

Competition: Gene expression as a diagnostic tool

STK4030/9030 - Fall 2015

November 3, 2015

In Nature Medicine, Spira et al. (2007) *Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer* presented a predictor or diagnostic tool for lung cancer based on profiles of gene expression (a measure of how active a gene is in a specific tissue sample). They write the following:

Lung cancer is the leading cause of death from cancer in the US and the world... A lung cancer diagnostic based on gene expression in large-airway epithelial cells might eliminate the need for additional diagnostic tests that are costly, incur risk and prolong the diagnostic evaluation of suspected lung cancer patients.

The data set `epithelial.RData` contains the expression of 22 215 genes in 128 subjects as a training set, recorded in `trainExpression`, and 35 subjects as a test set, recorded in `testExpression`. In the training samples there are 60 cases with lung cancer and 68 without lung cancer, recorded as `Cancer` and `No Cancer` in `trainStatus`. You can assume that the distribution of the test cases will be roughly as for the training data.

Select your favorite method to construct a classifier for the cancer status, for instance:

- Penalized logistic regression
- A boosted model
- Regularized LDA or QDA
- Classification tree
- Random forests

Use your selected classifier to predict the cancer status of the 35 samples in the test set. To enter a contribution to the competition, write your predictions in the table on the next page and deliver it to the postbox of Kristoffer Hellton at the Mathematics department post room, 7th floor before the **Friday 20th of November**.

Your predictions will be compared to the truth and evaluated using the 1-0 loss (misclassification), and the student achieving the highest prediction accuracy will win a prize!

The contribution of to the STK4030/STK9030
Prediction Competition. Mark predictions as either **Cancer** and **NoCancer**.

[illegible]