymetrix GeneChip Expression Analysis Technical Manual for 49/64 Format (standard) expression arrays and then hybridized to Affymetrix HG U133A plus 2.0 arrays for 16 hours. The arrays were washed, stained, and scanned according to Affymetrix protocols.

Alterations in Genome Structure (SNPs):

DNA isolation: DNA was isolated from frozen breast specimen sections and also from the blood cells/clots of the same subjects. The genomic DNA of the blood samples served as control for determining the genomic abnormalities in the tumors. The DNA from the cells/clots of the subjects was isolated using the Puregene Blood Kit (Qiagen Inc.) protocol. The DNA was isolated from the breast tissue sections using QIAamp DNA Mini Kit (Qiagen Inc.) protocol. Obtaining good quality genomic DNA suitable for running on the SNP 6.0 platform required standardization. Originally, the protocol in our lab was to let the tissue sections of the samples for DNA isolation remain for variable time periods and after a batch of samples are cut, then the DNA was isolated from the sections of these samples. Initially some of the samples for this project were processed similarly. But when the quality of the isolated DNA from these samples was tested on 1% agarose gel, the DNA was found to be degraded and unsuitable to run on SNP 6.0 platform. The protocol was later modified so that the DNA was isolated on the same day the tumor sections were cut. A representative 1% agarose gel run of the samples showing higher quality DNA with a prominent band >10 kb for both clots and tissue samples is shown in Appendix 2.

The concentration of the DNA of each sample was determined with the NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA). DNA thus isolated was used for both SNP/CNV studies as well as for whole genome methylation studies.

SNP analysis: DNA was genotyped using Affymetrix 6.0 microarrays according to manufacturer's instructions. About 500 ng of DNA was cleaved with either NspI (250 ng) or StyI (250 ng) restriction enzymes, ligated to specific linker and amplified with the Clontech Titanium Taq DNA polymerase. The quality of the PCR products is tested on 2% agarose gel, the PCR products appear as a smear between 200-1100 bp. The PCR products of the NspI reations (4 PCR reactions per sample) and the StyI reactions (3 reactions per sample) were combined and purified. The purified DNA concentration is determined with the NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA). DNA samples with concentrations above 3.5 µg/ml and a 260/280 ratio between 1.8 and 2.0 were fragmented (average fragment size <180 bp) and labeled with GeneChip DNA labeling reagent (Affymetrix). The samples were then hybridized to the Affymetrix Genechip Human 6.0 arrays. Following hybridization, the arrays were stained and scanned on a Genchip scanner 3000. Genotypes for each samples passed Affymetrix quality control metrics with contrast quality control threshold > 0.4.

Whole Genome Methylation:

Whole genome methylation protocol was standardized for use in our laboratory for the first time.

DNA isolation: An aliquot of DNA isolated as described above for SNP/CNVs was used for whole genome methylation studies.

Sonication: About 2.4µg of DNA was sonicated, 2µg of DNA for immunoprecipitation and 400ng for agarose gel run. Sonication was performed at 30% amplitude for 48 pulses (20 sec on and 20 sec off) in 10 mM Tris-HCL buffer, pH 8.5. The DNA was sheared to a range between 100-300 bp to minimize the number of CpG islands per fragment.

Immunoprecipitation: Enrichment of methylated DNA was done by immunoprecipitation using the MIRA assay method (MethylCollector Ultra, Active Motif, CA). The method involves the use of the

recombinant MBD2b/MBD3L1 protein complex which has a high affinity for the CpG-methylated DNA to immunoprecipitate the methylated DNA in the samples. The protein-DNA complexes are then captured using the nickel coated magnetic beads, washed and the methylated DNA is eluted. Patient DNA samples (1 microgram each) are immunoprecipitated in duplicate to have enough methylated DNA for downstream events. The eluted methylated DNA is cleaned up using the minielute purification kit (Qiagen) to remove traces of any degraded proteins and nucleotides.

Validation of Immunoprecipitation: To validate the recovery of methylated DNA by immunoprecipitation, control samples (200 ng of Human male genomic DNA) were used. The isolated CpG-methylated DNA fragments of the control sample after clean up step were amplified by PCR using the control PCR primers provided in the kit (Appendix 3), APC (the region amplified by this primer pair should be 338 base pairs and has about 29 CpGs); Xist (The region amplified by this primer pair is 178 base pairs and contains 8 CpGs) and NBR2 (The region amplified by this primer pair is 103 bp and contains 7CpGs). The PCR products were analyzed on 3% agarose gel run at 125v for 50 min. The APC is an unmethylated promoter and thus will not be amplified in the eluted fraction [Lane 8 of Appendix 3 but is seen in the unbound fraction (lane 9) and input (lane 10)]. Xist is a methylated promoter and thus is amplified in the eluted fraction (lane 5). Lane 6 and 7 are the unbound (faint band) and the input samples. NBR2 is also a methylated promoter and is enriched in the eluted fraction (lane 2) while the unbound fraction does not have any band (lane 3). Lane 4 is the input sample.

Amplification of the methylated DNA: As the quantity of the methylated DNA is not sufficient for the fragmentation, labeling and hybridization to the Affymetrix human promoter arrays, it is amplified using the GenomePlex Complete Whole Genome Amplification Kit (WGA2; Sigma) protocol. The amplified DNA is subjected to a clean up step using the QIAquick PCR purification kit (Qiagen) to remove smaller fragments of DNA and residual primers in the samples (The results of the amplification for 2 test samples are shown in Appendix 4A).

Fragmentation, labeling, hybridization to the human promoter 1.0 R arrays, washing, staining and scanning: The amplified methylated DNA is fragmented with the addition of UDG and APE1 (Appendix 4B) and labeled with biotin-labeled TdT as described in the Affymetrix protocol. The labeled DNA is hybridized to the GeneChip Promoter 1.0R array for 16 hrs at 45^0C as described in the Affymetrix protocol. The arrays were washed, stained, and scanned according to Affymetrix protocols.
Analysis: The sample data was analyzed using the Affymetrix Tiling Analysis software. To the date of writing our final report we analyzed only test samples and the analysis report is shown in Appendix 5.
Microarray data analysis

Data pre-processing and quality assessment

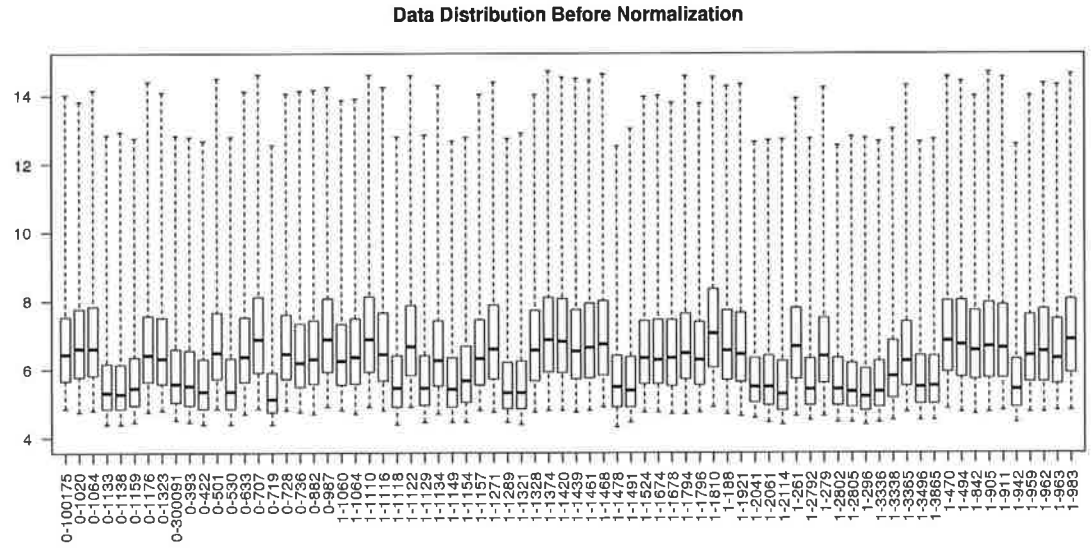
The microarray raw data (.CEL files) contain probe pixel intensities which were generated in the final scanning process of Affymetrix microarray experiment. These files are the starting data for pre-processing and quality assessment and quality control (QA/QC).

