**Analysis Planning Worksheet**

**Evaluation Question**

Using dataset from this link: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE70770>

Can I use the 29 genes to classify good prognostic biochemical recurrent cases from bad prognostic biochemical recurrent cases?

Here are the 29 genes:

MYBPC1, SLC7A4, CDC42EP5, BCAS1, PAK1IP1, ANPEP, SLC23A1, HS.270778, GLB1L2, GCNT2, TNFRSF19, CUX2, POTEG, NCAPD3, LOC728606, SLC22A3, C9ORF61, SRD5A2, FLJ31568, SERPINF2, PGM5, LOC645993, ZDHHC8P, CNTNAP2, F12, CXCL14, CTHRC1, ZNF467, CDC20

**Independent Variable(s)**

These variable(s) are causing something or creating an effect. List what each is and whether it is categorical or continuous. It is ok to only have one.

**Variable**

Time to biochemical recurrence

□ Categorical: # of levels: 2 □ Continuous

**Variable**

□ Categorical : # of levels \_\_\_\_\_ □ Continuous

**Variable**

□ Categorical: # of levels \_\_\_\_\_ □ Continuous

Variable

□ Categorical: # of levels \_\_\_\_\_ □ Continuous

Dependent Variable(s)

These variable(s) are influenced by your independent variable and *depend* on them. List what each is and whether it is categorical or continuous. Unless they are related, you should have only one.

**Variable**

Gene expression values

□ Categorical: # of levels \_\_\_\_\_ □ Continuous

**Variable**

□ Categorical: # of levels \_\_\_\_\_ □ Continuous

**Variable**

□ Categorical: # of levels \_\_\_\_\_ □ Continuous

Variable

□ Categorical: # of levels \_\_\_\_\_ □ Continuous

Now that you know the type and number of independent and dependent variables, you are ready to use the analysis flow charts to choose your analysis!

**Analysis:**

**Generalized linear model followed by step-forward and step-backward parameter inclusion. Lastly, conduct ROC analysis to assess the accuracy, sensitivity, and specificity of the model.**