Project Title: A Bioinformatics Approach to Develop an Application for an Interactive Gene Co-Expression Network for Breast Cancer
Student Name: Catherine Ding
Mentor: Dr. Krishna Karuturi, Director of Computational Sciences at the Jackson Laboratory
Dates: Monday, January 8, 2018 to Friday, June 1, 2018

Summary/Abstract:
A gene co-expression network (GCN) is a valuable tool that can be used to for differential co-expression analysis by determining the differences in co-expression profiles of various sets of genes viewed in multiple contexts. These analyses can be useful in inferring potential disease genes such as those related to breast cancer. With the coding language R, I will be using the both the Shiny and igraph packages to build an interactive web application that will allow for network analysis based upon user input of total number of co-expression relationships, gene name, and context (estrogen receptors, p53, and grade).

Background and Significance:
The tumor suppressor gene TP53, which encodes the protein p53, regulates biological processes in order to maintain genome stability. In 80% of all cancer cases, a mutation in the this gene occurs. While the frequency of this mutation occurring in breast cancer is 20%, they are most common in the most aggressive breast cancer subgroup: triple negative breast cancer. Estrogen receptors serve as a useful prognostic marker to predict the course of breast cancer or any treatment targeting the disease.

Biological systems respond to either internal or external stimuli, which control regulation or expression pathways in genes. In one of the earliest works performed by Butte and Kohane, they proposed that highly mutual information between two genes suggests that they are non-randomly associated and are thus biologically related.

Gene co-expression networks are constructed from datasets obtained from microarrays or RNA sequencing (RNA-seq). In these networks, there are nodes or vertices that correspond to a specific gene, and edges that represent a certain correlation or relationship in co-expression profiles. Though edges are undirected and thus do not prove causality, they are valuable in elucidating whether clusters of genes with similar co-expression patterns are controlled by the same transcriptional regulatory program, functionally related, or members of the same pathway or protein complex. Using the guilt-by-association (GBA) approach, potential disease genes may be identified if they are that are co-expressed with other disease genes. By identifying the genes that play an important role in breast cancer, an effective treatment may be more efficiently developed.

Experimental Approach:
I will be using R packages Shiny and igraph to build an interactive web application of a gene co-expression network. Using differentially co-expressed gene sets, the network will contain a node for each unique gene and distinct edge colors that correspond to positive or negative coefficients of estrogen receptors, the p53 protein, and tumor grade.

I will first debug the currently existing code and later add new features. In the server and user interface scripts, I will code for a slider that takes an integer as user input. This number will determine the genes displayed at one time based upon whether the total number of relationships one gene has with other genes is above or below this integer. Another feature will be a search bar for a gene name that will display all the relationships of that one gene. Lastly I will create tick boxes for each one of the three contexts (ER, p53, grade).

References Cited:

1. Herty Liany, Jagath C. Rajapakse, and R. Krishna Murthy Karuturi. “MultiDCoX: Multi-factor analysis of differential co-expression.” *BMC Bioinformatics*. The Author(s). 2017, published 28 December 2017. Web. Accessed 10 Feb 2018.
2. Sipko van Dam, Urmo Võsa, Adriaan van der Graaf, Lude Franke, João Pedro de Magalhães. “Gene co-expression analysis for functional classification and gene–disease predictions.” *Oxford Academic: Briefing in Bioinformatics.* N.p. Published 10 January 2017. Web. Accessed 10 Feb 2018.
3. A. J. Butte, and I. S. Kohane. “Mutual Information Relevance Networks: Functional Genomic Clustering Using Pairwise Entropy Measurements.” *Pacific Symposium on Biocomputing*., U.S. National Library of Medicine, published 2000. Web. Accessed 11. Feb 2018.
4. C. Berger, Y. Qian, and X. Chen. “The p53-Estrogen Receptor Loop in Cancer.” *Current Molecular Medicine*, U.S. National Library of Medicine, published Sept. 2013. Web. Accessed Feb 11 2018.
5. Sommer, S, and S A Fuqua. “Estrogen Receptor and Breast Cancer.” *Seminars in Cancer Biology*., U.S. National Library of Medicine, published Oct. 2001. Web. Accessed Feb 11 2018.