Detecting Aberrant Expression in Breast Cancer through Analysis of miRNA Microarrays
Shahrzad Eslamian, Gunter Tusch (PhD)
Medical and Bioinformatics Graduate Program
Grand Valley State University, Allendale, MI, USA

Introduction

Several studies have now shown that miRNAs are involved in the initiation and progression of cancer. miRNA expression is dysregulated in cancer by a variety of mechanisms including amplification, deletion, and epigenetic silence mutation. MicroRNAs usually make gene silencing by binding to the targeted mRNA. This interaction avoids protein production by initiating mRNA degradation. Stopping the NMD pathway will stabilize mutations like frame-shift or nonsense that shorten the protein products, this can be noticed using gene expression microarrays. The main purpose of the current research is to find out the specific DE-miRNA between all the microRNAs and be able to analyze the pathways of their target genes in order to find out the connection between the dysregulator miRNA and breast cancer. Datasets from GEO series “GSE37210”, which include non BRCA1/2 breast cancer caused by frameshift or nonsense mutations that changed the protein products. The significant genes analyzed between patients and controls in a family.

The miRNA expression profile of GSE37210 accession number was downloaded from the GEO database, which was collected by Johnson JK et al. This study tried to use breast cancer families in order to identify further high-risk breast cancer liability genes by using caffeine as a treatment agent.

Methodology

In the current research the expression profiles of miRNAs in 15 samples from 5 untreated patients with breast cancer and 15 samples from 5 healthy control subjects, which all belonged to family A, were chosen for the study. The series matrix text file of datasets was downloaded from GEO. The GEO series accession was GSE37210. This experiment was implemented on three different families. The chosen samples were from family A that had GIN1 methods implemented on them.

Their cell lines were lymphoblastoid and they were not treated with any caffeine agents. Gene series GSE37210 was downloaded from GEO, which was collected by Johnson JK et al (2015). This data set tried to identify novel breast cancer susceptibility genes by applying additional strategies beyond customary methods. Almost one third of congenital genetic diseases are initiated by frameshift or nonsense mutations that truncate the protein products.

For this purpose, the dysregulated genes between 5 cancer patients and 5 healthy control subjects in a family should be identified and each sample had 3 different representative data. The data consisted of 48804 genes and 30 samples fifteen samples from five individuals of Family A which were affected and not treated were between GSM913745 and GSM913759. Fifteen samples from five individuals of Family A which were healthy were between GSM913760 and GSM913774. For statistical analysis, in order to identify the dysregulator genes between cancer patients and healthy control subjects, since there are 48804 genes and 30 samples, multiple testing procedures on SAS were applied on the data and instead of raw P, the adjustment P was considered. After attempting multitest the data which had P ≤0.001, False Discovery Rate <0.01 and Bonferroni <0.05 were considered to indicate statistically significant differences.

Acknowledgements
I thank Dr. Johnathon Leidig for his guidance in visualization methods using Prefuse.

DAVID was used to discover the function of genes within the sections. Mirfocus intends to analyze miRNA-target gene pathways and the related miRNA annotations. The Mirbase also is a database which aims to provide integrated interfaces for comprehensive microRNA sequence data, annotation and predicted gene targets. By using Mirbase and Mirfocus 30 target genes from 52 miRNA were identified. KEGG database was used to discover the pathways of the products of these DE-miRNAs and their connection to non-BRCA1/2 breast cancer.

Result & Visualization

A total of 52 DE-miRNAs and 30 related genes were identified. In breast cancer most of these DE-miRNA were up regulated or down regulated. Among these miRNAs, micro RNAs let-7a-5p, miR-192-5p, miR-335-5p and miR-155 draw specific attention. Databases revealed that the target genes of miR-155 especially snx6 play crucial role in breast cancer.

To analyze the data, make them more clear, and easy to access, effective visualization can be helpful. The software toolkit that was used for building graphical user interfaces is Prefuse. Prefuse is an extensible user interface toolkit for visualizing both structured and unstructured data.

References:

Conclusion

By analyzing the results, this study reached some of the answers regarding the responsible genes and pathways which are related to breast cancer, but there should be still more research on the pathways.