National Surgical Adjuvant Breast and Bowel Project (NSABP) Foundation

Annual Progress Report: 2010 Formula Grant

Reporting Period

July 1, 2013 – June 30, 2014

Formula Grant Overview

The National Surgical Adjuvant Breast and Bowel Project (NSABP) Foundation, Inc. received $967,922 in formula funds for the grant award period January 1, 2011 through December 31, 2014. Accomplishments for the reporting period are described below.

Research Project 1: Project Title and Purpose

Next-generation Sequencing-based RNA Expression Profiling for Archived Tumor Tissue - The purpose of this project is to develop a protocol for next-generation (NexGen) sequencing of RNAs isolated from formalin-fixed, paraffin-embedded tissues.

Anticipated Duration of Project

1/1/2011 - 12/31/2014

Project Overview

The broad goals of the NSABP pathology laboratory are to find molecular markers that can be used for prognostic or predictive information that can improve the diagnosis and treatment of breast and colon cancers. The procedure currently used for the discovery of prognostic and predictive biomarkers involves several steps. First, whole genome expression analysis is done on a large discovery cohort consisting of approximately 1000 samples from NSABP's vast resources of archived tissues; second, prognostic and predictive genes are identified. Third, the expression profiles on the discovery cohort are repeated with a platform that is amenable as a clinically appropriate test. This step, referred to as platform validation, is important because genes discovered on the whole transcriptome platform are not always validated on a smaller clinically appropriate platform; also, this step adds a great deal of time, and cost to biomarker discovery. The final step is the clinical validation which interrogates a new cohort of patients. Next-generation (NexGen) RNA sequencing is a powerful new technology that allows for an unbiased and complete examination of the transcriptome. Furthermore, this technology is being developed as a clinical test and would simplify the process of discovery and validation.

To expedite the process of biomarker discovery and to lead to new biological insights into breast cancer, we propose the following specific aims: 1) determine the best platform for NexGen

analysis of RNAs isolated from formalin-fixed, paraffin-embedded (FFPE) tissue material; 2) purchase and install instruments and workflow for NexGen sequencing at the NSABP; and 3) evaluate NexGen gene-expression profiling data from RNAs isolated from formalin-fixed, paraffin-embedded tissues (FFPET) collected in NSABP clinical trial B-20.

To address aim 1: RNAs from cell lines, from both fresh and FFPE preparations, will be isolated. cDNA libraries will be prepared and profiled on both the Illumina® and Life Technologies instruments. Data will be evaluated for reproducibility and accuracy to determine the best platform for use with RNAs isolated from FFPET. In the second phase we will purchase the equipment and perform NexGen RNA sequencing analysis using 200 RNAs isolated from FFPET from NSABP clinical trial B-20. The data will be evaluated for biological accuracy and will provide preliminary data for development of a prognostic gene signature for ER+ and node-negative breast cancer patients.

Principal Investigator

Katherine L. Pogue-Geile, PhD
Assistant Director of Molecular Profiling
NSABP Foundation, Inc.
Four Allegheny Center, 5th Floor
Pittsburgh, PA 15212

Other Participating Researchers

Patrick Gavin, BS; Nan Song, PhD – employed by NSABP Foundation, Inc.

Expected Research Outcomes and Benefits

The following are expected outcomes and benefits of this research:

1. To develop a process and protocol for NexGen RNA sequencing (RNA-Seq) analysis which can be used to profile gene expression utilizing RNAs isolated from archival, FFPE clinical trial samples. The development of such a protocol can unleash a wealth of information in a single experiment that is not possible with other technologies.
2. To provide preliminary information for the development of a gene expression signature that will identify ER+, node-negative breast cancer patients who have good prognoses. Such a test may identify breast cancer patients who only need to be treated with tamoxifen and may spare patients the side effects of chemotherapeutic treatment.

