Customized Tool to visualize protein coverage after Maxquant search for MS data

Wael Kamel

Tool to show protein coverage across different samples/condition, using the Maxquant output file "peptdies.txt"

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 3.5.2
```

```
## -- Attaching packages ---------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.0.5     v dplyr   0.8.5
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0
```

```
## Warning: package 'dplyr' was built under R version 3.5.2
```

```
## Warning: package 'stringr' was built under R version 3.5.2
```

```
## Warning: package 'forcats' was built under R version 3.5.2
```

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

```
library(reshape2)
```

```
##
## Attaching package: 'reshape2'
```

```
## The following object is masked from 'package:tidyr':
##
##     smiths
```

```
library(scales)
```

```
##
## Attaching package: 'scales'
```

```
## The following object is masked from 'package:purrr':
##
##     discard

## The following object is masked from 'package:readr':
##
##     col_factor
```

```
library(splitstackshape)
```

```
## Warning: package 'splitstackshape' was built under R version 3.5.2
```

```
peptides<- as.data.frame(read.delim("peptides.txt", sep="\t",header=TRUE))

names(peptides)
```

```
##  [1] "Sequence"               "N.term.cleavage.window" "C.term.cleavage.window"
##  [4] "Amino.acid.before"      "First.amino.acid"       "Second.amino.acid"
##  [7] "Second.last.amino.acid" "Last.amino.acid"        "Amino.acid.after"
## [10] "A.Count"                "R.Count"                "N.Count"
## [13] "D.Count"                "C.Count"                "Q.Count"
## [16] "E.Count"                "G.Count"                "H.Count"
## [19] "I.Count"                "L.Count"                "K.Count"
## [22] "M.Count"                "F.Count"                "P.Count"
## [25] "S.Count"                "T.Count"                "W.Count"
## [28] "Y.Count"                "V.Count"                "U.Count"
## [31] "O.Count"                "Length"                 "Missed.cleavages"
## [34] "Mass"                   "Proteins"               "Leading.razor.protein"
## [37] "Start.position"         "End.position"           "Gene.names"
## [40] "Protein.names"          "Unique..Groups."        "Unique..Proteins."
## [43] "Charges"                "PEP"                    "Score"
## [46] "Fraction.Average"       "Fraction.Std..Dev."     "Fraction.1"
## [49] "Experiment.1"           "Experiment.2"           "Experiment.3"
## [52] "Experiment.4"           "Experiment.5"           "Experiment.6"
## [55] "Intensity"              "Intensity.1"            "Intensity.2"
## [58] "Intensity.3"            "Intensity.4"            "Intensity.5"
## [61] "Intensity.6"            "Reverse"                "Potential.contaminant"
## [64] "id"                     "Protein.group.IDs"      "Mod..peptide.IDs"
## [67] "Evidence.IDs"           "MS.MS.IDs"              "Best.MS.MS"
## [70] "Oxidation..M..site.IDs" "MS.MS.Count"
```

Select the following columns into new dataframe, "Sequence", "Gene.names", "Start.position", "End.position" and the samples intensities (in this case will be columns between 56:61)

```
select_col_peptides<- select(peptides, 1, 39,37,38,56:61)

