

Informe PEC1 - Anàlisi òmics

Marina Ribes Santolaria

2024-11-06

Taula de continguts

1	ABSTRACT	1
2	OBJECTIUS DE L'ESTUDI	1
3	MATERIALS I MÈTODES	2
4	RESULTATS	2
4.1	Creació del SummarizedExperiment	4
4.2	Anàlisi del SummarizedExperiment creat	8
5	DISCUSSIÓ, LIMITACIONS I CONCLUSIONS DE L'ESTUDI	29
URL REPOSITORI GITHUB : https://github.com/marinaribessantolaria/Ribes-Santolaria-Marina-PEC1		

1 ABSTRACT

La fosforilació és una modificació posttraduccional que juga un rol essencial en la senyalització cel·lular i la regulació de processos cel·lulars. Per aquest motiu la comparativa de pèptids fosforilats resulta d'un ampli interès, ja que pot revelar diferències clau en la funció i estat d'una proteïna en condicions diferents. Per aquest motiu, en aquest treball s'han interpretat diferències de perfils fosfoprotèics comparant dos fenotips de xenografts derivats de pacients (PDX) (fenotip MSS i fenotip PD). L'estudi estadístic de les dades proporcionades suggereix que hi ha diferències en el perfil fosfopeptídic segons el fenotip i entre aquestes diferències destaquem els pèptids P35269 i O00193, expressats diferencialment en les mostres MSS respecte a les PD.

2 OBJECTIUS DE L'ESTUDI

L'objectiu d'aquest estudi és:

- L'ús de SummarizedExperiments per tal d'estudiar i descriure les dades de fosfoproteòmica.
- Determinar si hi ha diferències entre mostres de dos fenotips diferents (MSS o PD) de PDX pel que fa al perfil de fosfopèptids enriquits en cada mostra.

3 MATERIALS I MÈTODES

En aquesta activitat es treballa amb les dades proporcionades en el repositori de github: <https://github.com/nutrimetabolomics/metaboData/>. Més concretament, s'ha seleccionat el dataset de 2018-Phosphoproteomics.

Les dades contingudes en aquest dataset es tracten de 3 mostres de model PDX de fenotip MSS (mostres M1, M5 i T49) i 3 mostres de fenotip PD (M42, M43 i M64), amb un replicat tècnic de cada mostra. La forma en què s'han obtingut les dades és mitjançant l'extirpació del PDX del ratolí, l'enriquiment de fosfopèptids en les mostres i finalment l'anàlisi mitjançant LC-MS. D'aquesta manera s'ha obtingut les seqüències modificades obtingudes (accessió) i el seu valor d'abundància per cada mostra i rèplica.

Podem visualitzar un resum del procediment que s'ha seguit en aquest diagrama que s'inclou en el repositori:

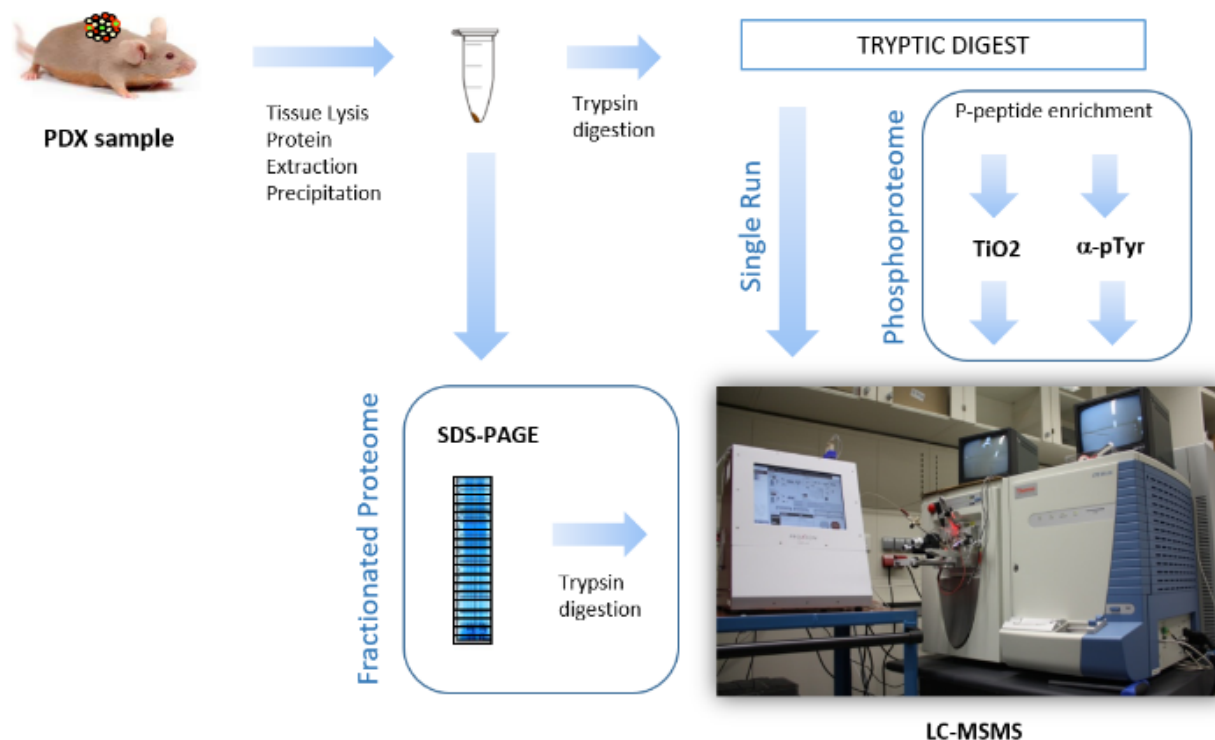


Figure 1: A phosphoproteomics experiment

Respecte al processament de dades, s'ha utilitzat el programa RStudio i les eines Bioconstructor i SummarizedExperiments pel processament de dades. El projecte creat es troba sota control de versions gràcies a l'eina Git i el repositori Github (<https://github.com/marinaribessantolaria/Ribes-Santolaria-Marina-PEC1>).

4 RESULTATS

Per tal de treballar amb les dades generem un SummarizedExperiment, que ha de constar dels següents elements:

- Una matriu que serà utilitzada com a assaig o “assay”, on s'inclouen els valors d'abundància de les diferents accessions obtingudes per cada mostra.

- Una taula on es descriuen les mostres que s'han utilitzat.
- Una taula on es descriuen les accessions, és a dir, els pèptids fosforilats.

Font: Pàgina web Bioinfo Training, Chapter 2: RNA-Seq analysis with R and Bioconductor. Obtingut en l'enllaç: <https://uclouvain-cbio.github.io/bioinfo-training-02-rnaseq/sec-se.html> (Consultat per última vegada el 6 de novembre de 2024)

Per tal de generar el Summarized Experiment, en primer lloc, obtenim les dades del repositori:

Llegim el primer full de càlcul i ho assignem al dataframe 'phData':

```
require(readxl) #utilitzem el paquet per llegir excels.
```

```
## Cargando paquete requerido: readxl
```

```
#Generem el dataframe 'phData'
```

```
phData <- read_excel(path= "TI02+PTYR-human-MSS+MSIvsPD.XLSX", sheet=1)
```

```
head(phData) #Visualitzem els primers valors del dataframe
```

```
## # A tibble: 6 x 18
##   SequenceModifications Accession Description Score M1_1_MSS M1_2_MSS M5_1_MSS
##   <chr>                <chr>    <chr>      <dbl>    <dbl>    <dbl>    <dbl>
## 1 LYPELSQYMGLSLNEEEIR[2]~ 000560 Syntenin-1~ 48.1      24.3    44476.      0
## 2 VDKVIQAQTAFSANPANPAILS~ 000560 Syntenin-1~ 67.0       0    43139.    2102.
## 3 VIQAQTAFSANPANPAILSEAS~ 000560 Syntenin-1~ 77.7    3413.   172143.   77323.
## 4 HADAEMTGYVVTR[6] Oxida~ 015264 Mitogen-ac~ 44.9  220431.  145657.  104288.
## 5 HADAEMTGYVVTR[9] Phosp~ 015264 Mitogen-ac~ 67.4  18255.    8530.   35956.
## 6 STGPGASLGTGYDR[12] Pho~ 015551 Claudin-3 ~ 63.7  644513.  261938.  187023.
## # i 11 more variables: M5_2_MSS <dbl>, T49_1_MSS <dbl>, T49_2_MSS <dbl>,
## #   M42_1_PD <dbl>, M42_2_PD <dbl>, M43_1_PD <dbl>, M43_2_PD <dbl>,
## #   M64_1_PD <dbl>, M64_2_PD <dbl>, CLASS <chr>, PHOSPHO <chr>
```

Tal com podem veure aquest primer full de càlcul conté les seqüències de pèptids fosforilats, el seu número d'accessió, la descripció de la modificació a la qual correspon, un score que es dona per cada modificació seguida de la quantitat obtinguda per cada mostra i rèplica analitzada, finalment també s'inclouen la variable classe i fosfo.