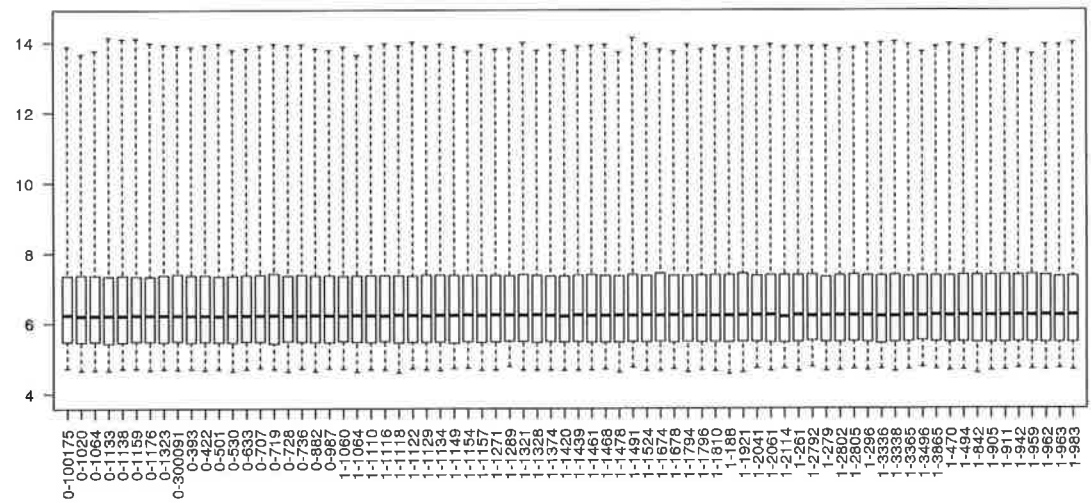
QA/QC is a crucial first step in successful data analysis. Before any comparison can be preformed, it is necessary to check that there are no problems with sample processing, and that arrays are of sufficient quality to be included in a study

These raw image data were processed using R programming and Bioconductor packages for quality assessment (QA) and calculation of the gene expression matrix.

The quality assessment was performed based on several methods which are widely used for microarray data quality control: Affymetrix recommended QC metrics (average background, scale factors, percentage of present, 3'/5' ratios for Gapdh and Actin, etc), RNA degradation, Normalized Unscaled Standard Error (NUSE), Relative Log Expression (RLE), residual images after fitting probe level model (PLM), MA plot, Spatial distribution of feature intensities, standard deviation versus rank of the mean, box plot and density plot for data distribution, clustering heat map of distance between arrays, Perfect match (PM) and mismatch (MM) probe expression pattern, etc.

The outlier chips with possible quality problems were repeated or removed from the following analysis. After the QA/QC process, 74 samples (microarrays) are included in the following data analysis process. The Gene Expression Matrix was obtained using the Robust Multi-chip Average (RMA) method. Specifically, background correction (RMA convolution model), quantile normalization across the array, and summarization of probe intensities in each probe set (PLM model) were carried out to get the final gene expression values. (See the following box plot for data distribution before and after normalization).





RESULTS:

CLINICAL CHARACTERISTICS

The clinical and pathological characteristics of the study subjects are given in Appendices 6 and 7. In the postmenopausal node negative (PMNN) women, the age at diagnosis was higher compared to node positive women (PMNP) (68 vs 61 yrs). The age at diagnosis in premenopausal groups was not different (Appendix 6). The body weight and BMI in PMNN women were also marginally lower compared to PMNP women, while in the premenopausal groups no such differences were noted. It is interesting to note that significant number of women belonging all groups had history of breast cancer in a primary relative and secondary relative, though they were not different between the groups (Appendix 6). Also all the groups had high percentage of women with history of other cancers. Another interesting observation was that node positive women compared to node negative women in both post (1357 vs 854) and premenopausal (1008 vs 739) groups had higher caffeine scores (Appendix 6).

In all groups of women, the rate of cancer was higher in the right breast than in the left breast (Appendix 7). The node negative women compared to node positive women of both pre and postmenopausal groups had higher percentage of women with well differentiated tumors (Post- 40% vs 20%; Pre - 40% vs 21%) and lower percentage of women with poorly differentiated tumors (Post -20% vs 35%; Pre- 11% vs 42%) (Appendix 7). The tumor sizes were marginally higher in node positive groups compared to node negative groups (Appendix 7). As expected, Stage I and Stage 2A cancers were the highest percentage of cancers in node negative post and premenopausal groups. On the other hand the node positive groups in both pre and postmenopausal women had higher AJCC pathological stages of breast cancer (Appendix 7).

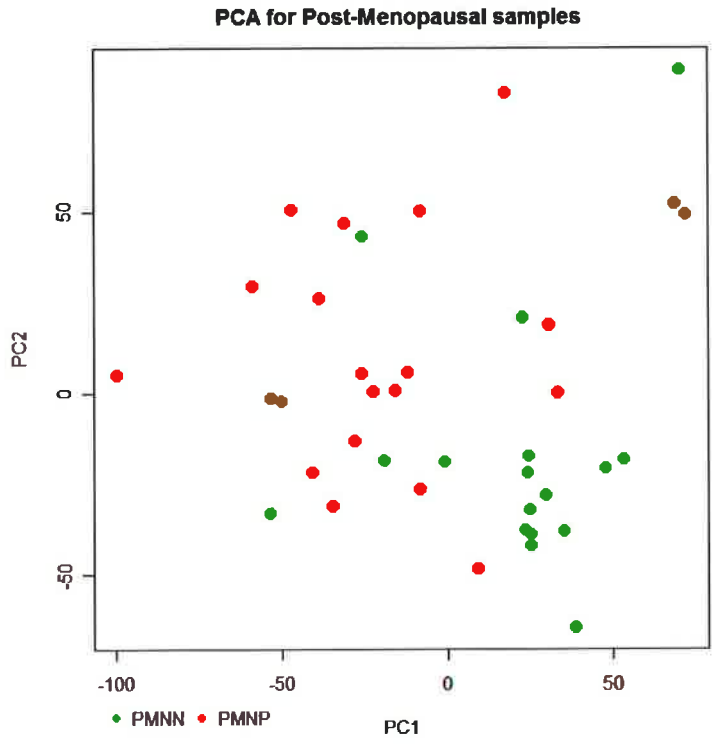
GENE EXPRESSION ANALYSIS

There are a number of possible analyses that can be performed on these data. Our initial analysis was carried out on the postmenopausal sub group because there was strong differentiation of the node positive (+) and node negative (–) tumors based on the genes that are differentially expressed within this group.

Exploratory multivariate analysis

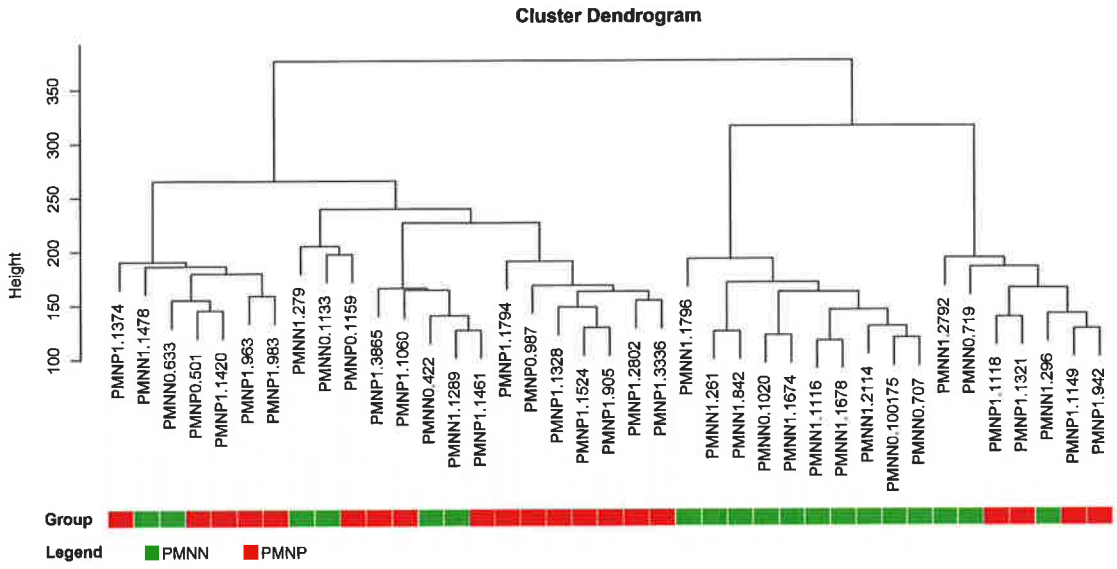
Microarray data are highly-dimensional data. Multivariate analyses are useful to reveal the data patterns for such data, including principal component analysis (PCA), clustering, etc.

