National Surgical Adjuvant Breast and Bowel Project (NSABP) Foundation

Annual Progress Report: 2009 Formula Grant

Reporting Period

July 1, 2013 – December 31, 2013

Formula Grant Overview

The National Surgical Adjuvant Breast and Bowel Project (NSABP) Foundation, Inc. received $1,255,581 in formula funds for the grant award period January 1, 2010 through December 31, 2013. Accomplishments for the reporting period are described below.

Research Project 1: Project Title and Purpose

Discovery and Validation of MicroRNAs as Biomarkers in Breast and Colon Cancer - Messenger RNA (mRNA) expression profiling has yielded useful prognostic tests for breast cancer (e.g., Onco*type* DX® and MammaPrint®), but has not been very useful in identifying definitive treatment targets. This is because mRNA expression profiling does not necessarily provide insight about which changes may be susceptible to therapies. For colon cancer, no clinically useful gene-expression-based prognostic tests exist yet.

MicroRNAs (miRNAs) are a subset of RNAs which have regulatory functions and, so, could be viewed as master switches for many coordinately regulated genes. This project's purpose is to identify miRNAs that are prognostic or predictive of benefit from systemic therapies for breast and colon cancers, using archived tumor tissues from completed NSABP clinical trials. Identified markers may point us to new treatment strategies.

Duration of Project

1/1/2010 - 12/31/2013

Project Overview

The long-term goal of this research project is to develop biomarkers that will improve the treatment of colon and breast cancers. To achieve this goal, we propose to examine the expression of microRNAs (miRNAs), which are small, non-coding RNAs that control translation of many mRNAs. miRNA expression analysis may not only allow for the identification of better biomarkers but may provide a greater understanding of current expression data.

One of the fundamental problems associated with the treatment of breast and colon cancers is clinical heterogeneity. Gene-expression mRNA signatures that distinguish several subtypes of

breast cancer have been identified, and, while no specific signatures currently exist for colon cancer, the molecular and clinical heterogeneity of colon tumors is known. The subtyping of breast cancers has vastly improved breast cancer treatment because different subtypes respond differently to different breast cancer therapies. Current molecular signatures that distinguish these subtypes of breast cancer involve the profiling of a large number of genes and have primarily been identified by whole genome mRNA expression analysis. Whole genome expression analyses are too costly and too complicated, and have too much background interference to be used as routine clinical tests.

miRNAs may be able to give prognostic and predictive information because they act as major switches that regulate the expression of many mRNAs. It may be possible to better characterize a tumor with a small number of miRNAs than is possible with mRNAs. Colon cancer lacks biomarkers for identifying tumors that have a high risk for recurring, and predicting response to chemotherapy is not possible. Given these considerations, we have adopted the following specific aims:
1) Discover and validate prognostic miRNA markers for breast cancer patients with resectable tumors and histologically positive axillary lymph node(s);
2) Discover and validate predictive and prognostic miRNA markers for colon cancer patients with stage II and III resectable colon cancer; and
3) Initiate the development of an integrated colon cancer molecular database using miRNA data as well as mRNA and mutation data from the NSABP clinical trial C-07 samples which will provide unique information for hypothesis generation.

Principal Investigator

Katherine L. Pogue-Geile, PhD
Assistant Director of Molecular Profiling
NSABP Foundation, Inc.
Two Allegheny Center, Suite 1200
Pittsburgh, PA 15212

Other Participating Researchers

Soonmyung Paik, MD; Patrick Gavin, BS; Chung Kim, MD – employed by NSABP Foundation, Inc.

Expected Research Outcomes and Benefits

In this project, we will examine the expression of miRNAs in breast and colon cancers.

The project has the following expected outcomes:
1) To identify the gene expression profile of miRNAs in a large number of colon and breast cancer samples;
2) To determine the prognostic significance of miRNAs in colon and breast cancers;
3) To determine the predictive significance of miRNAs in colon and breast tumor responses to specific anti-cancer therapies; and

4) To acquire new information for the development of a valuable integrated molecular database of colon tumors. The C-07 colon tumors that will be interrogated for miRNA expression have already been profiled for mRNA gene expression by Illumina® DASL, arrays and profiled for critical cancer mutations in KRAS, NRAS, PIK3CA, BRAF and MET. This information along with standard histological and clinical information makes this a valuable resource for the discovery of the mechanisms that determine the success or failure of specific cancer treatments and to generate hypotheses regarding control of gene expression and its role in cancer development.

