

### **Analysis of Microarray Data**

Perform a statistical analysis of a large microarray data set with a goal of identifying gene products that are possible targets in breast cancer. This data set is obtained from nine females. Three individuals have been diagnosed with breast carcinomas (BC1), three individuals have clinically normal breasts (BN), and three individuals served as a source of heart tissue (HN). In this set, the expression levels of 8180 mRNA fragments is measured.

For full credit, you should perform at least two of the four analysis discussed in the tutorial (scatter plot of means, ratio of means, difference in means, and the Volcano plot). From each analysis method, identify at least ten proteins that show abnormally high expression levels in the breast carcinoma as opposed to normal breast tissue. For each method used, identify the protein that shows the highest abnormality, and briefly discuss the biological role of this protein. Assess if this protein is likely to play a role in breast cancer, and if it would make a suitable target for treatment.

Create a subset of at least 12 proteins that, based on your analysis, are significantly expressed in the breast carcinomas and generate a heatmap showing their expression levels in all nine samples. Identify at least one protein from this data set that would be a good target for the treatment of breast cancer. Recall that a good target is one that has a high expression level in the cancer tissue, that is overexpressed significantly in breast tumors, and that is expressed at low levels in the heart. Perform a literature search about this target and summarize what is known about the role of this protein in breast cancer.