PCA can identify the data patterns by reducing data dimensions. The basic idea of PCA is to replace the original variables by a small number of "principal components", which are linear combinations of the initial variables, mutually uncorrelated, and are ordered according to the variance they captured from the original data. Typically it is only the first few that capture important amounts of the total variation. The scatter plot involving the first few PCs usually shows insightful pattern of the original data. PCA result shows the two groups could be roughly separated by first principal component (PC1) and the second PC. (See the figure below)



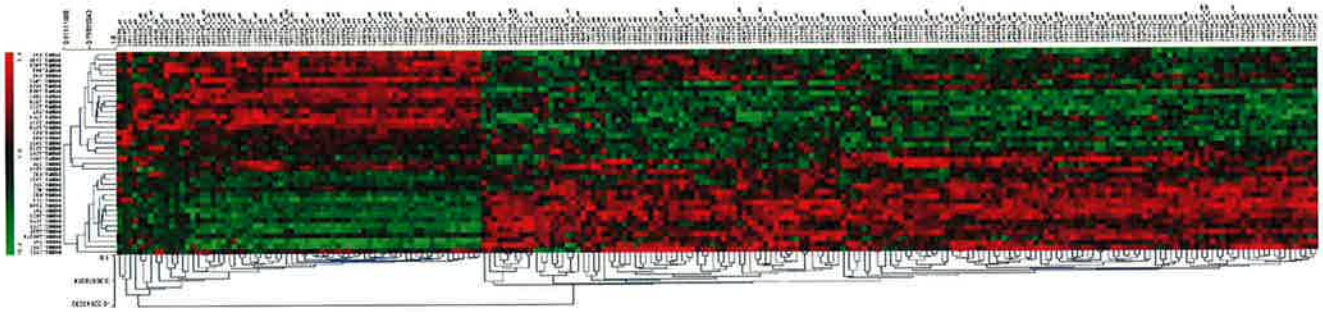
Clustering

Hierarchical clustering is performed on the whole gene expression profiles of post-menopausal samples (both LN+ and LN-). The result shows that these two groups are clustered relatively well, which indicates that these two groups have distinct expression profiles, with most of PMNP samples in left cluster and PMNN in the right one. (See the figure below)

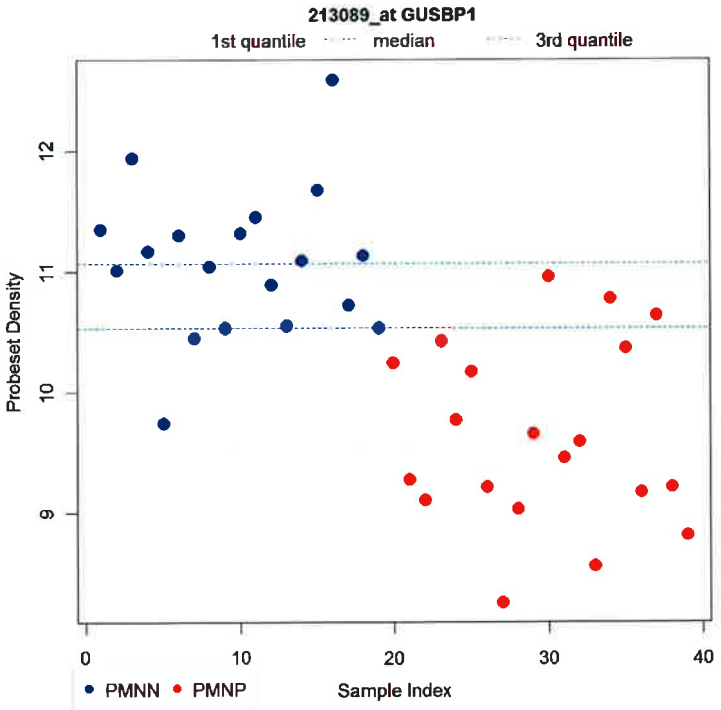


Identification of differentially expressed genes

Differentially expressed genes are identified using Wilcoxon test with Benjamini and Hochberg FDR control (0.1). Over 5,000 transcripts were differentially expressed between post-menopausal node positive and post-menopausal node negative group. Among these transcripts, about 230 have over 2 fold change of their expression levels. (see the gene table below and some gene plots).
Heat map for DEGs with FC2 (GREEN for low expression, RED for high expression)



Below are some examples of the distribution of selected differentially expressed genes in individual tumors from the post menopausal node negative and post menopausal node positive groups. Many of the differentially expressed probes discriminate the lymph node status of individual tumors based on their expression levels. We will use these data in the future to develop classifiers that will distinguish the lymph node status of ER+ tumors based on the express levels of selected genes.



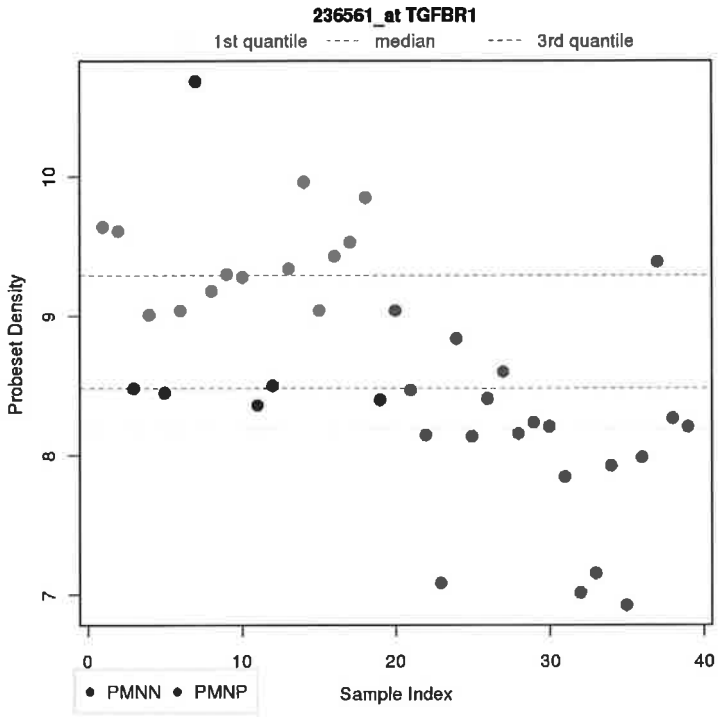


Table 1 Differentially Expressed Genes.

ProbesetID	Gene Symbol	Entrez ID	Raw P-value	adj.p FDR	Mean PMNN	Mean PMNP	FC (NP-NN)
220390_at	AGBL2	79841	0.0001	0.02	6.92	5.70	-2.3
222608_s_at	ANLN	54443	0.0022	0.05	7.02	8.49	2.8
1552619_a_at	ANLN	54443	0.0076	0.09	4.09	5.18	2.1
219918_s_at	ASPM	259266	0.0038	0.07	5.33	6.70	2.6
222740_at	ATAD2	29028	0.0015	0.05	7.05	8.12	2.1
202095_s_at	BIRC5	332	0.0012	0.04	7.17	8.41	2.4
209642_at	BUB1	699	0.0032	0.06	4.96	5.97	2.0
203755_at	BUB1B	701	0.0015	0.05	7.09	8.12	2.0
1561263_at	C1QTNF3	114899	0.0002	0.02	6.21	5.20	-2.0
239349_at	C1QTNF7	114905	0.0045	0.07	5.99	4.90	-2.1
239999_at	C21orf34	388815	0.0038	0.07	5.60	4.63	-2.0
1570571_at	CCDC91	55297	0.0001	0.02	5.86	4.86	-2.0
214710_s_at	CCNB1	891	0.0004	0.03	7.39	8.70	2.5
202705_at	CCNB2	9133	0.0016	0.05	7.57	8.67	2.2
210559_s_at	CDC2	983	0.0007	0.04	6.15	7.51	2.6
203213_at	CDC2	983	0.0008	0.04	7.80	9.15	2.6
203214_x_at	CDC2	983	0.0006	0.03	6.23	7.46	2.4
202870_s_at	CDC20	991	0.0059	0.08	7.19	8.27	2.1
223307_at	CDCA3	83461	0.0007	0.03	6.88	7.90	2.0