Seguidament, llegim el segon full de càlcul i ho assignem al dataframe 'targets':

```
#Repetim el mateix procés pel segon full de càlcul:
```

```
targets <- read_excel(path= "TI02+PTYR-human-MSS+MSIvsPD.XLSX", sheet=2)
```

```
## New names:
```

```
## * 'Sample' -> 'Sample...1'
```

```
## * 'Sample' -> 'Sample...2'
```

```
show(targets)
```

```
## # A tibble: 12 x 4
##   Sample...1 Sample...2 Individual Phenotype
##   <chr>      <chr>      <dbl> <chr>
## 1 M1_1      M1              1 MSS
```

```
## 2 M1_2      M1          1 MSS
## 3 M5_1      M5          2 MSS
## 4 M5_2      M5          2 MSS
## 5 T49_1     T49         3 MSS
## 6 T49_2     T49         3 MSS
## 7 M42_1     M42         4 PD
## 8 M42_2     M42         4 PD
## 9 M43_1     M43         5 PD
## 10 M43_2    M43         5 PD
## 11 M64_1    M64         6 PD
## 12 M64_2    M64         6 PD
```

Podem veure que en aquest segon dataframe la informació que es proporciona es correspon a les mostres enviades, ens indica el número de la rèplica tècnica, el tipus de mostra que es tracta així com el fenotip. Podem veure que hi ha alguna informació redundant, ja que a partir de la primera columna en pots deduir la segona i la tercera.

Per tal crear el Summarized Experiments, en primer lloc, hem de tenir clar les dimensions de les dades amb què treballem, podem determinar la dimensió de les dades proporcionades mitjançant la funció 'dim':

```
dim(phData) #Utilitzem el dataframe phData, ja que conté els valors d'interès
```

```
## [1] 1438    18
```

Podem veure que el dataframe conté 1438 files de dades, és a dir, 1438 fosfopèptids i 18 columnes, és a dir, 18 variables diferents.

4.1 Creació del SummarizedExperiment

Tal com he mencionat prèviament, per tal de crear el SummarizedExperiment s'han de generar els seus components de forma individual (la matriu, la informació sobre les mostres i la de les accessions).

4.1.1 Creem la matriu

Part de la matriu ja es troba en el data.frame proporcionat, com podem veure en les columnes de la 5 a la 16, que són les que contenen les mostres analitzades amb les abundàncies de cada accessió:

```
abundances<-phData[,5:16]
abundances #Visualitzem les columnes de la 5 a la 16 del dataframe phData
```

```
## # A tibble: 1,438 x 12
##   M1_1_MSS M1_2_MSS M5_1_MSS M5_2_MSS T49_1_MSS T49_2_MSS M42_1_PD M42_2_PD
##   <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
## 1    24.3  44476.      0      6269.   1136.   21934.      0        0
## 2      0   43139.   2102.   50355.    249.    3239.   1316.      0
## 3   3413.  172143.   77323.  307637.   98442.  192982.  24851.  16548.
## 4 220431.  145657.  104288.   75887.  773377.  481166. 1027196. 1163747.
## 5  18255.    8530.   35956.   44102.   57145.   34638.  21231.  49500.
## 6 644513.  261938.  187023.  124868.  4487444. 2572575. 535809.  434646.
## 7 686820.  331984.  252694.  322890. 34886738. 6651274. 832709.  827703.
## 8 815186.  728701.  267179.  251302. 26259011. 29240206. 686115. 1776036.
```

```
## 9 1578. 9836. 2033. 2221. 541270. 781087. 817. 0
## 10 2815. 13247. 2121. 638. 994154. 2864331. 1438. 37167.
## # i 1,428 more rows
## # i 4 more variables: M43_1_PD <dbl>, M43_2_PD <dbl>, M64_1_PD <dbl>,
## # M64_2_PD <dbl>
```

Per acabar de formar la matriu necessitem afegir-hi el nom del què correspon cada columna, és a dir, les diferents accessions dels fosfopèptids, per fer això, guardem les accessions en una variable que anomenem `Accession_names`:

```
Accession_names<-phData$Accession #Generem la variable
head(Accession_names) #En visualitzem els primers valors
```

```
## [1] "000560" "000560" "000560" "015264" "015264" "015551"
```

```
matrix <- as.matrix(abundances) #Convertim el dataframe en la matriu de valors
rownames(matrix)<-Accession_names #Afegim les accessions com a nom de les columnes
matrix[1:5,] #Visualitzem els primers resultats de la matriu
```

```
##          M1_1_MSS M1_2_MSS M5_1_MSS M5_2_MSS T49_1_MSS T49_2_MSS
## 000560    24.29438 44475.964    0.000  6269.141  1135.8169 21933.90
## 000560     0.00000 43138.904   2102.056 50355.051   248.9275  3239.16
## 000560   3412.60332 172143.040  77323.019 307637.429 98442.2773 192982.37
## 015264  220431.17880 145656.887 104287.815  75887.365 773377.4981 481165.54
## 015264  18254.77813   8529.755  35955.901  44102.316  57145.1682  34638.01
##          M42_1_PD M42_2_PD M43_1_PD M43_2_PD M64_1_PD M64_2_PD
## 000560     0.000     0.00   772.9056   2136.746   1820.724   1727.9098
## 000560   1315.904     0.00     0.0000     0.000     0.000     892.3565
## 000560   24851.344  16547.95   5565.2821     0.000   3264.563   5901.9577
## 015264 1027196.292 1163747.38 4080239.1820 4885818.113 3093786.793 2759104.5440
## 015264   21231.256  49499.70  666107.0448  379313.615  255792.117  579765.0018
```

Tal com podem veure en el resultat anterior, la matriu de valors té especificada en les columnes les mostres de l'estudi i en les files l'accessió del fosfopèptid quantificat en cada mostra.

Comprovem la dimensió de la matriu de valors obtinguda, que hem denominat 'matrix':

```
dim(matrix)
```

```
## [1] 1438 12
```

Podem veure que en aquest cas hem reduït de 18 a 12 el nombre de columnes respecte al dataframe original, ja que aquesta matriu solament conté les 6 mostres analitzades i els seus duplicats (12 mostres en total).