Summary of Research Completed

I. Progress to achieve specific aim 1) determine the best platform for NexGen analysis of RNAs isolated from formalin-fixed, paraffin-embedded (FFPE) tissue material:

Work for this aim has been completed and was reported in a prior reporting period.

II. Progress to achieve specific aim 2) purchase and install instruments and workflow for

NexGen sequencing at the NSABP:

Work for this aim has been completed and was reported in a prior reporting period.

III. Progress to achieve to achieve specific aim 3) evaluate NexGen gene-expression profiling data from RNAs isolated from FFPET collected in NSABP clinical trial B-20:

In our initial application we proposed to provide preliminary information for the development of a gene expression signature for node negative, ER+ breast cancer patients. However, given that there are already good clinical tests which are able to stratify these patients not only for prognosis but also for benefit from chemotherapy, we decided to focus on rectal cancer where no such test exists. We are collaborating with Yonsei University which has collected more than 300 pretreatment rectal cancer samples in the neoadjuvant setting. The purpose of our study is to provide preliminary data for the development of signatures that will be able to predict pathological complete response (pCR) in patients treated with 5-Fluorouracil (5-FU) alone or 5-FU plus oxaliplatin.

I. Progress to achieve outcome 1

During this report period we have optimized RNA isolation and RNA-Seq library construction with small amounts of FFPET. This optimization is essential to this study because the rectal samples that will be used for this study are from small biopsy samples and thus, the yield of RNA is limiting. However, before utilizing the clinical trial rectal samples, it was necessary to do additional optimization using non-clinical trial samples. RNA-Seq Libraries were built using RNAs isolated from 4 cell lines and from commercial FFPET samples purchased from ILS-bio. In these experiments, ribosomal RNA was removed using the Ribo-Zero kit (Epicentre), which represented a major cost of the library preparation. Therefore, we repeated library preparations using an ILS-bio sample and used only one quarter of the reagents recommended by the vendor. We concluded that ¼ of the ribosomal reagents was sufficient for ribosomal depletion because the ribosomal rate (dark blue circles, Figure 1) were nearly identical between run 2014-01-03 using the recommended amount of Ribo-Zero reagents and run 2014-01-24 in which only ¼ of the reagents was used.

As we expected, our exonic mapping rate was low in libraries prepared from FFPET due to the degraded and chemically modified nature of the RNA. To give us some idea of how well our procedure would work when using intact fresh RNA we used RNA from fresh cell lines (SW480, SW620, H747 and SW837). The spaghetti plot shows the rate at which the sequencing reads map to different regions of the genome [rRNA (dark blue), intergenic (light blue), exons (red), intragenic (black)]. The exonic mapping rate of the 4 cell lines was high (~85%) (Figure 1, red circles, runs 2014-02-05 and 2014-02-11). This experiment demonstrates that our procedure works well with intact RNA isolated from fresh samples. Note: The methodology for preparing RNA-seq libraries required that the RNA be fragmented before library preparation. This step was eliminated when libraries were prepared from RNA isolated from FFPET because the size of the RNA was already sufficiently small.

We also tested library preparations using cell lines (SW480, SW620, SW837 and H747) and two

tumor tissue RNAs (ILS-22435 and ILS-31043) isolated from fresh and FFPE tissues from the same patient and tumor. Tissue samples were obtained from ILS-bio. We show good correlation between replicate preparations from the same sample (Table 2). Libraries were also constructed using various amounts of ribosomal depleted RNA and with slight modifications in the library preparation as was dictated by necessity. For example, Life Technologies discontinued the version 1(v1) kits for templating and sequencing, necessitating our retesting the reproducibility of our methodology using the new v2 kit, and then later, v3 kits. Correlations between the same sample and different samples with these variations in the procedure and with different analytical methodologies are shown in Figure 2. It shows that, despite changes in the amount of RNA, source of RNA (FFPE versus fresh) and slight modifications in the wet lab or analytical approaches of methodologies, the correlations between separate preparations using the same sample (shown as solid lines) was higher than between different samples (shown as hatched lines). Analytical methods used in Figure 2, shown from left to right along the horizontal axis, are 1) with no quality filters either at a transcript level (gwg.mapq0.l2) or at a gene level (gwg.mapq0.gl) and 2) with a q20 filter at the transcript level (wg.mapq20.l2) and gene level (g.mapq20.l2gl). The next 2 analyses are based on the rpkm values which were supplied by the Life Technologies software at a transcript and gene level (cor.rpkm.l2 and cor.rpkm.l2gl). The analyses on the far right were aligned with TwoStep alignment (step2.rpkm.l2 and step2rpkm.l2gl).