names(select_col_peptides)
```

```
##  [1] "Sequence"       "Gene.names"     "Start.position" "End.position"
##  [5] "Intensity.1"    "Intensity.2"    "Intensity.3"    "Intensity.4"
##  [9] "Intensity.5"    "Intensity.6"
```

Clean Up and Melting the dataframe

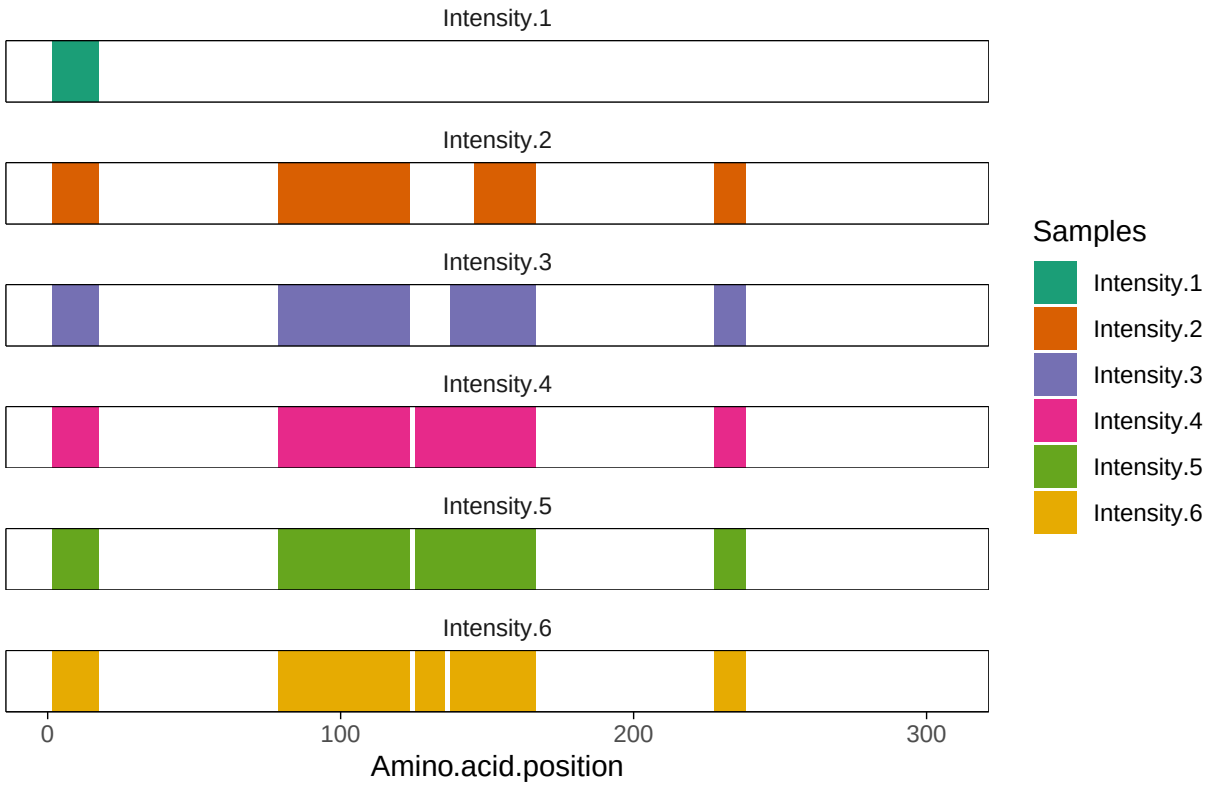
```
select_col_peptides[select_col_peptides == 0] <- NA
select_col_peptides[, 5:10] <- log(select_col_peptides[,5:10], 2)
select_col_peptides_melt<- melt(select_col_peptides,id.vars = c("Sequence","Gene.names", "Start.positio
```

```
#Calculate peptide lenght
select_col_peptides_melt$peptide_length <- (select_col_peptides_melt$End.position -select_col_peptides_m

#remove peptides witgh NA intensities
select_col_peptides_melt<- subset(select_col_peptides_melt, Log2_Intesnsity>0  )

#Since we are only mapping true or false coverage, we remove redundant peptide
select_col_peptides_melt$duplication_mark<-paste(select_col_peptides_melt$Samples,select_col_peptides_me
select_col_peptides_melt<-select_col_peptides_melt[!duplicated(select_col_peptides_melt$duplication_mark
#setting the bar height for visulization
select_col_peptides_melt$Bar_length = 1

#Finally breakdown the peptide into individula aa with corresponding corrdinates

select_col_peptides_melt= select_col_peptides_melt[complete.cases(select_col_peptides_melt), ]

select_col_peptides_melt_aa = select_col_peptides_melt %>%
group_by( Sequence,Gene.names,Samples, Start.position  ) %>%
expandRows("peptide_length") %>%
mutate(Amino.acid.position = Start.position - 1 + cumsum(Bar_length))
```

Visuliza the protein coverage for any protein of interest (insert Gene name in Gene.names==“XX” , in the code below)

```
ggplot( subset(select_col_peptides_melt_aa, Gene.names=="PABPN1" ), aes(x=Amino.acid.position,y= Bar_le
  geom_bar(stat = "identity", position = "fill")+
  facet_wrap(~Samples,ncol = 1)+
  #below you can set the full lenght of the protein of interest
  xlim(1, 306)+
  scale_y_continuous(expand = c(0,0))+
  scale_fill_brewer(palette="Dark2")+
  theme_classic()+
  labs(title = "PABPN1")+
  theme(axis.title.y=element_blank(),
        axis.text.y=element_blank(),
        axis.ticks.y=element_blank(),
        axis.line = element_line(colour = 'black', size = 0.25),
        axis.ticks = element_line(colour = 'black', size = 0.25),
        strip.background  = element_blank())+
  annotate("segment", x=-Inf, xend=Inf, y=-Inf, yend=-Inf, color="black",size=0.25, linetype="solid")+
  annotate("segment", x=-Inf, xend=Inf, y=Inf, yend=Inf, color="black",size=0.25, linetype="solid")+
  annotate("segment", x=Inf, xend=Inf, y=-Inf, yend=Inf, color="black",size=0.25, linetype="solid")
```

PABPN1


Intensity.1
Intensity.2
Intensity.3
Intensity.4
Intensity.5
Intensity.6
0
100
200
300
Amino.acid.position
Samples
Intensity.1
Intensity.2
Intensity.3
Intensity.4
Intensity.5
Intensity.6