1555758_a_at	CDKN3	1033	0.0035	0.06	6.60	7.75	2.2
209714_s_at	CDKN3	1033	0.0070	0.09	6.72	7.77	2.1
228868_x_at	CDT1	81620	0.0018	0.05	5.00	5.98	2.0
204962_s_at	CENPA	1058	0.0020	0.05	6.63	7.84	2.3
219555_s_at	CENPN	55839	0.0004	0.03	6.76	7.98	2.3
218542_at	CEP55	55165	0.0015	0.05	7.05	8.47	2.7
239629_at	CFLAR	8837	0.0006	0.03	8.45	7.33	-2.2
243751_at	CHD2	1106	0.0002	0.02	7.56	6.53	-2.0
205394_at	CHEK1	1111	0.0009	0.04	5.22	6.33	2.2
1552289_a_at	CILP2	148113	0.0012	0.04	7.08	6.10	-2.0
222043_at	CLU	1191	0.0045	0.07	9.42	8.39	-2.0
233109_at	COL12A1	1303	0.0004	0.03	6.76	5.63	-2.2
205713_s_at	COMP	1311	0.0083	0.09	10.55	9.39	-2.2
222958_s_at	DEPDC1	55635	0.0007	0.03	4.83	6.20	2.6
235545_at	DEPDC1	55635	0.0050	0.07	4.42	5.39	2.0
226980_at	DEPDC1B	55789	0.0032	0.06	4.23	5.27	2.1
1558828_s_at	DKFZp586C0721	153688	0.0006	0.03	7.46	6.35	-2.2
1556821_x_at	DLEU2	8847	0.0003	0.03	7.42	6.37	-2.1
230229_at	DLG1	1739	0.0001	0.02	8.25	7.28	-2.0
203764_at	DLG7	9787	0.0045	0.07	5.47	6.76	2.4
222680_s_at	DTL	51514	0.0026	0.06	5.50	6.58	2.1
225275_at	EDIL3	10085	0.0064	0.08	8.93	7.84	-2.1
236700_at	EIF3C	8663	0.0001	0.02	6.92	5.95	-2.0
233498_at	ERBB4	2066	0.0002	0.02	10.03	8.06	-3.9
206794_at	ERBB4	2066	0.0070	0.09	6.50	5.46	-2.0
228069_at	FAM54A	113115	0.0018	0.05	4.40	5.51	2.2
225687_at	FAM83D	81610	0.0015	0.05	6.51	8.07	3.0
204988_at	FGB	2244	0.0076	0.09	5.55	6.59	2.0
1556474_a_at	FLJ38379	285097	0.0090	0.10	9.05	7.60	-2.7
1559964_at	FLJ38717	401261	0.0003	0.03	6.13	5.07	-2.1
1558199_at	FN1	2335	0.0042	0.07	8.03	6.97	-2.1
215203_at	GOLGA4	2803	0.0004	0.03	7.65	6.63	-2.0
213089_at	GUSBP1	153561	0.0000	0.02	11.08	9.64	-2.7
232889_at	GUSBP1	153561	0.0000	0.02	9.00	7.68	-2.5
206548_at	hCG_1776259	79938	0.0004	0.03	6.97	5.78	-2.3
207165_at	HMMR	3161	0.0015	0.05	7.24	8.49	2.4
209709_s_at	HMMR	3161	0.0035	0.06	5.45	6.48	2.0
203424_s_at	IGFBP5	3488	0.0064	0.08	10.34	11.41	2.1
203426_s_at	IGFBP5	3488	0.0090	0.10	6.45	7.47	2.0
214927_at	ITGBL1	9358	0.0004	0.03	9.91	8.72	-2.3
1557080_s_at	ITGBL1	9358	0.0015	0.05	8.28	7.19	-2.1
202503_s_at	KIAA0101	9768	0.0007	0.04	9.51	10.52	2.0
216000_at	KIAA0484	57240	0.0000	0.02	6.88	5.58	-2.5
215268_at	KIAA0754	643314	0.0000	0.02	5.60	4.42	-2.3
243589_at	KIAA1267	284058	0.0001	0.02	8.02	6.95	-2.1

218755_at	KIF20A	10112	0.0009	0.04	6.93	8.22	2.5
208767_s_at	LAPTM4B	55353	0.0022	0.05	7.35	8.68	2.5
203276_at	LMNB1	4001	0.0005	0.03	5.66	6.72	2.1
239965_at	LOC151878	151878	0.0001	0.02	6.49	5.49	-2.0
235482_at	LOC400960	400960	0.0000	0.02	7.99	6.88	-2.2
1566079_at	LOC647190	647190	0.0009	0.04	8.04	6.97	-2.1
213605_s_at	LOC728411	728411	0.0000	0.02	11.12	9.69	-2.7
230653_at	LOC728555 | LOC730391	728555 | 730391	0.0003	0.03	9.70	8.64	-2.1
215599_at	LOC730390 | SMA4	11039 | 730390	0.0000	0.02	8.25	6.58	-3.2
243874_at	LPP	4026	0.0000	0.02	8.60	7.55	-2.1
203362_s_at	MAD2L1	4085	0.0001	0.02	6.50	7.81	2.5
226210_s_at	MEG3	55384	0.0004	0.03	7.52	6.33	-2.3
232568_at	MGC24103	158295	0.0003	0.03	7.44	6.38	-2.1
218883_s_at	MLF1IP	79682	0.0018	0.05	6.56	7.59	2.0
229305_at	MLF1IP	79682	0.0004	0.03	5.15	6.12	2.0
1554906_a_at	MPHOSPH6	10200	0.0018	0.05	4.79	5.91	2.2
228608_at	NALCN	259232	0.0004	0.03	6.60	5.61	-2.0
218663_at	NCAPG	64151	0.0022	0.05	4.39	5.45	2.1
218662_s_at	NCAPG	64151	0.0018	0.05	5.57	6.55	2.0
204641_at	NEK2	4751	0.0045	0.07	7.16	8.36	2.3
223381_at	NUF2	83540	0.0059	0.08	6.55	7.80	2.4
236930_at	NUMB	8650	0.0006	0.03	6.42	5.39	-2.0
218039_at	NUSAP1	51203	0.0003	0.03	8.44	9.64	2.3
219978_s_at	NUSAP1	51203	0.0004	0.03	5.95	7.13	2.3
213568_at	OSR2	116039	0.0035	0.06	7.60	6.63	-2.0
219148_at	PBK	55872	0.0050	0.07	5.52	6.73	2.3
232304_at	PELI1	57162	0.0002	0.03	6.78	5.64	-2.2
204086_at	PRAME	23532	0.0020	0.05	4.58	5.67	2.1
218009_s_at	PRC1	9055	0.0026	0.06	8.12	9.17	2.1
237180_at	PSME4	23198	0.0001	0.02	8.42	7.34	-2.1
203554_x_at	PTTG1	9232	0.0004	0.03	8.64	9.76	2.2
228613_at	RAB11FIP3	9727	0.0010	0.04	9.93	8.95	-2.0
205024_s_at	RAD51	5888	0.0016	0.05	7.04	8.01	2.0
204146_at	RAD51AP1	10635	0.0010	0.04	7.16	8.17	2.0
230742_at	RBM6	10180	0.0020	0.05	7.18	6.11	-2.1
238047_at	RP13-102H20.1	158763	0.0015	0.05	2.82	4.39	3.0
236621_at	RPS27	6232	0.0016	0.05	6.41	5.44	-2.0
201890_at	RRM2	6241	0.0070	0.09	8.99	10.13	2.2
204051_s_at	SFRP4	6424	0.0059	0.08	9.11	7.84	-2.4
204052_s_at	SFRP4	6424	0.0050	0.07	8.74	7.56	-2.3
219215_s_at	SLC39A4	55630	0.0011	0.04	7.39	8.61	2.3
209891_at	SPC25	57405	0.0001	0.02	4.56	5.97	2.6
226086_at	SYT13	57586	0.0029	0.06	7.07	8.87	3.5

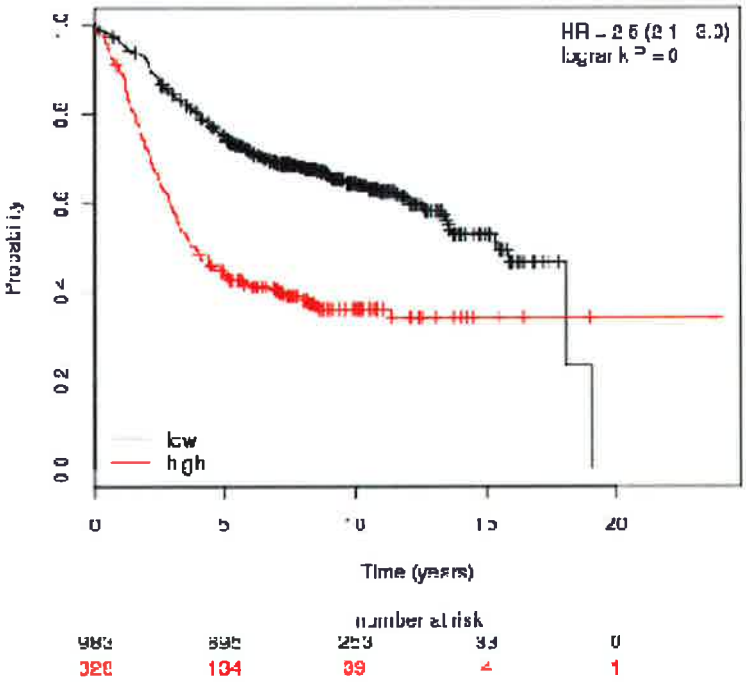
209277_at	TFPI2	7980	0.0083	0.09	4.49	5.93	2.7
236561_at	TGFBR1	7046	0.0000	0.02	9.21	8.10	-2.2
219580_s_at	TMC5	79838	0.0070	0.09	8.54	7.20	-2.5
216005_at	TNC	3371	0.0045	0.07	6.20	4.92	-2.4
201292_at	TOP2A	7153	0.0018	0.05	6.90	8.30	2.6
201291_s_at	TOP2A	7153	0.0090	0.10	8.11	9.30	2.3
238688_at	TPM1	7168	0.0016	0.05	8.72	7.59	-2.2
210052_s_at	TPX2	22974	0.0059	0.08	6.19	7.25	2.1
204822_at	TTK	7272	0.0024	0.05	5.24	6.38	2.2
215898_at	TTLL5	23093	0.0001	0.02	5.33	4.32	-2.0
202954_at	UBE2C	11065	0.0014	0.04	7.40	8.56	2.2
222357_at	ZBTB20	26137	0.0000	0.02	8.12	7.14	-2.0
239757_at	ZFAND6	54469	0.0001	0.02	8.09	7.04	-2.1
239243_at	ZNF638	27332	0.0000	0.02	7.68	6.47	-2.3

Table 1 (above) lists the probes (genes) that are differentially expressed between the post menopausal node negative (PMNN) and post menopausal node positive (PMNP) groups. Examining this list by a number of methods including reviewing the literature as summarized in the On-line Mendelian Inheritance in Man (OMIM) and using bioinformatics techniques such as Gene Set Enrichment Analysis (Subramanian, *et al.* 2005. *PNAS*, **102**: 15545-15550) reveal a number of biologically significant associations. Genes that are preferentially expressed in the PMNP group are frequently associated with cell cycle, kinetochore and spindle formation and the interaction of these processes with DNA repair. We will explore the implications of these preliminary findings in future projects.