Potential benefits include the following:

1) Improved prognostication for patients with diagnoses of breast or colon cancer;
2) Improved prediction of response of patients with colon cancer to oxaliplatin; and
3) Generation of information for the development of integrated molecular databases. C-07 miRNA data will add to other molecular information on the same clinical samples, including data on whole genome mRNA expression and somatic cancer mutations. This database can be used for hypothesis-generation as it will provide a large database for correlating these molecular characteristics.

Summary of Research Completed

The information included in this report refers only to the work that was done between July 1, 2013 and December 31, 2013. The new information in this report includes the analysis of our molecular data with 5-year clinical follow-up data. This should improve on previous analyses, which used 3-year follow-up data (due to the larger number of events in the 5-year clinical data) and will help us to better evaluate patients' outcomes.

A. The following description addresses specific aim 2 which was to "<u>Discover and validate predictive and prognostic miRNA markers for colon cancer patients with stage II and III resectable colon cancer";</u>

Assessing the Prognostic and Predictive Value of Mirs

We assessed the prognostic and predictive value of each mir individually using expression as continuous or categorical values in Cox models in 50 bootstrap data sets, each with 2/3 randomly selected samples. Twelve mirs were found to be significant for prognosis at a mean p value of less than 0.1, by continuous or categorical p value. None of the mirs were significant for bevacizumab prediction when they were analyzed as a continuous variable, but eighteen mirs were significant when they were analyzed as categorical values using a p value below 0.1.

We also used mir data to subtype samples with Non-negative Matrix Factorization unsupervised clustering of mir data. The optimum number of clusters was determined by cophenetic correlation coefficients which indicated that our samples could best be divided into 3 to 5 clusters. Unsupervised cluster analysis of samples using mir data identified 3 mir cluster groups with a trend for differential benefit from bevacizumab, with cluster group 1 receiving some benefit, group 2 receiving no benefit, and group 3 harmed by treatment (Figure 1). Red lines are plots for the control patients and black lines are for patients from the bevacizumab arm. These

results are unlikely to be clinically meaningful since the results were not significant in any of the groups, and the degree of benefit in group 1 is small (hazard ratio [HR] = 0.857).

The last specific aim on our proposal was to initiate the development of an integrated colon cancer molecular database using mir data as well as mRNA and mutation data. Thus, we have compared the prognostic and predictive p values for mirs and for genes within the discovery cohort (Figure 2). The genes were profiled with a custom code set using the nCounter platform and were selected as either prognostic in C-07 using Illumina DASL arrays or predictive for bevacizumab benefit in C-08 using Agilent arrays. Expression of some mirs was undetectable or other technical problems resulted in missing data. Thus, in Figure 2 we have removed mirs for which data were missing from more than 20% of the samples. In general, the p values for prognosis and prediction are lower in the nCounter or NanoString data than in the mir data, as would be expected since there was a pre-selection for such genes in the NanoString code set.

B. The following analyses were performed to address specific aim 3 which was "Initiate the development of an integrated colon cancer molecular database using miRNA data as well as mRNA and mutation data from the C-07 samples which will provide unique information for hypothesis generation". We changed our focus to use the National Surgical Breast and Bowel Project (NSABP) clinical trial C-08 cases and have discussed in other reports the reason for this change. Below we discuss our work to integrate mRNA and micro RNA data.

Prognostic and Predictive Model Development within the Discovery Cohort using mir and NanoString Data

Mir and or NanoString data were used to build prognostic models using only the control arm of C-08 and the SuperPC method (Figure 3). Fifty bootstrapping datasets, each with two thirds of randomly selected samples were used. Genes with mean p value less than 0.05 were treated as stable genes. This analysis demonstrates that mirs or NanoString genes are able to segregate patients with good or poor prognoses. However, perhaps the best model is from the NanoString data because it yielded the largest HR (3.898) with a p value =0. Combining the NanoString data with the mir data did not add to the prognostic model; the hazard ratio was essentially the same (HR=3.636).