4.1.2 Creem la informació de les mostres

Editem la informació referent a les mostres per tenir-la de forma concisa:

```
#Generem una variable que contingui el nom de les columnes de la matriu:
Samples<-colnames(matrix)

#Afegim aquesta variable a la informació que ja teníem de les mostres:
targets$Samples = Samples

#Reordenem les dades perquè la variable de les columnes aparegui la primera:
Targets <- targets[,c(5,1,2,3,4)]
Targets #visualitzem les dades que corresponen a les mostres
```

```
## # A tibble: 12 x 5
##   Samples   Sample...1 Sample...2 Individual Phenotype
##   <chr>      <chr>      <chr>          <dbl> <chr>
## 1 M1_1_MSS M1_1         M1              1 MSS
## 2 M1_2_MSS M1_2         M1              1 MSS
## 3 M5_1_MSS M5_1         M5              2 MSS
## 4 M5_2_MSS M5_2         M5              2 MSS
## 5 T49_1_MSS T49_1       T49             3 MSS
## 6 T49_2_MSS T49_2       T49             3 MSS
## 7 M42_1_PD M42_1       M42             4 PD
## 8 M42_2_PD M42_2       M42             4 PD
## 9 M43_1_PD M43_1       M43             5 PD
## 10 M43_2_PD M43_2       M43             5 PD
## 11 M64_1_PD M64_1       M64             6 PD
## 12 M64_2_PD M64_2       M64             6 PD
```

Aquesta és la informació que se'ns ha proporcionat sobre les mostres i que volem incloure en el Summarized-Experiment. Per això, hem de tenir informació sobre 12 mostres, ja que aquesta és la quantitat de columnes que té la nostra matriu de valors.

Com podem veure, tenim tantes files com nombre de columnes que té la matriu de valors, una per cada mostra analitzada:

```
dim(Targets)
```

```
## [1] 12 5
```

4.1.3 Creem la informació de les accessions

Seguim un procés similar a l'anterior per tal de generar la informació que volem introduir en el Summarized-Experiment sobre els fosfopèptids detectats en la fosfoproteòmica:

```
accession <- phData[,1:3] #Seleccionem les 3 primeres columnes de phData

#Reordenem les dades perquè l'accessió aparegui com a primera columna:
Phospho<- accession[,c(2,1,3)]
head(Phospho) #Visualitzem les dades
```

```
## # A tibble: 6 x 3
##   Accession SequenceModifications Description
##   <chr>      <chr>                      <chr>
```

```
## 1 000560 LYPELSQYMGLSLNEEEIR[2] Phospho|[9] Oxidation Syntenin-1 OS=Hom~
## 2 000560 VDKVIQAQTAFSANPANPAILSEASAPIPHDGNLYPR[35] Phospho Syntenin-1 OS=Hom~
## 3 000560 VIQAQTAFSANPANPAILSEASAPIPHDGNLYPR[32] Phospho Syntenin-1 OS=Hom~
## 4 015264 HADAEMTGYVVTR[6] Oxidation|[9] Phospho Mitogen-activated~
## 5 015264 HADAEMTGYVVTR[9] Phospho Mitogen-activated~
## 6 015551 STGPGASLGTGYDR[12] Phospho Claudin-3 OS=Homo~
```

Donat que aquestes dades corresponen a la informació sobre els fosfopèptids obtinguts, la quantitat de files que hi ha d'haver en les dades 'Phospho' han de ser equivalents a les de la matriu de valors:

```
dim(Phospho) #Visualitzem les dimensions de les dades
```

```
## [1] 1438 3
```

Podem veure que en els dos casos tenim 1438 columnes, per tant, podem seguir amb la creació del SummarizedExperiment.

Finalment, creem el SummarizedExperiment, que denominem com a 'se' utilitzant el comando 'SummarizedExperiment':

```
#Creem el SummarizedExperiment
se <- SummarizedExperiment(assays = list(counts = matrix), colData = Targets, rowData = Phospho)
se #Visualitzem el seu contingut
```

```
## class: SummarizedExperiment
## dim: 1438 12
## metadata(0):
## assays(1): counts
## rownames(1438): 000560 000560 ... Q13283 Q9NYF8
## rowData names(3): Accession SequenceModifications Description
## colnames(12): M1_1_MSS M1_2_MSS ... M64_1_PD M64_2_PD
## colData names(5): Samples Sample...1 Sample...2 Individual Phenotype
```

Podem veure en el codi anterior que en l'apartat 'assays' hi hem afegit la matriu de valors, en 'colData' les dades referents a les mostres i en rowData, la informació dels fosfopèptids. Quan en visualitzem el resum, podem veure que s'ha creat exitosament la classe amb tota la informació proporcionada. No obstant, podem apreciar que hi ha un apartat on hi manca informació, que es tracta de l'apartat metadata. L'apartat metadata conté informació rellevant sobre les dades disponibles en el SummarizedExperiment, com ara annexos o fórmules. En el nostre cas, en les dades proporcionades no hi ha massa més informació adicional, per aquest motiu hi afegim una breu descripció de l'experiment:

```
#Afegim en la metadata una descripció de l'experiment:
metadata(se)$description <- c("The accompanying dataset has been obtained from a phosphoproteomics exper
The experiment has analyzed (3 + 3) PDX models of two different subtypes using Phosphopeptide enriched LC-MS analysis of 2 technical duplicates has been performed on each sample.
The results set consisted of Normalized abundances of MS signals for ca. 1400 phosphopeptides
Goal of the analysis: *search phosphopeptides that allow differentiation of the two tumor groups*
This should be made with both Statistical Analysis and visualization.
Data have been provided as an excel file: TIO2+PTYR-human-MSS+MSIvsPD.XLSX

Groups are defined as:
- MSS group: Samples M1, M5 and T49,
```

- PD group: Samples M42, M43 and M64 with two technical replicates for each sample
The first column, *SequenceModification* contains abundance values for the distinct phosphopeptides. Other

#Tornem a visualitzar per tal de confirmar que ara a metadata hi ha informació
se

```
## class: SummarizedExperiment
## dim: 1438 12
## metadata(1): description
## assays(1): counts
## rownames(1438): 000560 000560 ... Q13283 Q9NYF8
## rowData names(3): Accession SequenceModifications Description
## colnames(12): M1_1_MSS M1_2_MSS ... M64_1_PD M64_2_PD
## colData names(5): Samples Sample...1 Sample...2 Individual Phenotype
```

Font: Pàgina web RDRR. Obtingut en l'enllaç: <https://rdrr.io/bioc/SummarizedExperiment/f/vignettes/SummarizedExperiment.Rmd> (Consultat per última vegada el 6 de novembre de 2024)

4.2 Anàlisi del SummarizedExperiment creat

4.2.1 Comprovació de la informació disponible en el SummarizedExperiment

De la classe SummarizedExperiment creada en podem obtenir tota la informació disponible de la següent manera:

1. Obtenció de la informació sobre les mostres:

Accedim a la informació de les mostres mitjançant colData:
colData(se)

```
## DataFrame with 12 rows and 5 columns
##           Samples Sample...1 Sample...2 Individual Phenotype
##           <character> <character> <character> <numeric> <character>
## M1_1_MSS      M1_1_MSS      M1_1         M1             1          MSS
## M1_2_MSS      M1_2_MSS      M1_2         M1             1          MSS
## M5_1_MSS      M5_1_MSS      M5_1         M5             2          MSS
## M5_2_MSS      M5_2_MSS      M5_2         M5             2          MSS
## T49_1_MSS     T49_1_MSS     T49_1        T49            3          MSS
## ...          ...          ...          ...          ...          ...
## M42_2_PD      M42_2_PD      M42_2        M42            4          PD
## M43_1_PD      M43_1_PD      M43_1        M43            5          PD
## M43_2_PD      M43_2_PD      M43_2        M43            5          PD
## M64_1_PD      M64_1_PD      M64_1        M64            6          PD
## M64_2_PD      M64_2_PD      M64_2        M64            6          PD
```