For the above series of experiments, we prepared RNA-Seq libraries using Life Technologies' Ion Total RNA-Seq kits and following the vendor's recommendations.

Having optimized the library and sequencing procedures with cell lines and commercial samples we then prepared samples using the clinical biopsy samples from rectal cancer patients. RNAs were initially prepared using the MagMAX kit procedure from Life Technologies, which was the same method we had used to prepare the ILS and cell-line RNAs for our pilot studies. We attempted to prepare 4 rectal cases but only 2 of the samples yielded RNA. The 2 samples that yielded no RNA were from cases in which the biopsy samples were very small. Therefore, we changed our RNA preparation and used the RNeasy FFPE Kit from Qiagen. This procedure yielded 409 and 834 ng of RNA, which were good yields based on the amount of starting material; however, because these are precious clinical trial samples that will also be used for mutation analysis, it was necessary to also prepare DNA. Therefore, we prepared the next 2 samples with the AllPrep DNA/RNA FFPE Kit from Qiagen, with a slight modification. The xylene wash was eliminated because these small samples were easily lost during this wash step. The RNeasy and AllPrep kits are nearly identical but a pellet collected after protease digestion is thrown away in the RNeasy kit and is used in the AllPrep kit to make DNA. The modified AllPrep procedure was the best choice for RNA and DNA purification for these small rectal biopsy samples (Table 1).

During this report period we prepared RNA-Seq libraries using 2 rectal samples according to the procedure developed with the ILS samples; however, the mapping rate was lower than what we had seen in the ILS tissue and cell line samples (green circles, Figure 1). In the first run of the rectal sample the test fragment, which is a control to monitor the success of the sequencing reaction, showed very poor performance and could have indicated that the sequencing run was responsible for the low mapping rate of the rectal sample. Therefore, we repeated the sequencing

with the same library preparation; on this run, while the test fragment passed quality control, the mapping rate of the rectal sample did not improve. To improve the mapping rate, we adjusted the procedure used to purify the library and while our adjustment of the ethanol volume largely eliminated the smaller contaminating pieces of DNA within the library, it did not improve the mapping rate. To date, all of the RNA-Seq libraries have been prepared using all Life Technologies reagents and suggested procedures with only small modifications.

Additionally, to further improve our mapping rate we altered the RNA-Seq library preparation by amplifying and converting the RNA into cDNA with the SeqPlex RNA Amplification Kit for whole genome transcriptome amplification from Sigma-Aldrich. This kit was designed to be used with extremely small quantities of degraded/highly fragmented RNA. The SeqPlex kit provides a method to pre-amplify RNA samples while maintaining patterns of differential expression found in the unamplified sample. Once our RNA was amplified and converted to cDNA, then libraries were constructed with Life Technologies kit for the preparation of genomic DNA sequencing libraries. Specifically, the Ion Plus Fragment Library kit was used, but we eliminated the fragmentation step because the amplified cDNA was already of appropriate size for library preparation. We prepared 5 rectal samples using this procedure which were run in 3 different sequencing runs, (Figure 1, runs 2014-08-05, 2014-05-27, 2014-05-30). The mapping rate for the rectal samples using the SeqPlex RNA amplification shows a higher mapping rate than the library preparations constructed with the Life Technologies RNA-Seq library kit without any pre-amplification. Labels on Figure 1 designate the procedure (either Life Technologies or SeqPlex) and the types of samples that were used (cell line, ILS, or colorectal FFPET).