Another analysis that we carried out in a subset of the genes identified in table 1, was to examine how the levels of expression of these genes affect the progression of breast cancer and the ultimate survival of the patient. This analysis uses a web-based tool, KMplot, based on work (Gyorffy B, Lanczky A, Eklund AC, Denkert C, Budczies J, Li Q, Szallasi Z. An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1809 patients, **Breast Cancer Res Treatment,** 2010 Oct;123(3):725-31) calculates survival curves (Kaplan-Meier Plots) based on the expression levels of Affymetrix gene probes in publicly available gene expression sets (GEO) from 1908 cases with known outcomes. Two examples of this analysis are shown below. The role of the expression of two genes, ASPM, which is associated with spindle assembly, and CCBN1, cyclin B1 involved in cell cycle, on the survival of patients with ER+ tumors was examined. In both cases high expression of the gene was an indicator of poor prognosis as shown by the survival curves. We will continue to us this analysis as we attempt to understand the biological implications of the gene expression results generated by this project.

Affy ID or Gene Symbol:	ASPM
Survival:	relapse free survival
Split patients by:	upper quartile
Restrict analysis to subtypes...	
ER status:	positive
Derive ER status from gene expression data (n=1809):	on
Lymph node:	all
Grade:	all

Gene Symbol: 219918_s_at



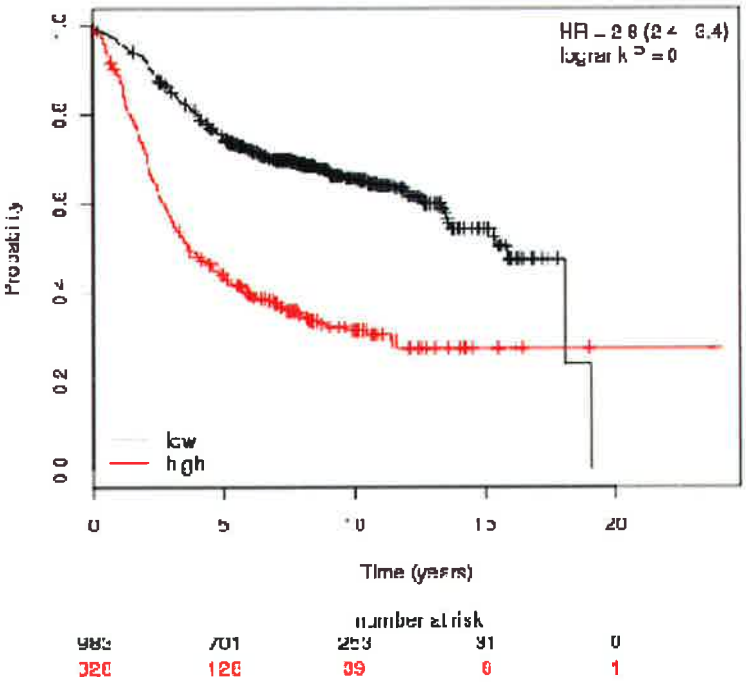
Raw p value = 0

Multiple testing corrected p value = 0

KM plot for ASPM gene

Affy ID or Gene Symbol:	214710_s_at
Survival:	relapse free survival
Split patients by:	upper quartile
Restrict analysis to subtypes...	
ER status:	positive
Derive ER status from gene expression data (n=1809):	on
Lymph node:	all
Grade:	all

Affy ID: 214710_s_at



Raw p value = 0

Multiple testing corrected p value = 0

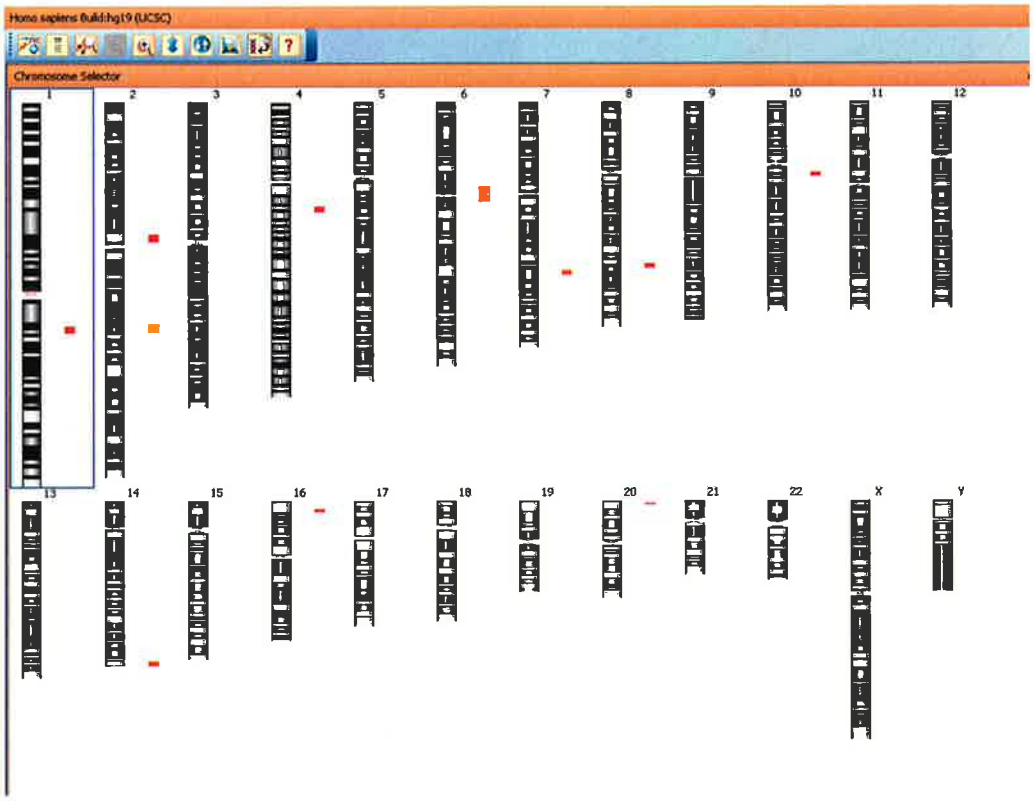
KMplot for CCBN1 gene

ANALYSIS OF SNP CHIPS TO EXAMINE COPY NUMBER VARIATION (CNV)

