**Project Description**

Project: (Personalized Medicine) A Priori Predicting Reaction to FOLFOX Therapy in Patient with Colorectal Cancer using Random Forests/Decision Trees/Support Vector Machines

**Materials.** You will need to obtain the gene express dataset, GSE32536, from Gene Expression Omnibus (GEO). Also, review the associated paper [Tsuji

2012].

**Tasks.** Complete the following tasks.

Differentially Expressed Genes (DEGs).

* Run script to generate p-values for all genes.
* Open results in Excel and sort by p-value.
* Create a heat map of the top 25 DEGs (that is, genes with the lowest *p*-value)

Training & Testing

* Select the top 25 genes.
* Use Weka Attributes to further reduce the number of genes to 25? Or use multiple attribute selectors and generate a Venn diagram and select the genes that intersect?
* Randomize the dataset; create two files (2/3 for training; 1/3 for testing). Optionally, allow the algorithm you use to automatically divide the data into training and testing using cross validation.
* Convert the data into the correct format for Weka or R. Perform classification using J48 Decision Tree or Random Forests or another of your choice. If using a decision tree, retain the image.

Data Description

* Create a bar plot for CR, PR, PD, and SD (see the reference paper in sections Materials & Methods and Results).
* Create a Venn diagram?

Evaluation

* Create a confusion matrix of TP, TN, FP, FN
* Calculate Sensitivity and Specificity
* Calculate Precision, Recall, and F-score