We have compared the clustering samples from mir data (Figure 2) to NanoString data using non-negative matrix factorization (NMF) (Figure 4). As in the analysis of the mir data, the optimum number of clusters for NanoString data, determined by cophenetic correlation coefficients, was 3 to 5 clusters. Similar to clustering using mir data, the NanoString data also show 3 clusters with differential benefit from bevacizumab: one for benefit, one for harm, and one with no effect. Furthermore, there is significant overlap between clusters identified with mir or NanoString data (Table 1). Cluster group 1 with mir data was the group that received benefit and it is largely found in groups 1 and 2 in NanoString data, but in the NanoString cluster 1, only group 1 received benefit so NanoString was better at identifying patients who received benefit. With NanoString data, cluster group 3 received harm and overlapped with group 2 and 3 of mir data. In this instance, mir data did a better job of identifying patients who would receive harm. This demonstrates that by integrating the mir and NanoString data, we may be able to refine the subset with differential benefit.

Models built with immune mirs and genes
Several lines of evidence suggest that the immune reaction plays a critical role in the prognosis of colon cancer and any treatment that affects the immune reaction may alter the therapeutic effect. Our pathway analyses of whole genome expression of 447 tumors from Stage II and III colorectal tumors collected in NSABP clinical trial C-08 is consistent with these observations. The 5 most significantly differentiated pathways between deficient mismatch repair (dMMR) and proficient MMR (pMMR) tumors were 4 T-cell pathways (unpublished data).

Nineteen immune related mirs and 4-immune related NanoString genes were used to cluster samples within C-08. Clustering by K means was used to arrange patients into 3, 4, 5 or 6 clusters, and the benefit from bevacizumab was evaluated in each cluster. Clustering into either 4 or 5 different clusters identified a subset with a significant benefit from bevacizumab. The best results occurred when cases were clustered into 4 groups. One cluster in this analysis showed significant benefit from bevacizumab of patients had hazard ratio (HR) =0.484, p=0.006; two other clusters had a trend toward harm and another showed no effect (Figure 5).

Table 1. Overlap of clusters identified by micro RNA (MIR) and NanoString using non-negative matrix factorization (NMF).

		NanoString		
		1Benefit	2 None	3 Harm
MIR	1 Benefit	295	212	44
	2 None	42	54	144
	3 Harm	21	22	119

Figure 1. Clustering of (micro RNA) (MIR) data. Hazard ratios (HR) are shown.

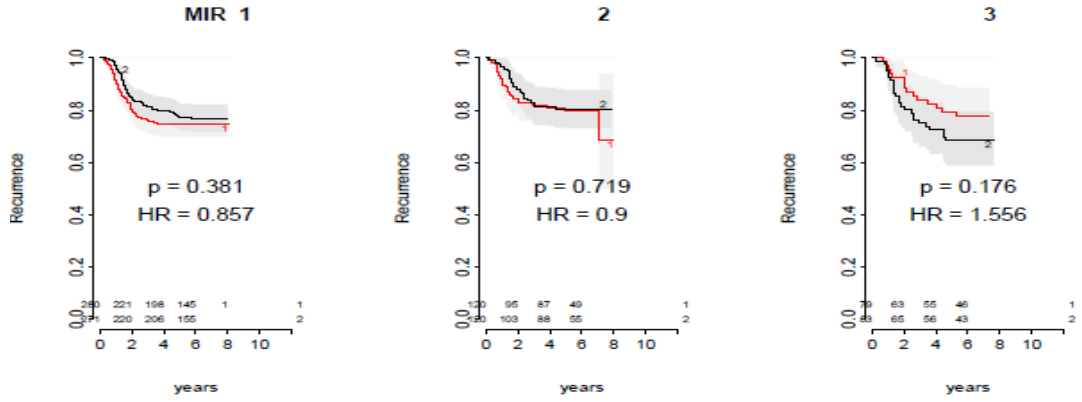

Figure 2. Prognostic and predictive P values for individual micro RNA (mirs) and NanoString genes (nano).

