ClustermRNA

David Kaplan

12/07/2021

```
setwd('/Users/davidkaplan/Desktop/Ewing/Ahood')
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyr)
library(tibble)
library(cluster)
library(NbClust)
library(factoextra)
```

```
## Loading required package: ggplot2
```

```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

```
library(dendextend)
```

```
##
## ---------------------
## Welcome to dendextend version 1.15.1
## Type citation('dendextend') for how to cite the package.
##
## Type browseVignettes(package = 'dendextend') for the package vignette.
## The github page is: https://github.com/talgalili/dendextend/
##
## Suggestions and bug-reports can be submitted at: https://github.com/talgalili/dendextend/issues
## Or contact: <tal.galili@gmail.com>
##
##  To suppress this message use:  suppressPackageStartupMessages(library(dendextend))
## ---------------------
```

```
##
## Attaching package: 'dendextend'
```

```
## The following object is masked from 'package:stats':
##
##     cutree
```

Loading dataset files

```
mRNA <- read.csv('CCLE_expression.csv')
sample_info <- read.csv('sample_info.csv')
```

Filtering sample_info to only retrieve DepMap_ID of colorectal cancer lineages

```
target_diseases <- filter(
  .data=sample_info,
  primary_disease == 'Colon/Colorectal Cancer'
)
```

Applying the filter in the CCLE expression dataset

```
mRNA_colon <- mRNA %>%
  filter(X %in% target_diseases$DepMap_ID)
transformed_mRNA <- mRNA_colon[,names(mRNA_colon)]
rownames(transformed_mRNA) <- transformed_mRNA$X
transformed_mRNA <- transformed_mRNA[,-1]
```

Running through hierarchical clustering, cutting the data into 70 clusters.

```
transformed_mRNA <- t(transformed_mRNA)
distancemRNA <- dist(transformed_mRNA,method='maximum')
clustermRNA <- hclust(distancemRNA, method='ward.D')
clustermRNAs = cutree(clustermRNA,k=70)
clustermRNAs <- as.data.frame(clustermRNAs)

clustermRNAs <- cbind(gene = rownames(clustermRNAs), clustermRNAs)
clustermRNAs$gene <- sub("\\..*","",clustermRNAs$gene)
```