2. Obtenció de la informació sobre els fosfopèptids detectats:

Accedim a la informació de les accessions mitjançant rowData:
rowData(se)


```
## DataFrame with 1438 rows and 3 columns
##      Accession SequenceModifications      Description
##      <character>      <character>      <character>
## 000560      000560 LYPELSQYMGLSLNEEEIR[... Syntenin-1 OS=Homo s...
## 000560      000560 VDKVIQAQTAFSANPANPAI... Syntenin-1 OS=Homo s...
## 000560      000560 VIQAQTAFSANPANPAILSE... Syntenin-1 OS=Homo s...
## 015264      015264 HADAEMTGYVVTR[6] Oxi... Mitogen-activated pr...
## 015264      015264 HADAEMTGYVVTR[9] Pho... Mitogen-activated pr...
## ...      ...      ...      ...
## Q12792      Q12792 YLLSQSSPAPLTAAEEELR[... Twinfilin-1 OS=Homo ..
## Q13177      Q13177 YLSFTPPEK[3] Phospho Serine/threonine-pro...
## 095218      095218 YNLDASEEEDSNK[6] Pho... Zinc finger Ran-bind...
## Q13283      Q13283 YQDEVFGGFVTEPQEESEEE... Ras GTPase-activatin...
## Q9NYF8      Q9NYF8 YSPSQNSPIHHIPSRR[1] .. Bcl-2-associated tra...
```

3. Obtenció de la matriu de valors:

```
# Accedim a la matriu de valors:
head(assays(se)$counts) #En visualitzem solament els primers valors
```

```
##      M1_1_MSS  M1_2_MSS  M5_1_MSS  M5_2_MSS  T49_1_MSS  T49_2_MSS
## 000560  24.29438  44475.964    0.000  6269.141  1135.8169  21933.90
## 000560   0.00000  43138.904  2102.056  50355.051   248.9275   3239.16
## 000560  3412.60332 172143.040  77323.019 307637.429  98442.2773 192982.37
## 015264 220431.17880 145656.887 104287.815  75887.365  773377.4981 481165.54
## 015264  18254.77813   8529.755  35955.901  44102.316   57145.1682  34638.01
## 015551 644513.31840 261938.025 187023.484 124867.715 4487443.6920 2572575.27
##      M42_1_PD  M42_2_PD  M43_1_PD  M43_2_PD  M64_1_PD  M64_2_PD
## 000560    0.000    0.00    772.9056   2136.746   1820.724   1727.9098
## 000560   1315.904    0.00    0.0000    0.000    0.000    892.3565
## 000560   24851.344   16547.95   5565.2821    0.000   3264.563   5901.9577
## 015264 1027196.292 1163747.38 4080239.1820 4885818.113 3093786.793 2759104.5440
## 015264   21231.256   49499.70  666107.0448  379313.615  255792.117  579765.0018
## 015551  535809.187  434645.89   91361.8781   65997.913  243250.439 206632.6444
```

4. Obtenim les metadades relacionades:

```
# Accedim a les metadades:
metadata(se)$description
```

```
## [1] "The accompanying dataset has been obtained from a phosphoproteomics experiment\nThe experiment h...
```

Amb aquest últim pas podem veure que la classe funciona correctament i conté tota la informació que li hem proporcionat.

4.2.2 Descripció general de les dades

Mitjançant el Summarized Experiment es permet d'una forma fàcil i senzilla respondre a diferents preguntes sobre les dades de la fosfoproteòmica.

Per exemple, si llegim la descripció que es proporciona en la metadata podem averiguar el nombre variables de l'experiment:

```
metadata(se)$description
```

```
## [1] "The accompanying dataset has been obtained from a phosphoproteomics experiment\nThe experiment ha
```

En aquest cas, ens queda clar que són dues condicions diferents (dos fenotips, MSS i PD) compostos per 3 mostres cada un amb dues repeticions per mostra. Les informacions sobre les mostres estudiades es podria accedir ràpidament tal com hem fet anteriorment amb 'colData' així com la informació dels pèptids obtinguts que es podria obtenir amb 'rowData'. D'una forma més breu, per determinar la quantitat de mostres amb què treballem, ho podem veure mitjançant el càlcul de les dimensions de la matriu de valors:

```
dim(assays(se)$counts)
```

```
## [1] 1438 12
```

Podem veure que hi ha 1438 columnes, és a dir, 1438 fosfopèptids estudiats i 12 mostres, que corresponen a les columnes.

També veiem que en la matriu no hi ha cap NA, fet que és imprescindible de determinar abans de prosseguir amb l'anàlisi de les dades:

```
table(is.na(assays(se)$count))
```

```
##  
## FALSE  
## 17256
```

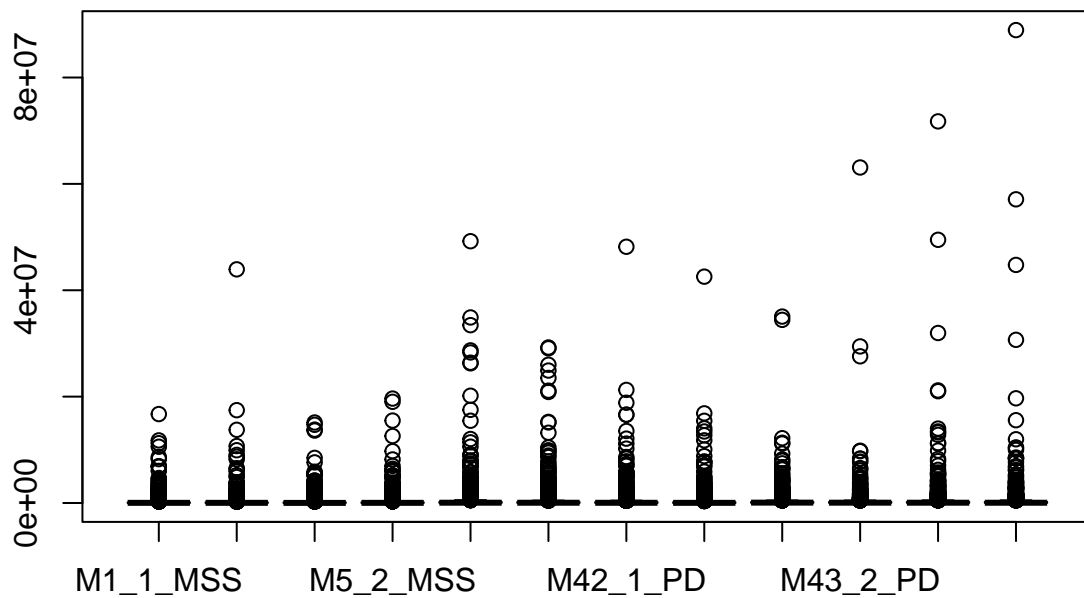
Podem obtenir fàcilment estadístiques bàsiques sobre les abundàncies obtingudes per cada mostra: Obtindre estadístiques i gràfics simples:

```
summary(assays(se)$count)
```

```
##      M1_1_MSS      M1_2_MSS      M5_1_MSS      M5_2_MSS  
## Min.   :      0   Min.   :      0   Min.   :      0   Min.   :      0  
## 1st Qu.:    5653   1st Qu.:    5497   1st Qu.:    2573   1st Qu.:    3273  
## Median :   30682   Median :   26980   Median :   20801   Median :   26241  
## Mean   :  229841   Mean   :  253151   Mean   :  232967   Mean   :  261067  
## 3rd Qu.:  117373   3rd Qu.:  113004   3rd Qu.:  113958   3rd Qu.:  130132  
## Max.   :16719906   Max.   :43928481   Max.   :15135169   Max.   :19631820  
##      T49_1_MSS      T49_2_MSS      M42_1_PD      M42_2_PD  
## Min.   :      0   Min.   :      0   Min.   :      0   Min.   :      0  
## 1st Qu.:    9306   1st Qu.:    8611   1st Qu.:    5341   1st Qu.:    4216  
## Median :   55641   Median :   46110   Median :   36854   Median :   30533  
## Mean   :   542449   Mean   :   462616   Mean   :   388424   Mean   :   333587  
## 3rd Qu.:  223103   3rd Qu.:  189141   3rd Qu.:  180252   3rd Qu.:  152088  
## Max.   :49218872   Max.   :29240206   Max.   :48177680   Max.   :42558111  
##      M43_1_PD      M43_2_PD      M64_1_PD      M64_2_PD  
## Min.   :      0   Min.   :      0   Min.   :      0   Min.   :      0  
## 1st Qu.:   19641   1st Qu.:   17299   1st Qu.:   11038   1st Qu.:    8660  
## Median :   67945   Median :   59607   Median :   52249   Median :   47330  
## Mean   :   349020   Mean   :   358822   Mean   :   470655   Mean   :   484712  
## 3rd Qu.:  205471   3rd Qu.:  201924   3rd Qu.:  209896   3rd Qu.:  206036  
## Max.   :35049402   Max.   :63082982   Max.   :71750330   Max.   :88912734
```