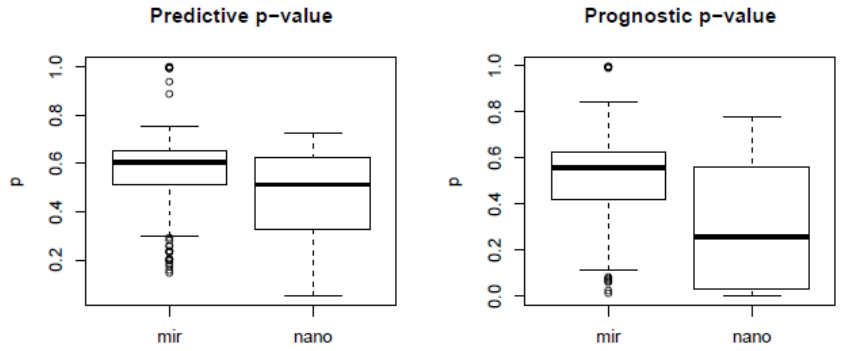

Figure 3. Prognostic models built with SuperPC using Micro RNA (MIR) and Nanostring data Hazard ratios (HR) are shown.

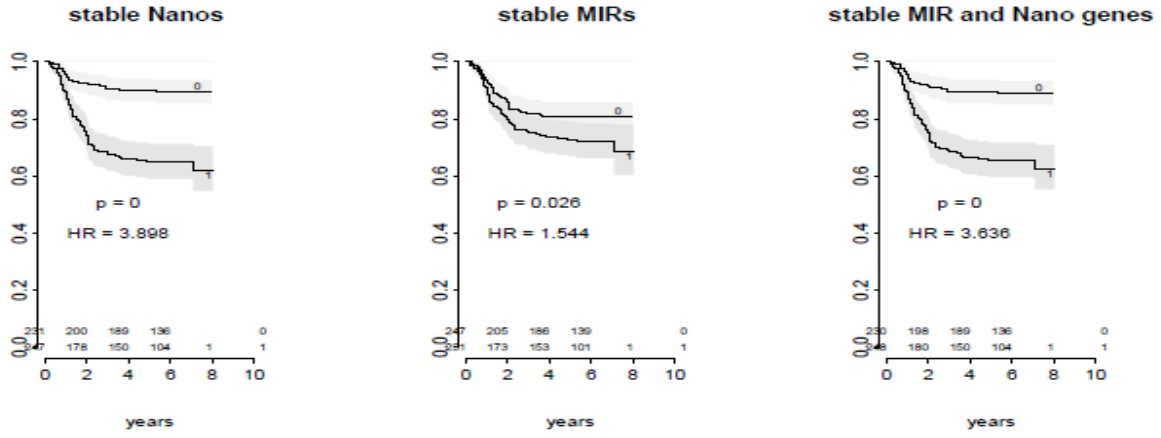

Figure 4. Clustering of colon clinical trial number 8 (C-08) using micro RNA (MIR) or NanoString data. Hazard ratios (HR) are shown.