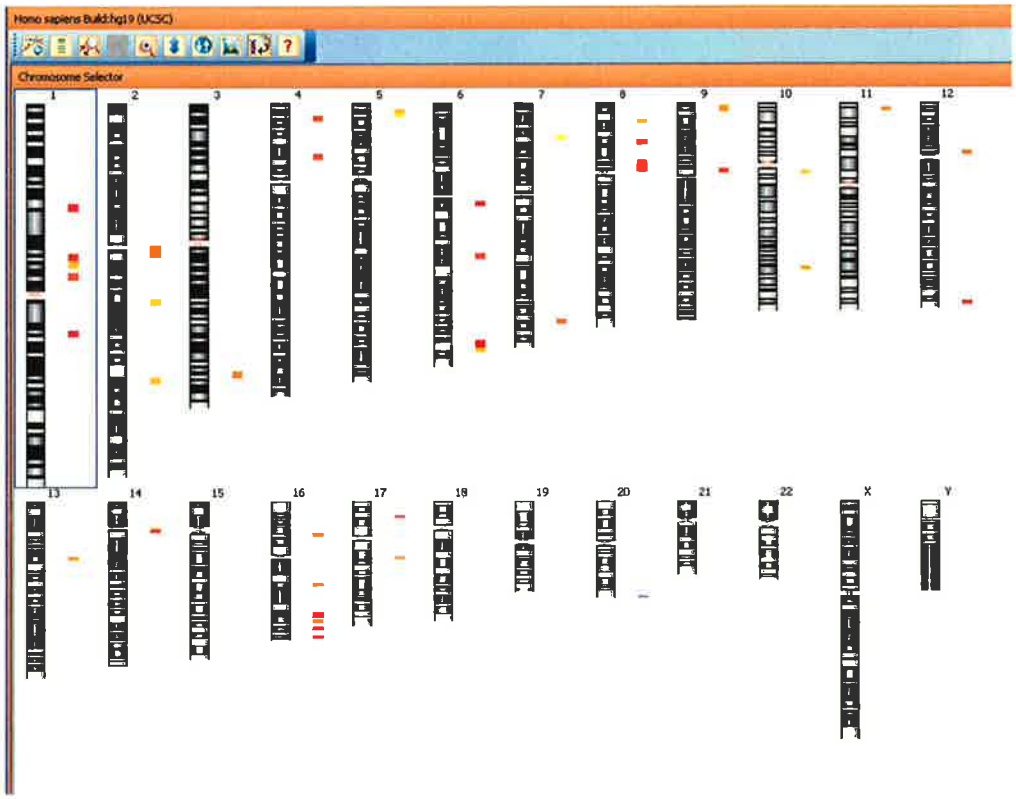
The summary with respect to SNP data generation is given in Appendix 8. As noted before due to budget and time constraints, we proposed to obtain SNP data using Affymetrix SNP 6.0 arrays in 12 samples for each group of this study. As seen from Appendix 8, SNP data has been developed for 10 subjects each in postmenopausal node positive and node negative groups and 11 subjects each for premenopausal node positive and node negative groups. The average contrast quality control threshold for the breast tissue SNPs was 1.27 and for the blood clot/cell control samples was 1.52. These results show that we have this technology working in the lab and that we will be able to use these data in future studies.

The Affymetrix 6.0 platform is designed to yield data, based on the intensity of the hybridization signal, which can be interpreted to measure the copy number of various regions of the genome. We have used these data to determine changes in copy number of various genomic regions in tumors by comparing the single obtained in the tumor DNA with the signal obtained from DNA extracted from a blood sample from the same patient. In this way variation is normalized to the patients own normal DNA and amplifications or deletions seen in the tumor are a result of the growth of the tumor.

Below is an example of such an analysis for one for the tumors used in this study.



Regions Amplified in the genome of the tumor from case CBCP 905.



Regions Deleted from the genome of the tumor from case CBCP 905 (both deletions and amplifications are relative the germ line DNA of the patient)

KEY RESEARCH ACCOMPLISHMENTS:

The clinical and pathological characteristics of the study subjects are given in Appendices 6 and 7.

In the postmenopausal node negative (PMNN) women, the age at diagnosis was higher compared to node positive women (PMNP) (68 vs 61 yrs). The age at diagnosis in premenopausal groups was not different (Appendix 6). The body weight and BMI in PMNN women were also marginally lower compared to PMNP women, while in the premenopausal groups no such differences were noted. It is interesting to note that significant number of women belonging all groups had history of breast cancer in a primary relative and secondary relative, though they were not different between the groups (Appendix 6). Also all the groups had high percentage of women with history of other cancers. Another interesting observation was that node positive women compared to node negative women in both post (1357 vs 854) and premenopausal (1008 vs 739) groups had higher caffeine scores (Appendix 6).

In all groups of women, the rate of cancer was higher in the right breast than in the left breast (Appendix 7). The node negative women compared to node positive women of both pre and postmenopausal groups had higher percentage of women with well differentiated tumors (Post- 40% vs 20%; Pre - 40% vs 21%) and lower percentage of women with poorly differentiated tumors (Post -20% vs 35%; Pre- 11% vs 42%) (Appendix 7). The tumor sizes were marginally higher in node positive groups compared to node negative groups (Appendix 7). As expected, Stage I and Stage 2A cancers were the highest percentage of cancers in node negative post and premenopausal groups. On the other hand the node positive groups in both pre and postmenopausal women had higher AJCC pathological stages of breast cancer (Appendix 7).

Our initial analysis was carried out on the postmenopausal sub group because there was strong differentiation of the node positive (+) and node negative (–) tumors based on the genes that are differentially expressed within this group. Over 5,000 transcripts were differentially expressed between post-menopausal node positive and post-menopausal node negative group. Among these transcripts, about 230 have over 2 fold change of their expression levels. Genes that are preferentially expressed in the PMNP group are frequently associated with cell cycle, kinetochore and spindle formation and the interaction of these processes with DNA repair. We will explore the implications of these preliminary findings in future projects. We were also able to examine how the levels of expression of a sub set of these genes affect the progression of breast cancer and the ultimate survival of the patient.

We have also been able to use the Affymetrix 6.0 SNP platform to identify the regions of the genome in various tumors that have undergone amplification or deletion during the development of the tumor. In future studies we will integrate these data with gene expression data to gain insight into how copy number variation in the tumor genome might effect gene expression and tumor progression.

REPORTABLE OUTCOMES:

1. From the tissue repository at WRI, which contains the samples collected and stored as a part of the ongoing Clinical Breast Care Project (CBCP), we were able to successfully select invasive breast tissue samples meeting our inclusion criteria (post and premenopausal node positive and node negative ER+ breast tumors from European American women). The uniqueness of these tissues is that they are high quality research grade specimens collected for research use using high quality, IRB approved tissue collection and preservation protocols. Extensive clinical and demographic information are also available for CBCP patients, which allowed us to carefully select and stratify the patient samples included in this study.

2. Gene expression profiling was successfully completed for the postmenopausal node positive (n=20) and node negative (n=20) patients as well as for premenopausal node positive (n=19) and node negative (n=18) patients. Tumor cells were isolated from the breast tissues using laser capture micro dissection. Total RNA was isolated from these tumor cells, its quality assessed using an Agilent bioanalyzer and the was quantity determined by a Thermo Nanodrop spectorphotometer. Only high quality RNA was used for gene expression profiling. High quality RNA was amplified (two rounds) and during the amplification process, the quality of RNA was assessed at various steps and measures were taken to obtain high quality aRNA. The gene expression data was generated using the Affymetrix HG U133A plus 2.0 arrays.

3. SNP data was generated using the Affymetrix SNP 6.0 platform for the postmenopausal node positive (n=12) and node negative (n=12) patients as well as for premenopausal node positive (n=12) and node negative (n=12) patients. SNP data was generated for fewer patients because of time and budget constraints. By the end of next month, we will be successfully completing the SNP data generation. High quality DNA (showing a band >10kb on 1% agarose gels) isolated from frozen breast specimen sections and also from the blood cells/clots of the same subjects (control) was used to determine the genomic abnormalities in the tumor tissue. The genotypes for each sample passed Affymetrix quality control metrics with contrast quality control threshold >0.4. These data are being used to generate copy number variation data (CNV), which when compared to the data generated from non tumor DNA (from blood), can be used to determine the region of the tumor genome which have undergone deletion or amplification.

4. We have standardized the whole genome methylation methodology in our laboratory. The number of pulses required to shear the DNA to required length was standardized first. We then standardized the enrichment of methylated DNA by immunoprecipitation using the MIRA assay method using a control DNA. The recovery of the methylated DNA was confirmed using the control samples. As the tumor DNA is limiting and the immunoprecipiated methylated DNA will not be sufficient for the hybridization to the Affymetrix human promoter arrays, we also standardized the whole genome amplification methodology using the GenomePlex Complete Whole Genome Amplification Kit. The fragmentation, labeling and hybridization methodology for the Affymetrix human promoter 1.0R array were standardized using test samples. Ultiminately as a followup to this preliminary study we plan to generate the whole genome methylation data for the postmenopausal node positive (n=10) and node negative (n=10) patients as well as for premenopausal node positive (n=10) and node negative (n=10) patients. An aliquot of DNA isolated from the tumor tissues will be used for methylation studies. The sample data will be analyzed using the Affymetrix Tiling Analysis software.

CONCLUSIONS:

The aim of this pilot project was to obtain comprehensive data on gene expression, epigenetic changes and genomic abnormalities from breast tumor samples in the Windber Research Institute (WRI) tissue repository, collected as part of the Clinical Breast Care Project (CBCP), from patients stratified by lymph node status (diseased or free of disease). This is the first step in a broader project whose goal is to obtain comprehensive molecular data on breast tumors, stroma, and sentinel nodes using all available measurement modalities including gene expression, genomic abnormalities, epigenetic properties such as methylation state, protein abundance, and protein localization. This project focused on measuring alternations in genome structure, the status of DNA methylation and an analysis of gene expression in breast tumors. Other modalities such as protein expression in tumors will be address in other projects.

We were able to complete microarray based gene expression analysis on 77 cases and these data are being combined with data from other CBCP projects for analysis and publication. As with many pilot projects we underestimated the time and cost required to complete the project. As a result we were not able to complete the analysis of copy number variation in the tumor under study but we were able to generate the data for this analysis on 48 of the cases. These data will be analyzed and combined with other CBCP data for further analysis and publication. For the proposed methylation analysis we were only able to validate the technology in our laboratory and we look forward to generating the data from these cases with future funding.

