Untitled

My Name

4/14/2020

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

summary(cars)

## speed dist   
## Min. : 4.0 Min. : 2.00   
## 1st Qu.:12.0 1st Qu.: 26.00   
## Median :15.0 Median : 36.00   
## Mean :15.4 Mean : 42.98   
## 3rd Qu.:19.0 3rd Qu.: 56.00   
## Max. :25.0 Max. :120.00

## Including Plots

You can also embed plots, for example:



Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.

## Tutorial on gene assignment

We illustrate with an example. Tannous et al. (2020) presented abundance levels of proteins among six fractions from a differential centrifugation experiment: N, M, L1, L2, P, and S, and three fractions from a Nycodenz density gradient separation of the differential fraction L1 (Nyc1, Nyc2, and Nyc3). Eight subcellular compartments were considered: nucleus, mitochondria, lysosomes, peroxisomes, Golgi apparatus, plasma membrane, and cytosol. The CPA method assigns each of a large number of genes to one or more of these compartments, based on profiles from a set of reference genes.

## Installation

Start by installing the devtools package from CRAN, by typing

install.packages("devtools")

Then install the protlocassign package from the github repository by typing

devtools::install\_github("mooredf22/protlocassign0p1p1")

In this example we will work with the TMT ms2 data from Tannous et al. (2020). To get started, load the package and look at the dimensions of the data set and then the first few rows:

library(protlocassign)

## Loading required package: lme4

## Loading required package: Matrix

## Loading required package: BB

dim(geneProfileSummaryTMTms2)

## [1] 7893 12

options(digits=3)  
head(geneProfileSummaryTMTms2)

## geneName N M L1 L2 P S Nyc.1 Nyc.2  
## 1 2900026A02RIK 0.1025 0.0851 0.1125 0.0930 0.1051 0.1254 0.1524 0.1484  
## 2 A1CF 0.1577 0.0805 0.0765 0.1264 0.2554 0.1118 0.0657 0.0634  
## 3 A930018M24RIK 0.0544 0.0503 0.1129 0.0760 0.0399 0.0507 0.1358 0.3969  
## 4 AAAS 0.3692 0.0388 0.0619 0.0995 0.2070 0.0336 0.0709 0.0699  
## 5 AABR07001519.1 0.0579 0.0528 0.0732 0.0687 0.0504 0.3653 0.1381 0.1431  
## 6 AABR07002683.1 0.0720 0.0749 0.1059 0.1189 0.1090 0.0787 0.2232 0.1334  
## Nyc.3 Nspectra Nseq  
## 1 0.0758 4 3  
## 2 0.0624 25 16  
## 3 0.0830 1 1  
## 4 0.0492 3 3  
## 5 0.0505 1 1  
## 6 0.0839 2 2