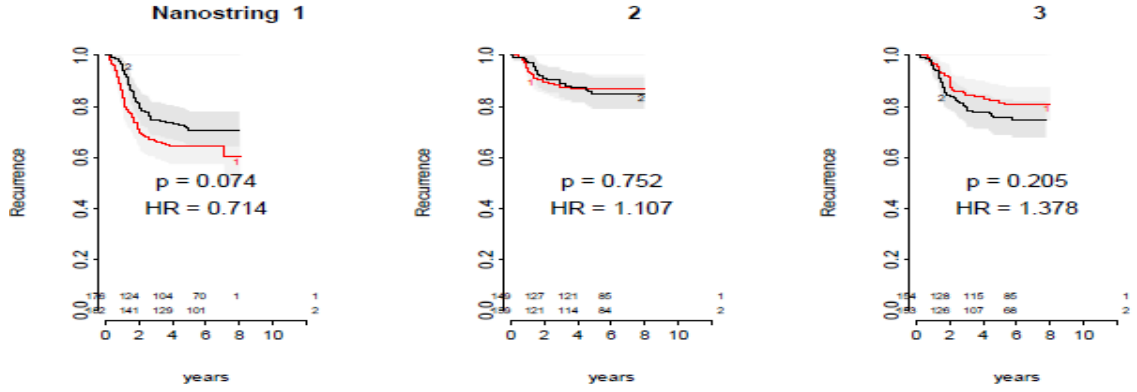
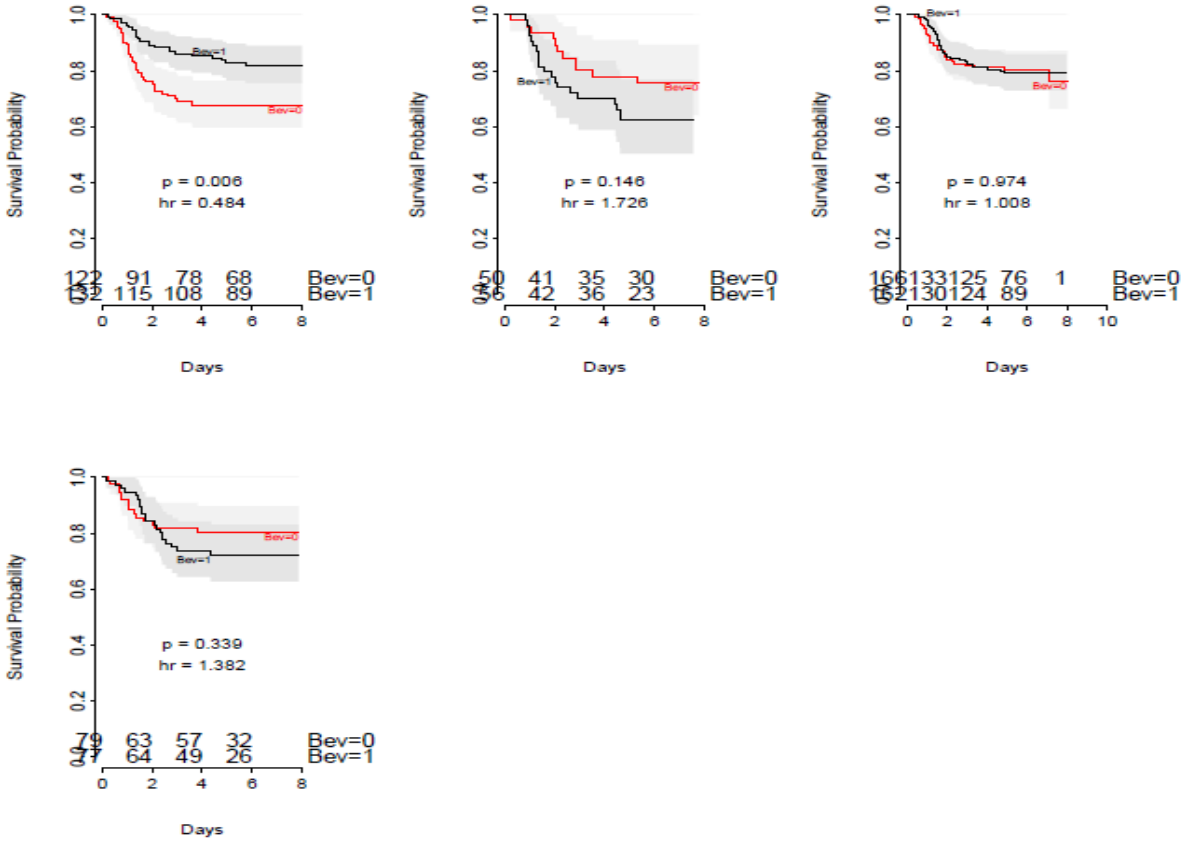
Figure 5. Clustering of samples using immune-related micro RNA (mirs) and genes. Hazard ratios (hr) are shown for patients treated with bevacizumab (bev=1) versus those treated with chemotherapy alone (bev=0).