Aquestes dades també poden ser representades gràficament:

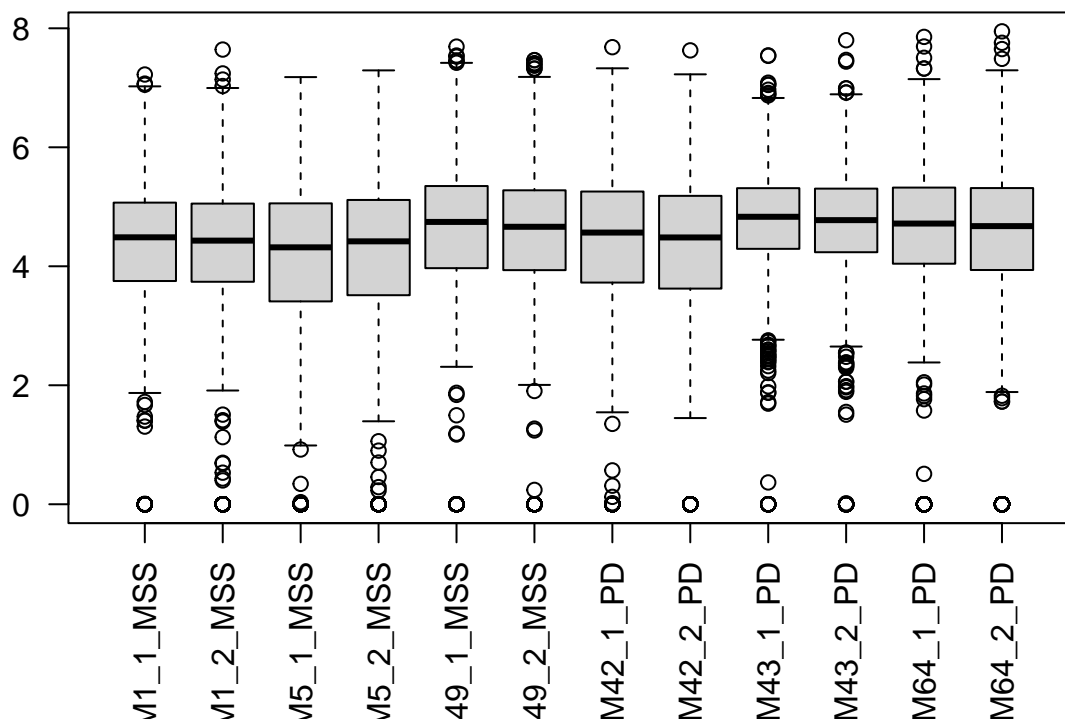
```
boxplot(assays(se)$count)
```



De la representació de les diferents abundàncies en podem extreure dues coses; que les abundàncies detectades entre mostres varien àmpliament entre elles, i que degut a que es tracten a dades en exponents, seria interessant manipular les dades en base logarítmica. A continuació visualitzem les dades, però en escala log10:

```
logmatrix<-log10(assays(se)$count+1) #Transformem la matriu, hi sumem 1 per evitar prOblemes amb valors
#Visualitzem en forma de boxplot les dades transformades:
boxplot(logmatrix, las=2, main="Phosphoproteomics Experiment. Abundance in log 10 scale")
```

Phosphoproteomics Experiment. Abundance in log 10 scale



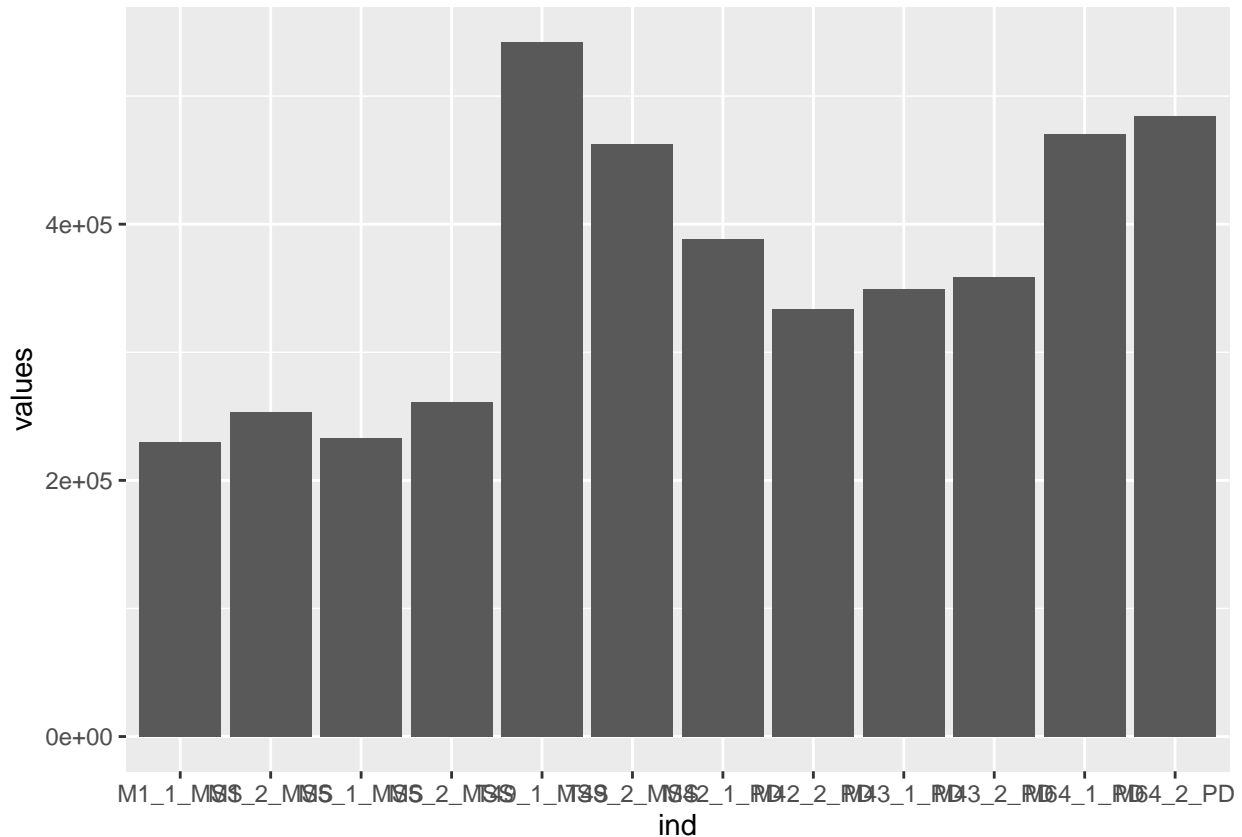
També podem visualitzar les mitges d'abundància de les diferents mostres:

```
#Creem un vector que conté les mitges de les columnes de la matriu de valors
mean_sample<-colMeans(assays(se)$count)
mean_sampledf<-stack(mean_sample) #Creem un dataframe amb les mostres i les mitges
mean_sampledf #Visualitzem el dataframe
```

```
##      values      ind
## 1  229841.2  M1_1_MSS
## 2  253151.4  M1_2_MSS
## 3  232966.5  M5_1_MSS
## 4  261066.7  M5_2_MSS
## 5  542448.9  T49_1_MSS
## 6  462616.5  T49_2_MSS
## 7  388424.0  M42_1_PD
## 8  333587.0  M42_2_PD
## 9  349019.6  M43_1_PD
## 10 358821.7  M43_2_PD
## 11 470655.0  M64_1_PD
## 12 484712.4  M64_2_PD
```

Podem veure que aquesta informació concorda amb la visualitzada anteriorment amb la funció 'summary' de la matriu de valors. Podem representar gràficament aquests valors de la forma següent:

```
library(ggplot2)
ggplot(mean_sampledf, aes(x=ind, y=values))+
  geom_col()
```



Podem veure un resultat similar al vist amb els boxplots.