These preliminary studies have given us enough data to be convinced that this multi-modality approach will yield valuable insights into understanding the mechanisms of breast tumor progression and lead, in the future, to more robust methods for managing breast cancer cases.

REFERENCES:

1. Jemal A, Siegel R, Ward E, Hao Y, Xu J, Thun MJ. Cancer statistics, 2009. CA Cancer J Clin 2009;59:225-49.

2. Punglia RS, Morrow M, Winer EP, Harris JR. Local therapy and survival in breast cancer. N Engl J Med 2007;356:2399-405.

3. Perou CM, Sorlie T, Eisen MB et al. Molecular portraits of human breast tumours. Nature 2000;406:747-52.

4. Sorlie T, Perou CM, Tibshirani R et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 2001;98:10869-74.

5. Liu J, Campen A, Huang S et al. Identification of a gene signature in cell cycle pathway for breast cancer prognosis using gene expression profiling data. BMC Med Genomics 2008;1:39.

6. Liu R, Wang X, Chen GY et al. The prognostic role of a gene signature from tumorigenic breast-cancer cells. N Engl J Med 2007;356:217-26.

7. Ma XJ, Salunga R, Dahiya S et al. A five-gene molecular grade index and HOXB13:IL17BR are complementary prognostic factors in early stage breast cancer. Clin Cancer Res 2008;14:2601-8.

8. Paik S, Shak S, Tang G et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 2004;351:2817-26.

9. Sotiriou C, Wirapati P, Loi S et al. Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. J Natl Cancer Inst 2006;98:262-72.

10. van 't Veer LJ, Dai H, Van de Vijver MJ et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002;415:530-6.

11. Wang Y, Klijn JG, Zhang Y et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet 2005;365:671-9.

12. Albertson DG, Collins C, McCormick F, Gray JW. Chromosome aberrations in solid tumors. Nat Genet 2003;34:369-76.

13. Chin K, DeVries S, Fridlyand J et al. Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. Cancer Cell 2006;10:529-41.

14. Knuutila S, Autio K, Aalto Y. Online access to CGH data of DNA sequence copy number changes. Am J Pathol 2000;157:689.

15. Chin SF, Teschendorff AE, Marioni JC et al. High-resolution aCGH and expression profiling identifies a novel genomic subtype of ER negative breast cancer. Genome Biol 2007;8:R215.

16. Jain AN, Chin K, Borresen-Dale AL et al. Quantitative analysis of chromosomal CGH in human breast tumors associates copy number abnormalities with p53 status and patient survival. Proc Natl Acad Sci U S A 2001;98:7952-7.

17. Zhang Y, Martens JW, Yu JX et al. Copy number alterations that predict metastatic capability of human breast cancer. Cancer Res 2009;69:3795-801.

18. Romano G, Michell P, Pacilio C, Giordano A. Latest developments in gene transfer technology: achievements, perspectives, and controversies over therapeutic applications. Stem Cells 2000;18:19-39.

19. Sigalotti L, Fratta E, Coral S et al. Epigenetic drugs as pleiotropic agents in cancer treatment: biomolecular aspects and clinical applications. J Cell Physiol 2007;212:330-44.

20. Jones PA, Laird PW. Cancer epigenetics comes of age. Nat Genet 1999;21:163-7.

21. Li J, Gao F, Li N et al. An improved method for genome wide DNA methylation profiling correlated to transcription and genomic instability in two breast cancer cell lines. BMC Genomics 2009;10:223.

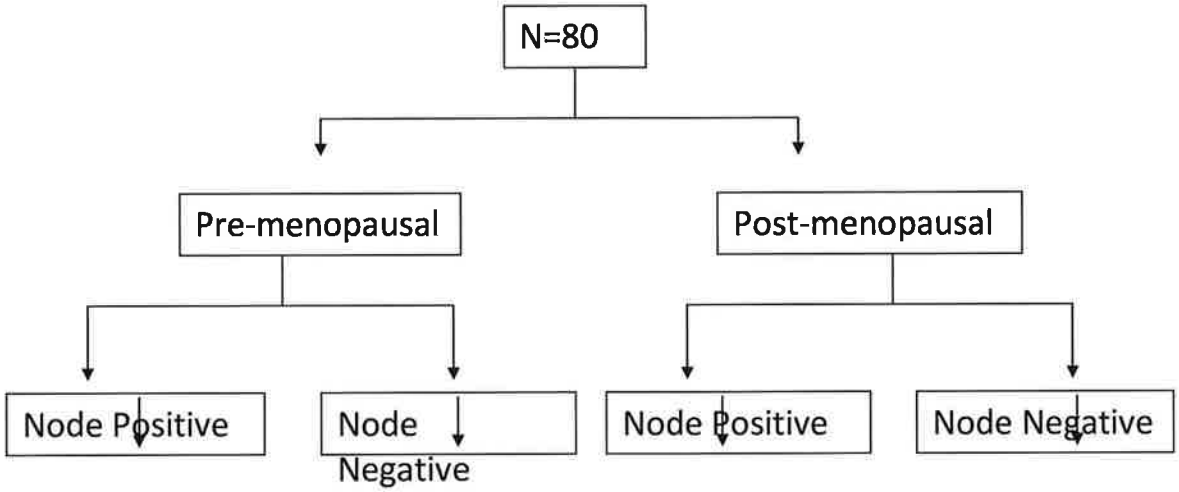
22. Rodenhiser DI, Andrews J, Kennette W et al. Epigenetic mapping and functional analysis in a breast cancer metastasis model using whole-genome promoter tiling microarrays. Breast Cancer Res 2008;10:R62.

23. Sadikovic B, Yoshimoto M, Al-Romaih K, Maire G, Zielenska M, Squire JA. In vitro analysis of integrated global high-resolution DNA methylation profiling with genomic imbalance and gene expression in osteosarcoma. PLoS One 2008;3:e2834.

24. Andrews J, Kennette W, Pilon J et al. Multi-platform whole-genome microarray analyses refine the epigenetic signature of breast cancer metastasis with gene expression and copy number. PLoS One 2010;5:e8665.

APPENDICES:

Appendix 1

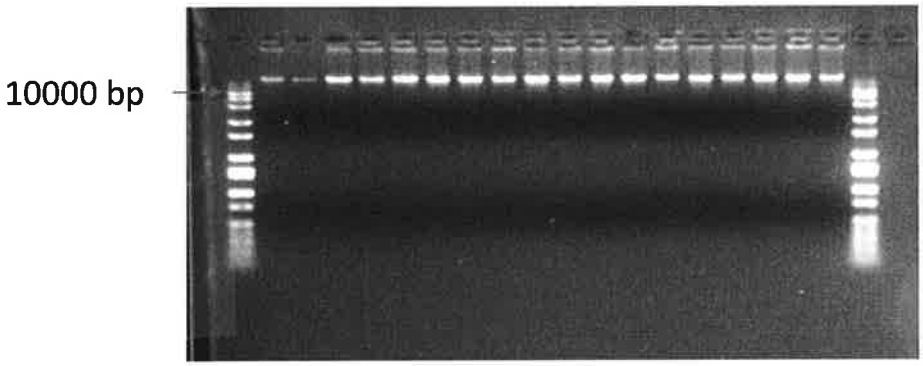


Gene expression	N=19	N=18	N=20	N=20
SNPs	N=12	N=12	N=12	N=12
Methylation profiling	N=10	N=10	N=10	N=10

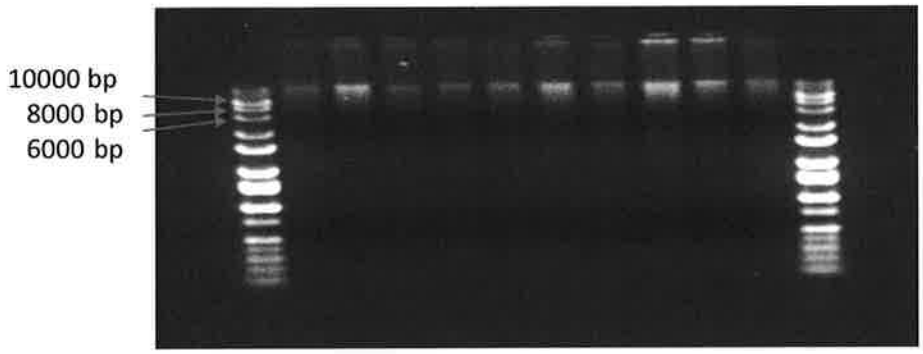
Appendix 2

Representative 1% agarose gel runs of the clot and tissue DNA samples of the ISB subjects

