

R Notebook

Here, we want to take the sup. tables from Vemulapalli et al (<https://www.biorxiv.org/content/10.1101/598664v2.full>) and group proteins by their change in phosphorylation

```
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.2.1 --
## v ggplot2 3.2.1      v purrr  0.3.3
## v tibble  2.1.3      v dplyr  0.8.3
## v tidyr   1.0.0      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.4.0

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

Read in the data and keep sites quantified in both reps

readRepData <- function(sheet, pid="Protein Id"){
  x <- readxl::read_excel('../raw/Vemulapalli_sup.xlsx', sheet=sheet)

  x %>% separate(!!(pid), into=c('sp', 'UniprotAc', 'name')) %>%
    select(UniprotAc, `Protein description`, Motif, DMSO_0, DMSO_5, DMSO_10, DMSO_30)
}

rep1 <- readRepData(1)

## Warning: Expected 3 pieces. Additional pieces discarded in 497 rows [1, 2, 3, 4,
## 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].

rep2 <- readRepData(2, pid="Protein ID")

## Warning: Expected 3 pieces. Additional pieces discarded in 748 rows [1, 2, 3, 4,
## 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].

both_reps <- intersect(rep1$Motif, rep2$Motif)
print(length(both_reps))

## [1] 294

rep1 <- rep1 %>% filter(Motif %in% both_reps)
rep2 <- rep2 %>% filter(Motif %in% both_reps)

Identify the maximum changes in phosphorylation for each replicate

getMaxDiff <- function(obj){

  diffs <- obj %>%
    mutate(diff5=DMSO_5-DMSO_0, diff10=DMSO_10-DMSO_0, diff30=DMSO_30-DMSO_0) %>%
    select(diff5, diff10, diff30)

  return(diffs %>% apply(MARGIN=1, function(x) x[which(abs(x)==max(abs(x)))]))
}

rep1$max_diff <- getMaxDiff(rep1)
rep2$max_diff <- getMaxDiff(rep2)
```

Merge the replicates and obtain the average change in phosphorylation

```
combined_replicates <- rep1 %>% merge(rep2, by=c('Motif', 'UniprotAc', 'Protein description')) %>%
  rowwise() %>%
  mutate(ave_max_diff=mean(max_diff.x, max_diff.y)) %>%
  select(UniprotAc, Motif, `Protein description`, max_diff.x, max_diff.y, ave_max_diff) %>%
  arrange(desc(ave_max_diff))
```

```
print(head(combined_replicates))
```

```
## Source: local data frame [6 x 6]
```

```
## Groups: <by row>
```

```
##
```

```
## # A tibble: 6 x 6
```

##	UniprotAc	Motif	`Protein description`	max_diff.x	max_diff.y	ave_max_diff
##	<chr>	<chr>	<chr>	<dbl>	<dbl>	<dbl>
## 1	Q9UQC2	SDEKVD~	GAB2_HUMAN GRB2-associat~	23.5	20.3	23.5
## 2	Q9Y6I3	KNIVHN~	EPN1_HUMAN Isoform 2 of ~	19.8	15.0	19.8
## 3	Q13480	ADERVD~	GAB1_HUMAN Isoform 2 of ~	15.9	13.4	15.9
## 4	P19174	GSFESR~	PLCG1_HUMAN Isoform 2 of ~	14.3	3.13	14.3
## 5	Q9NZM3	YQNKLI~	ITSN2_HUMAN Intersectin-2	14.1	14.4	14.1
## 6	Q13191	IPDLSI~	CBLB_HUMAN E3 ubiquitin~	14.1	5.93	14.1

```
print(tail(combined_replicates))
```

```
## Source: local data frame [6 x 6]
```

```
## Groups: <by row>
```

```
##
```

```
## # A tibble: 6 x 6
```

##	UniprotAc	Motif	`Protein description`	max_diff.x	max_diff.y	ave_max_diff
##	<chr>	<chr>	<chr>	<dbl>	<dbl>	<dbl>
## 1	Q05655	SEPVGIY~	KPCD_HUMAN Protein kina~	-13.2	-17.2	-13.2
## 2	Q13113	SEHENAY~	PDZ1I_HUMAN PDZK1-inter~	-13.5	-5.35	-13.5
## 3	P16070	NGEASKS~	CD44_HUMAN CD44 antigen	-14.4	-13.0	-14.4
## 4	P07947	LIEDNEY~	YES_HUMAN Tyrosine-prot~	-15.3	-23.8	-15.3
## 5	Q8N3R9	ITDERVY~	MPP5_HUMAN MAGUK p55 su~	-15.5	-20.4	-15.5
## 6	Q9H5V8	ESESEPY~	CDCP1_HUMAN CUB domain~	-17.6	-33.6	-17.6