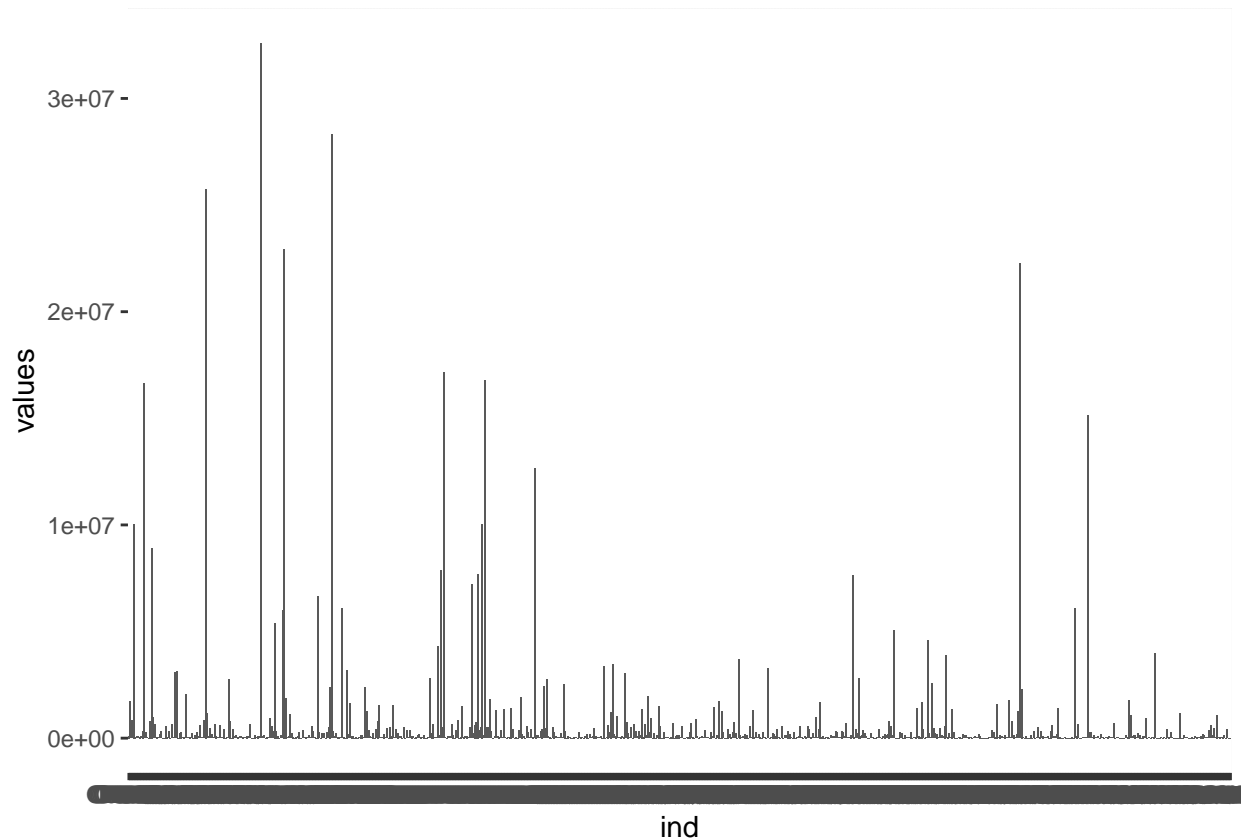
Pel que fa a les dades dels fosfopèptids, podem obtenir una informació similar:

```
mean_accession<-rowMeans(matrix) #Calcuem la mitja d'abundància per accesió
mean_accessiondf<-stack(mean_accession) #Creem el dataframe amb la informació
head(mean_accessiondf) #Visualitzem el dataframe
```

```
##      values    ind
## 1    6691.45 000560
## 2    8441.03 000560
## 3   75672.65 000560
## 4 1567558.22 015264
## 5  179194.55 015264
## 6  821338.29 015551
```

Podem representar gràficament aquesta informació mitjançant el barplot següent:

```
ggplot(mean_accessiondf, aes(x=ind, y=values))+
  geom_col()
```



En tenir tanta informació resulta una mica entendre el resultat obtingut. Tot i que podem veure que les abundàncies varien enormement segons el fosfopèptid. Per tal de determinar quins són els pèptids més abundants seguim els següents passos:

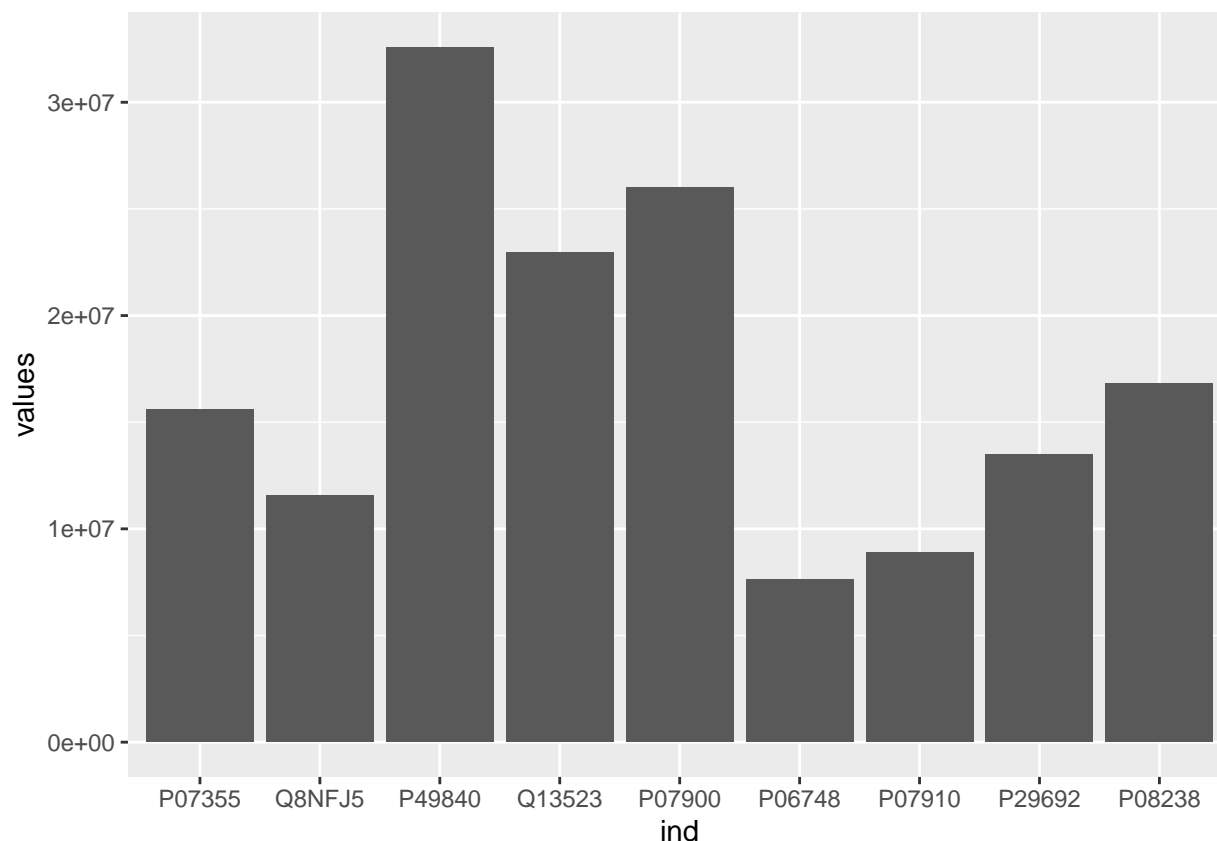
```
#Ordenem les dades per quantitat
quant_accession<-mean_accessiondf[order(mean_accessiondf$values,
                                         decreasing = TRUE), ]
#Visualitzem les dades i comprovem que estan en l'ordre adequat
head(quant_accession)
```

```
##      values   ind
## 164 32574211 P49840
## 186 22943961 Q13523
## 1194 16803191 P08238
## 22 15603857 P07355
## 383 14568440 P07900
## 446 13492704 P29692
```