To assess the reproducibility of this library preparation that included the SeqPlex amplification, we prepared 3 rectal samples in two different preparations. Figure 3 shows that the concordance between different preps of the same sample was higher (0.94, 0.93, 0.92) than between different samples (0.86-0.89). The higher concordance of mapped reads within different preparations of the same sample versus between different samples indicates success of our RNA-Seq procedure.

Table 1 Summary of RNA Yields from Rectal Cancer Biopsies

Case no	sections scraped	Extraction Kit	total RNA yield	total DNA yield
KR1-0001	12	MagMax	745	na
KR1-0002	6	MagMax	0	na
KR1-0003	9	MagMax	0	na
KR1-0004	9	MagMax	330	na
KR1-0005	12	RNEasy	408.8	na
KR1-0006	9	RNEasy	834.4	na
KR1-0007	3	AllPrep no xylene	968.8	26
KR1-0008	3	AllPrep no xylene	548.8	21

Table 2 Correlation (R values) for Cell lines and ILS tissue Samples											
		ILS-22435_FF	ILS-22435_F	ILS-31043_0	ILS-31043_1	ILS-31043_01-03T19	ILS-31043_F	SW480	SW620	SW837	H747
	Run Date	2014-01-03T19	2014-01-07T18	2014-01-24T19	2014-01-24T19		2014-01-07T18	2014-02-05T18	2014-02-05T18	2014-02-05T18	2014-02-05T18
ILS-22435	2014-01-03T19	1.00	0.99	0.77	0.84	0.79	0.77	0.44	0.37	0.43	0.43
ILS-22435	2014-01-07T18	0.99	1.00	0.72	0.82	0.75	0.73	0.38	0.32	0.39	0.38
ILS-31043	2014-01-24T19	0.77	0.72	1.00	0.96	0.98	0.99	0.75	0.63	0.71	0.65
ILS-31043	2014-01-24T19	0.84	0.82	0.96	1.00	0.98	0.96	0.61	0.50	0.62	0.54
ILS-31043	2014-01-03T19	0.79	0.75	0.98	0.98	1.00	0.99	0.70	0.58	0.69	0.59
ILS-31043	2014-01-07T18	0.77	0.73	0.99	0.96	0.99	1.00	0.73	0.60	0.70	0.59
SW480	2014-02-05T18	0.44	0.38	0.75	0.61	0.70	0.73	1.00	0.93	0.91	0.90
SW620	2014-02-05T18	0.37	0.32	0.63	0.50	0.58	0.60	0.93	1.00	0.88	0.90
SW837	2014-02-05T18	0.43	0.39	0.71	0.62	0.69	0.70	0.91	0.88	1.00	0.91
H747	2014-02-05T18	0.43	0.38	0.65	0.54	0.59	0.59	0.90	0.90	0.91	1.00
Alignment was determined using the default iontorrent software.											