Clot DNA Samples

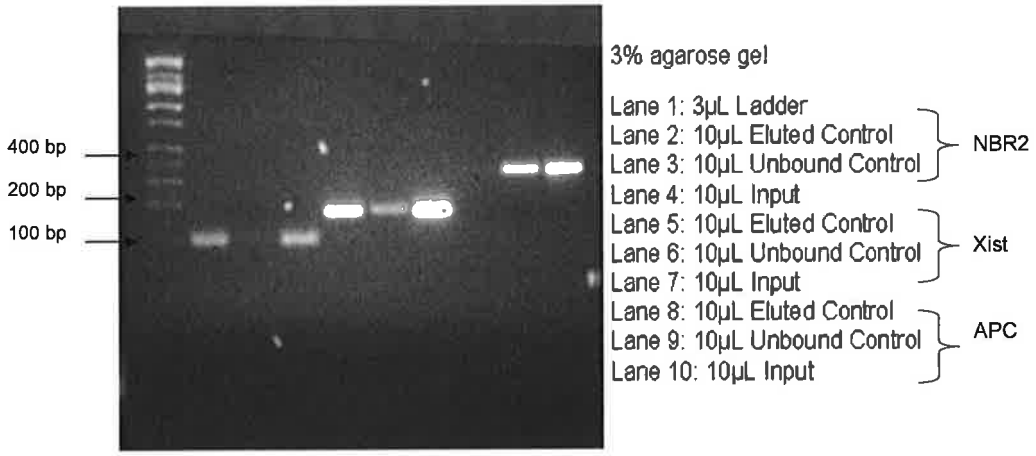


Tissue DNA Samples



Appendix 3

Immunoprecipitation Validation using a control DNA



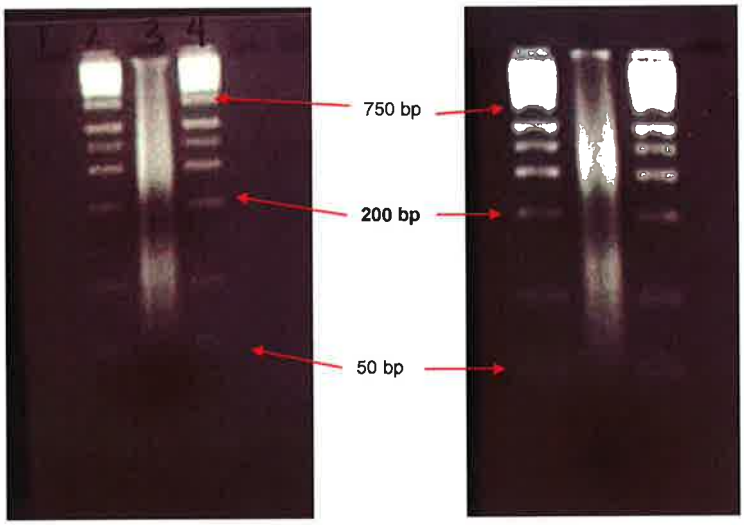
Appendix 4

A. Amplification

1.5% agarose gel

Lane 1: 5µL Ladder
Lane 2: 5µL Test 1
Lane 3: 5µL Test 1
Lane 4: 5µL Ladder

B. Fragmentation



Fragmentation –Test sample 1

Lane 1: Empty

Lane 2: 5µL Ladder

Lane 3: ~3µL sample + 1µL Loading dye

Lane 4: 5µL Ladder

Fragmentation – Test sample 2

Lane 1: 5µL Ladder

Lane 2: ~3µL sample + 1µL Loading dye

Lane 3: 5µL Ladder

Appendix 5

- Analysis based upon two test samples: Test 1 and Test 2
- Test 1 and 2 are the sample tissue sample.
- Test 2 is the duplicated chip used for comparative analysis.

Test 2

TAG File: Test 2.TAG
Genome Map: C:\Documents and Settings\Gerianne Brillhart\Desktop\Library Files\Hs_PromPR_v02-3_NCBI36.bpmap\Hs_PromPR_v02-3_NCBI36.bpmap
GRC File: C:\Documents and Settings\Gerianne Brillhart\Desktop\Library Files\human_promoter_librarytile\CD_Hs_PromP_v02\Full\Hs_PromPR_v02\LibFiles\Hs_PromPR_v02.grc
Analysis Type: One sample detection analysis

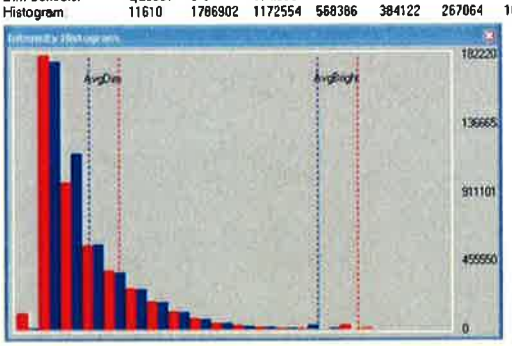
Hist Scale: LOG2
Min X Value: 5.72792
Max X Value: 16

Test 2_(Hs_PromPR_v02).CEL

Minimum Value:	53																			
Maximum Value:	65534																			
Mean Value:	393.328																			
Bright Controls:	2/28768	0.00695217%		14660.9																
Dim Controls:	0/28531	0%	287.982																	
Histogram	107631	1822203	980977	561366	396817	279264	191093	124285	77373	45700	26913	15773	10153	7028	5522	28208	7429	3114	577	130

Test 1_(Hs_PromPR_v02).CEL

Minimum Value:	62																			
Maximum Value:	65491																			
Mean Value:	321.321																			
Bright Controls:	3/28768	0.0104283%		7601.21																
Dim Controls:	0/28531	0%	174.835																	
Histogram	11610	1786902	1172554	568386	384122	267064	181557	115894	69818	41030	22993	13994	9286	27834	9040	5854	2352	1024	182	60



Appendix 6: Clinical Characteristics of the ISB subjects

Variable	PMNN	PMNP	PRNN	PRNP
N	20	20	18	19
Age at diagnosis (yrs)	68.05±9.59	61.35±6.93	44.65±4.03	44.74±6.05
Weight (lbs)	152.6±25.99	169.85±37.3	154.82±34.76	157.05±37.51
BMI	26.72±4.84	29.15±5.33	26.02±5.99	27.25±6.56
History of Breast Cancer in Primary relative(%)				
Yes	30	35	27.8	21
No	70	65	72	78.95
History of Breast Cancer in secondary relative(%)				
Yes	30	40	44.99	42
No	70	60	44.44	57.89
History of Ovarian Cancer (%)				
Yes	5	5	11.1	10.5
No	95	95	83.3	84.2
History of Other Cancers (%)				
Yes	60	50	55.6	73.7
No	40	50	38.9	26.3
Age at Menopause	51.71±2.78	49.6±3.73	N/A	N/A
HRT use (%)				
Yes	35	60	11.1	5.3
No	65	40	88.9	94.7
History of smoking (yrs)	11.19	14.45	5.82	9.17
Alcohol usage- past year (%)				
never	30	20	22.2	10.5
1 time per month or less	40	35	16.6	42.1
more than 2 times/ month or more	30	40	55.5	47.4
Exercise frequency				
Never	30	50	27.8	31.6
1-3 times	20	20	16.7	26.3
3 times or more	50	30	49.99	42.1
Caffeine score	854.55±754.9	1357.33±1128.6	739.19±591.7	1007.71±619.1
Fat intake score	26±6.03	30.6±2.97	28.88±6.99	30.47±5.00

PMNN: Postmenopausal Node Negative; PMNP: Postmenopausal Node Positive

PRNN: Premenopausal Node Negative; PRNP: Premenopausal Node Positive

Appendix 7: Pathological Characteristics of the ISB subjects

Variable	PMNN	PMNP	PRNN	PRNP
N	20	20	18	19
Location of cancer (%)				
Right Breast	70	60	55.6	57.9
Left Breast	30	40	44.4	42.1
Invasive grade (%)				
Well differentiated	40	20	38.9	21
Moderately differentiated	40	45	50	36.8
poorly differentiated	20	35	11.1	42.1
Invasive size (cms)	2.27	2.56	1.97	2.77
PR (%)				
Positive	80	80	88.9	94.7
Negative	20	20	11.1	5.26
Her2 (%)				
Positive	35	35	33.3	15.8
Negative	40	35	55.6	57.9
Not performed	25	30	11.1	26.3
AJCC Path Stage (%)				
Stage I	40	0	50	0
Stage 2A	50	40	50	42
stage 2B	5	25	0	31.6
Stage 3A	0	25	0	21
Stage 3B	5	0	0	0
Stage 4	0	10	0	5.26

PMNN: Postmenopausal Node Negative; PMNP: Postmenopausal Node Positive

PRNN: Premenopausal Node Negative; PRNP: Premenopausal Node Positive

SUMMARY OF SNP DATA GENERATION

GROUP	Number of samples	Breast Tissue		Blood Clots (Control)	
		Number completed	Avg CQC	Number completed	Avg CQC
PMNN	12	10	1.36	7	1.59
PMNP	12	10	1.28	8	1.5
PRNN	12	11	1.24	11	1.48
PRNP	12	11	1.19	9	1.52