A continuació veiem els 10 fosfopèptids més abundants detectats en les mostres:

```
#Seleccióem les 10 primeres entrades, corresponents als pèptids més abundants
highaccession<-quant_accession[1:10,]

#Visualitzem gràficament aquestes dades mitjançant un barplot:
ggplot(highaccession, aes(x=ind, y=values))+
  geom_col()
```



Tot i que és interessant la visualització d'aquesta informació, no ens són de massa utilitat per la finalitat d'aquest treball, que és la comparació dels diferents perfils de fosfopèptids obtinguts en les diferents mostres. Per això és necessari l'ús d'eines per la comparació, que veurem en el següent apartat.

4.2.3 Comparació del perfil fosfopeptídic de les dades obtingudes

El primer pas per la comparació de les dades obtingudes és generar una matriu de correlació a partir de la matriu de valors del SummarizedExperiment:

```
#Generem la matriu de correlació amb la funció 'cor':
corrmatrix<-cor(assays(se)$count)
round(corrmatrix, 2) #La visualitzem arrodonint els valors a 2 decimals
```

```
##          M1_1_MSS M1_2_MSS M5_1_MSS M5_2_MSS T49_1_MSS T49_2_MSS M42_1_PD
## M1_1_MSS      1.00    0.69    0.83    0.79    0.70    0.73    0.71
## M1_2_MSS      0.69    1.00    0.83    0.86    0.46    0.58    0.45
## M5_1_MSS      0.83    0.83    1.00    0.99    0.61    0.72    0.61
## M5_2_MSS      0.79    0.86    0.99    1.00    0.60    0.71    0.59
## T49_1_MSS     0.70    0.46    0.61    0.60    1.00    0.92    0.80
## T49_2_MSS     0.73    0.58    0.72    0.71    0.92    1.00    0.78
## M42_1_PD      0.71    0.45    0.61    0.59    0.80    0.78    1.00
## M42_2_PD      0.64    0.36    0.48    0.44    0.78    0.72    0.91
## M43_1_PD      0.56    0.43    0.45    0.43    0.64    0.60    0.71
## M43_2_PD      0.42    0.79    0.54    0.57    0.43    0.50    0.46
## M64_1_PD      0.63    0.36    0.49    0.47    0.74    0.69    0.85
```

## M64_2_PD	0.52	0.81	0.64	0.67	0.57	0.62	0.64
##	M42_2_PD	M43_1_PD	M43_2_PD	M64_1_PD	M64_2_PD		
## M1_1_MSS	0.64	0.56	0.42	0.63	0.52		
## M1_2_MSS	0.36	0.43	0.79	0.36	0.81		
## M5_1_MSS	0.48	0.45	0.54	0.49	0.64		
## M5_2_MSS	0.44	0.43	0.57	0.47	0.67		
## T49_1_MSS	0.78	0.64	0.43	0.74	0.57		
## T49_2_MSS	0.72	0.60	0.50	0.69	0.62		
## M42_1_PD	0.91	0.71	0.46	0.85	0.64		
## M42_2_PD	1.00	0.83	0.52	0.86	0.61		
## M43_1_PD	0.83	1.00	0.75	0.89	0.76		
## M43_2_PD	0.52	0.75	1.00	0.55	0.93		
## M64_1_PD	0.86	0.89	0.55	1.00	0.71		
## M64_2_PD	0.61	0.76	0.93	0.71	1.00		

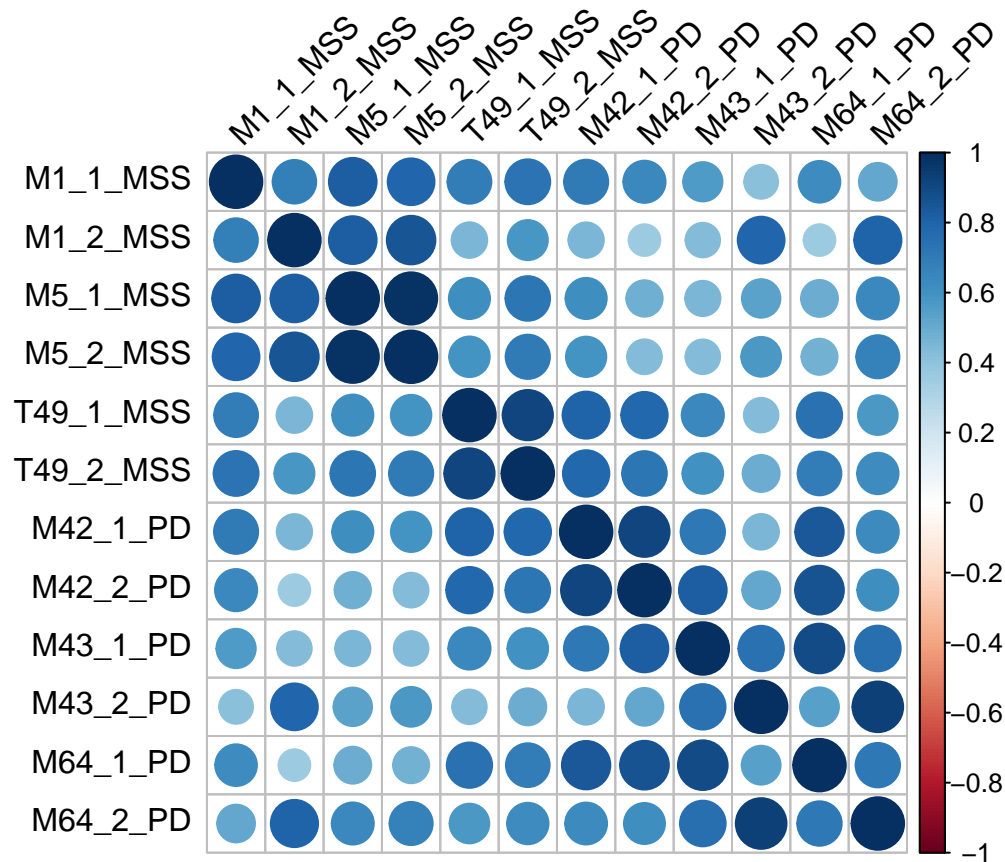
Aquesta matriu ens compara les diferents mostres entre elles i ens determina la seva similitud. Com podem veure, tenim una diagonal de valors 1 d'esquerra a dreta que corresponen a la comparació de la mateixa mostra, on la similitud és màxima.