Figure 1

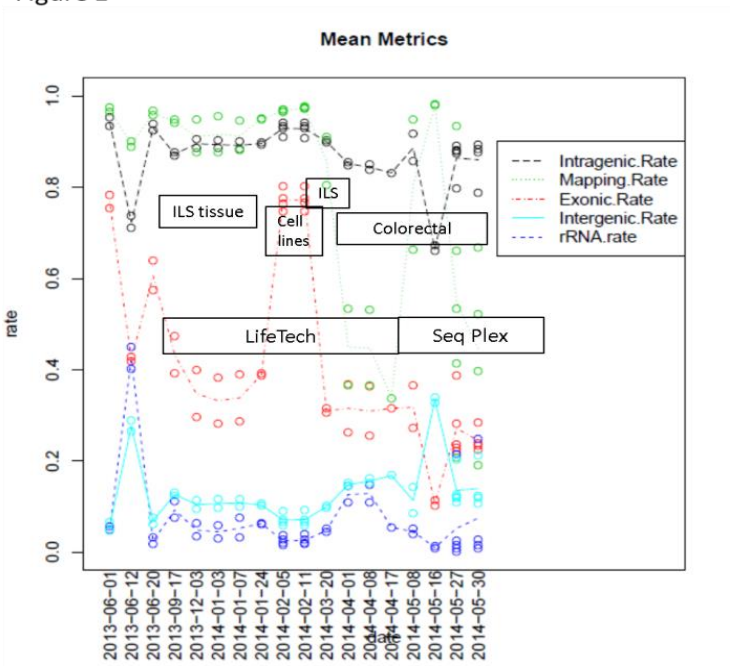

Figure 2 Concordance between and within Samples

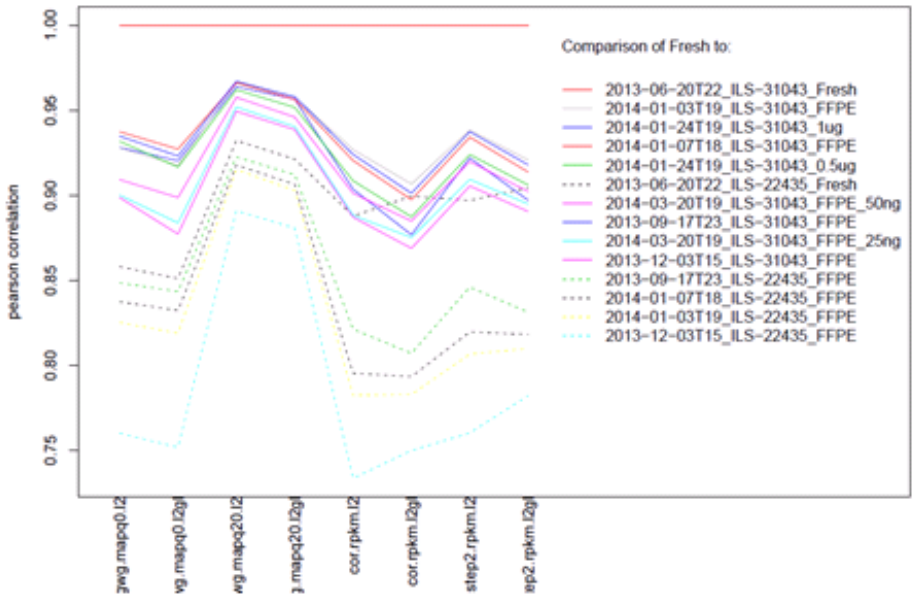
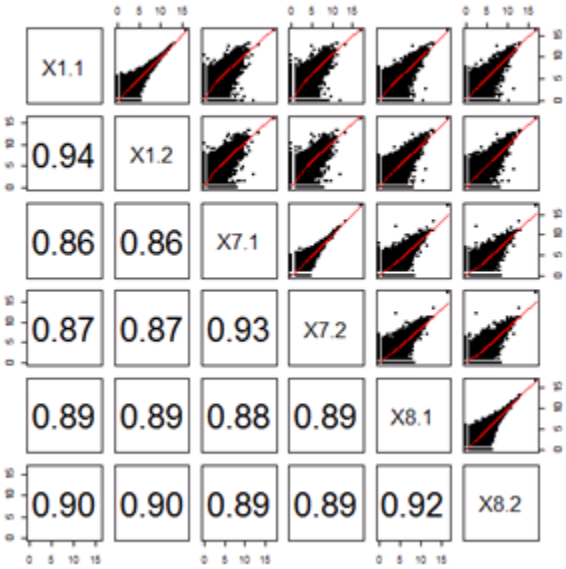
Fig. 3 Reproducibility of SeqPlex Method