**This folder contains three R scripts to run statistical tests for generating predictors for colorectal cancer based on PRM data.**

**Required software**

The scripts use core R functionality and several publicly R available packages listed below. Version numbers in brackets correspond to the versions of the packages that were used to develop and debug these scripts:

* **R** (4.0.3)
* **RStudio** (1.2.5033) - optional, testing functions and running the code step-by-step
* **reshape2** (1.4.4) - robustly and flexibly reshape data
* **randomForest** (4.6-14) – Breiman and Cutler’s random forests analysis for classification and regression
* **pROC** (1.16.2) – display and analyze ROC curves
* **ggplot2** (3.3.2) - generation heatmap plots

**Running the scripts**

1. Change path of the data folder on the top of the script.
2. Lauch the script. You can run the eitire script at once or execue commends one by one to see the intermediate outputs.
3. Output results are saved as *.csv* (tables) and *pdf* (plots) files.
4. **Script for correlation matrix of protein expression abundance to measure the intercorrelation between peptides：**

PRM corr matrix.R

* Data import (using the 23 upregulated proteins from 82 samples listed in Supplementary Table 3b)
* Compute the correlation matrix
* Get upper triangle of the correlation matrix
* Reorder the correlation matrix
* Melt the correlation matrix
* Create a ggheatmap and visualization

1. **Script for random forests analysis**

PRM random forests.R

* Data import (using the 14 upregulated proteins from 82 samples that filtered by the above correlation analysis)
* Create the forest
* View the forest results
* Look at variable importance based on mean decrease in accuracy index
* 100 random forest replicates of Mean Decrease Accuracy to minimize randomness

1. **Script for ROC analysis that combined any two proteins**

pROC combine 2 proteins corr matrix.R

* Data import (using the 14 upregulated proteins from 82 samples that filtered by the above correlation analysis)
* Calculate AUC values of single protein (generate Fig. 3c)
* Calculate AUC values of the combination of any two proteins
* Prepare for drawing corr. Matrix
* Get upper triangle of the correlation matrix
* Use correlation between variables as distance
* Reorder the correlation matrix
* Melt the correlation matrix
* Create a heatmap and add correlation coefficients on the heatmap