

Mayo Clinic's Breast Cancer Translational Genomics (BCTG) Program
Progress Report
January 2009 - December 2010

Prepared by: Edith A. Perez, M.D.

The Mayo Clinic BCTG program started in Jan. 2009, with support from the 26.2 with Donna Foundation. The support allowed us to assemble a team of investigators that is already serving as a model for how to develop and perform translational research here at Mayo Clinic. I am pleased to provide a summary of the project, including goals, accomplishments, and next steps.

The overarching goal of the Breast Cancer Translational Genomics Program is to use state-of-the-art genomic technology to develop 21st century therapeutic approaches to the identification and treatment of breast cancer. Specifically, we intend to identify every genomic abnormality in tumors and develop therapeutic strategies based on the full range of genomic abnormalities, beyond just estrogen receptor, progesterone receptor and HER2. The goals of our program also include utilization of these genomic technologies to serve as infrastructure for investigators here at Mayo Clinic to conduct translational research related to breast cancer as well as pancreatic, lung, thyroid, renal, and colon cancer, amongst others.

Our short-term objectives for the first two years of the program, from 2009 to 2010, included development of the technology and building the infrastructure. At the time of initiation of the program in January 2009, we had a number of key resources available here at Mayo Clinic in Jacksonville. Those resources included a highly motivated clinical leadership team with the vision of developing state-of-the-art breast cancer genomics, a well-established cellular and molecular technology program with strong background in microarray-based genomic profiling, a developing breast tumor bank approved by the IRB, as well as access to a new Illumina genome analyzer, which came online in Rochester in November of 2008. As important as all of these factors, we had support from the 26.2 with Donna Foundation.

We knew that genomic research prior to 2009 was mainly dependent upon microarray, with its inherent limitations: it is hybridization-based technology, restricted to known genes, has low dynamic range, low signal-to-noise ratio, and is incapable of quantifying low abundance transcripts. We also knew that, although we lacked the infrastructure needed for state-of-the-art genomic analysis, we could optimize our genomics/translational and clinical expertise to develop a strong program. At the time, we also did not have full-time bioinformatic support in Jacksonville, although support was available through collaborative interactions with colleagues in Rochester. Back in 2009, we understood that these next-generation gene sequencers could provide us the answers we are looking for to help unravel the problem of breast cancer. This

was based on the understanding that this technology is 1) sequence based, 2) not restricted to known genes, 3) capable of providing an infinite, high dynamic range, 4) able to detect and quantify all messenger RNAs in the cell (including low-abundance transcripts, splice variants, fusion-gene products, expressed somatic mutations, and non-coding RNAs, including microRNAs), and 5) potentially able to analyze gene copy and total genomic promoter methylation patterns.

We assembled a group of investigators that became the translational breast cancer genomics team. This group of investigators spans the three Mayo sites, as well as the Illumina Corporation and T-Gen (in a smaller scale). The team is composed of clinicians, biologists, biostatisticians, bioinformatics specialists, computational biologists and database managers.

In terms of accomplishments and plans, even in this fairly new program, I am pleased to report that success is already evident. We published a manuscript in 2009 describing new protocols for complementary DNA library preparation, describing that 3' tag digital gene expression profiling with massive parallel sequencing technology achieved high sensitivity and reproducibility for transcriptome profiling compared to RNA-SEQ, but was more affordable and clearly outperformed microarrays in detecting lower abundant transcripts. We sequenced eight well-characterized breast cancer cell lines in 2009, and then concentrated our efforts to develop analytical tools to mine the vast amount of new data, consisting of 80 gigabyte base pairs of sequence generated.

Additionally, we developed several new analytical tools to mine sequence data on messenger RNA expression, DNA methylation and gene copy number. We just submitted a manuscript to PLoS One in November, 2010, describing deep-sequence analysis and the relationship between gene expression, DNA manipulation and gene copy number in breast cancer cells. In this manuscript we are describing a new epigenetic signature that defines estrogen-receptor positive over estrogen-receptor negative tumors. Our work has extended beyond identification of gene expression to the exciting area of development of analytical pipelines to identify and quantify fusion-gene mutations (in breast tumors). A new analytical pipeline, termed "Snowshoes FD," a tool for fusion transcript detection using paired end transcriptome sequencing, is the subject of a manuscript in preparation; we anticipate submitting in the first quarter of 2011, with Dr. Asmann as the lead author.

We have now identified and experimentally validated 56 gene mutations that have never been identified in the setting of breast cancer. This type of work will be the basis of a significant number of validation experiments over the next few years. Moreover, we have contributed preliminary data, analytical protocols and specific aims for five new grants totaling \$4.2 million obtained by investigators at Mayo Clinic in Jacksonville. These are all collaborative, translational research studies that involve evaluation of translational genomics of triple-negative breast

cancer, evaluation of protein kinase D as a marker and target for invasive breast cancer, atypical protein kinase signaling in lung cancer stem cells, microRNA in lobular invasion in breast cancer, as well as fusion-gene mutations as biomarkers of pancreatic cancer and lymph node metastasis.

In conclusion, we have already completed total messenger RNA sequencing of 24 primary breast tumors, 8 non-transformed human breast epithelial cells, 8 pairs of tumor-normal pancreatic tumors and 15 lung cancers. We have a manuscript published, one submitted, and two in preparation. In addition, we have two abstracts submitted and two in preparation.

Now, to our plans: The task for the next two years will be to apply the technology to determine the impact of these genomic discoveries related to patients and their outcome. We are currently mining the data to identify the key genomic features including differentially expressed messenger RNAs, spliced variants, novel expressed polymorphisms (mutations), and fusion-gene products. Additionally, we are preparing RNA and DNA for sequence analysis of 40 triple-negative, 40 HER2-positive and 20 ER-positive tumors. We are preparing RNA and DNA from a series of benign breast disease samples from women who are at high risk for breast cancer, or low risk for progression. As part of this discovery project, we are set to continue to evaluate the novel sequencing protocols to detect genomic rearrangements in tumor cells, and are developing protocols for total messenger RNA sequence analysis in paraffin-embedded samples.

The work planned for the next few years will allow us to mine the newly obtained genomic data to identify new biomarkers and therapeutic targets in triple-negative, HER2-positive, and estrogen-receptor-positive breast cancer. We will also work on identifying risk factors for breast cancer and foster collaborative work to identify predictors of lymph node metastases in pancreatic cancer. The data will help identify new mechanisms of transformation by evaluating the role of fusion-gene mutations as drivers in breast, lung and pancreatic cancer. Last, but not least, our work will help unravel new mechanisms of drug resistance in breast cancer. We anticipate that we will be able to use the deep-sequence genomic data to build predictive models of the clinical behavior of breast and other malignancies, and develop translational clinical trials based on these models.

The financial support from the 26.2 with Donna Foundation has made this program possible, with oversight and critical instrumentation provided by Mayo Clinic. We look forward to continued, careful stewardship of the money in the context of our visionary strategy to use and develop state-of-the-art genomic technology to develop 21st century therapeutic approaches to the development and treatment of breast cancer and other malignancies.