Podem representar gràficament aquestes correlacions de diferents maneres:

```
#install.packages("corrplot")
library(corrplot)
```

```
## corrplot 0.95 loaded
```

```
corrplot(corrmatrix,
          tl.col = "black", tl.srt = 45)
```

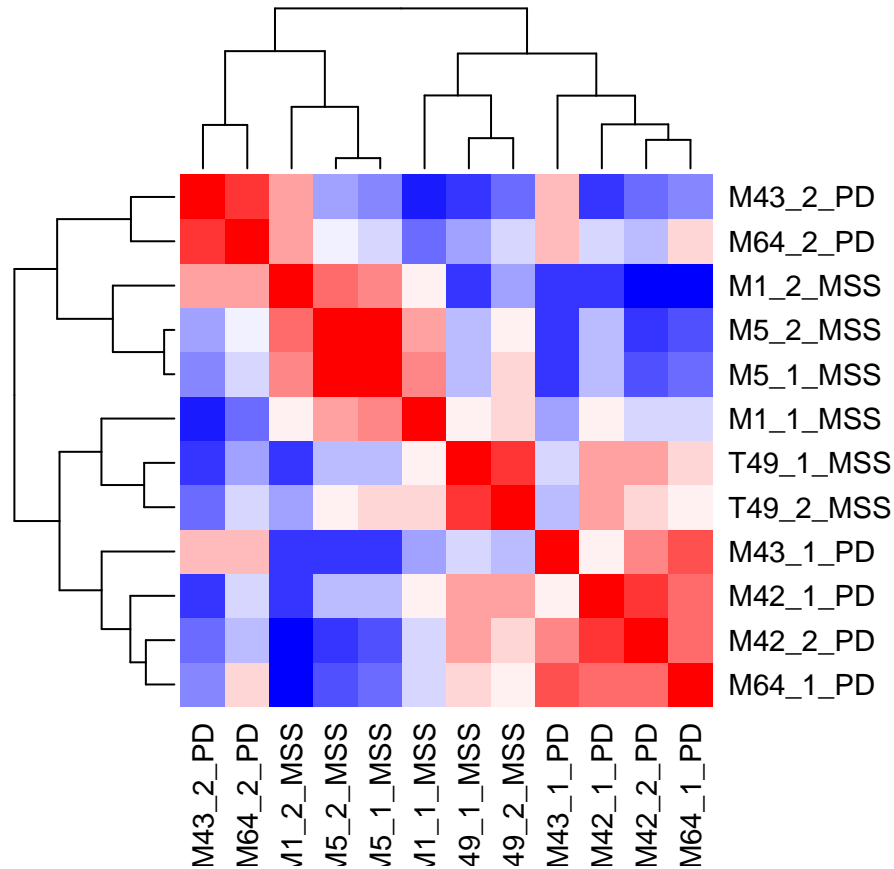



Gràcies a aquest gràfic podem veure com de similars són aquestes mostres entre elles. Veiem la diagonal de blau fosc que indica una màxima similitud a l'hora de comparar les mateixes mostres.

- Generalment podem veure que els replicats de les mostres són molt similars entre ells, sent la mostra M1 l'única excepció.
- Podem veure que hi ha certes diferències entre fenotip MSS i PD, però depenen en part de la mostra

Alternativament, també podem visualitzar aquestes dades amb un heatmap:

```
col<- colorRampPalette(c("blue", "white", "red"))(20)
heatmap(x = corrmatrix, col = col, symm = TRUE)
```



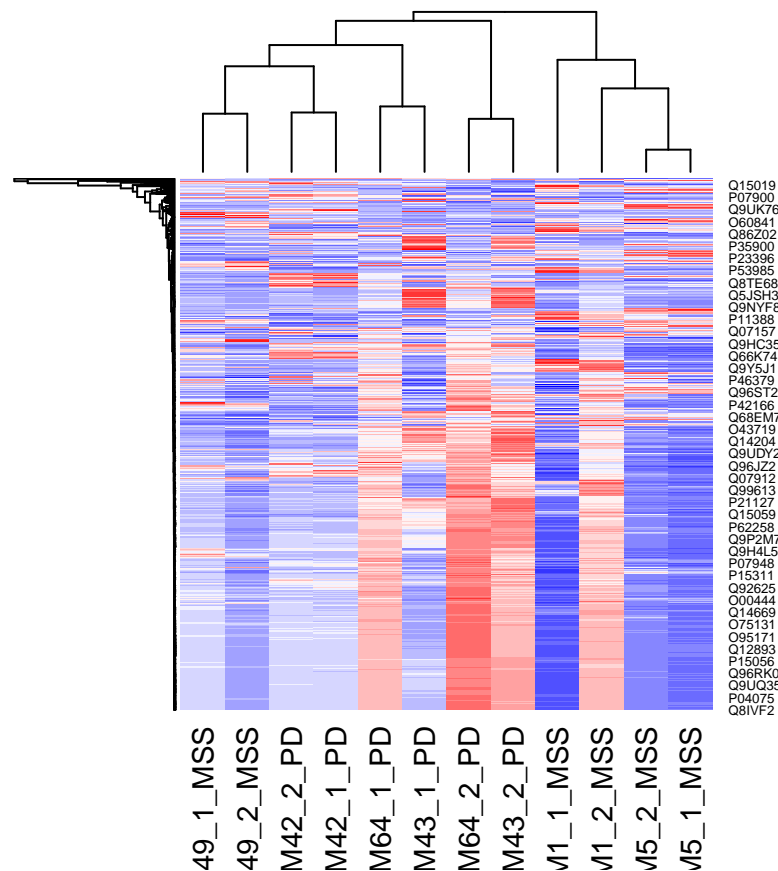
Si normalitzem les dades de la matriu de valors, en podem comparar l'abundància de pèptids per cada mostra, per això, utilitzem la funció 'scale':

```
data_normalized<-scale(assays(se)$count) #Normalitzem els valors de la matriu
head(data_normalized)#En visualitzem els primers valors
```

```
##           M1_1_MSS      M1_2_MSS      M5_1_MSS      M5_2_MSS      T49_1_MSS      T49_2_MSS
## 000560 -0.24829131 -0.139978619 -0.2378408 -0.22713246 -0.20435869 -0.211590122
## 000560 -0.24831756 -0.140875513 -0.2356948 -0.18783327 -0.20469351 -0.220566248
## 000560 -0.24463063 -0.054340082 -0.1589000  0.04151418 -0.16762317 -0.129462598
## 015264 -0.01016646 -0.072106882 -0.1313710 -0.16507317  0.08718108  0.008906192
## 015264 -0.22859533 -0.164091185 -0.2011326 -0.19340710 -0.18321382 -0.205490347
## 015551  0.44800660  0.005894016 -0.0469043 -0.12141096  1.48933023  1.013079467
##           M42_1_PD      M42_2_PD      M43_1_PD      M43_2_PD      M64_1_PD      M64_2_PD
## 000560 -0.20633416 -0.20472293 -0.2205279 -0.168470548 -0.16843312 -0.14524269
## 000560 -0.20563514 -0.20472293 -0.2210174 -0.169479783 -0.16908724 -0.14549395
## 000560 -0.19313291 -0.19456743 -0.2174932 -0.169479783 -0.16791441 -0.14398747
## 015264  0.33932136  0.50947080  2.3628028  2.138204835  0.94238486  0.68395329
## 015264 -0.19505594 -0.17434489  0.2007963  0.009678789 -0.07719150  0.02858414
## 015551  0.07829229  0.06201999 -0.1631623 -0.138307446 -0.08169722 -0.08362391
```

Si en comparem les abundàncies, sent de color blau els valors menys abundants i en vermell els més, obtenim la informació següent:

```
col<- colorRampPalette(c("blue", "white", "red"))(20)
heatmap(x = data_normalized, col=col)
```



Podem veure cada accessió i la seva quantitat per mostra. Podem veure que el heatmap ordena les mostres per similitud, i que aquestes estan ordenades per fenotip per l'excepció de la mostra M49, que és situada juntament a les mostres de fenotip PD. Podem veure que el perfil de fosfopèptids resulta bastant depenent de la mostra de PDX analitzada i que les rèpliques són altament similars entre elles a excepció de la mostra M1, com ja havíem vist abans.

Font: Pàgina web STHDA: Statistical tools for high-throughput data analysis. Obtingut en l'enllaç: <http://www.sthda.com/english/wiki/correlation-matrix-a-quick-start-guide-to-analyze-format-and-visualize-a-correlation-matrix-using-r-software> (Consultat per última vegada el 6 de novembre de 2024)

També podem estudiar les dades mitjançant una anàlisi de components principals (PCA). Primer fer el PCA i després en podem representar els primers components:

```
#Fem el PCA a partir de les dades normalitzades:
```

```
library(stats)
pca<-prcomp(data_normalized, scale.=TRUE)
summary(pca) #En visualitzem el resum
```

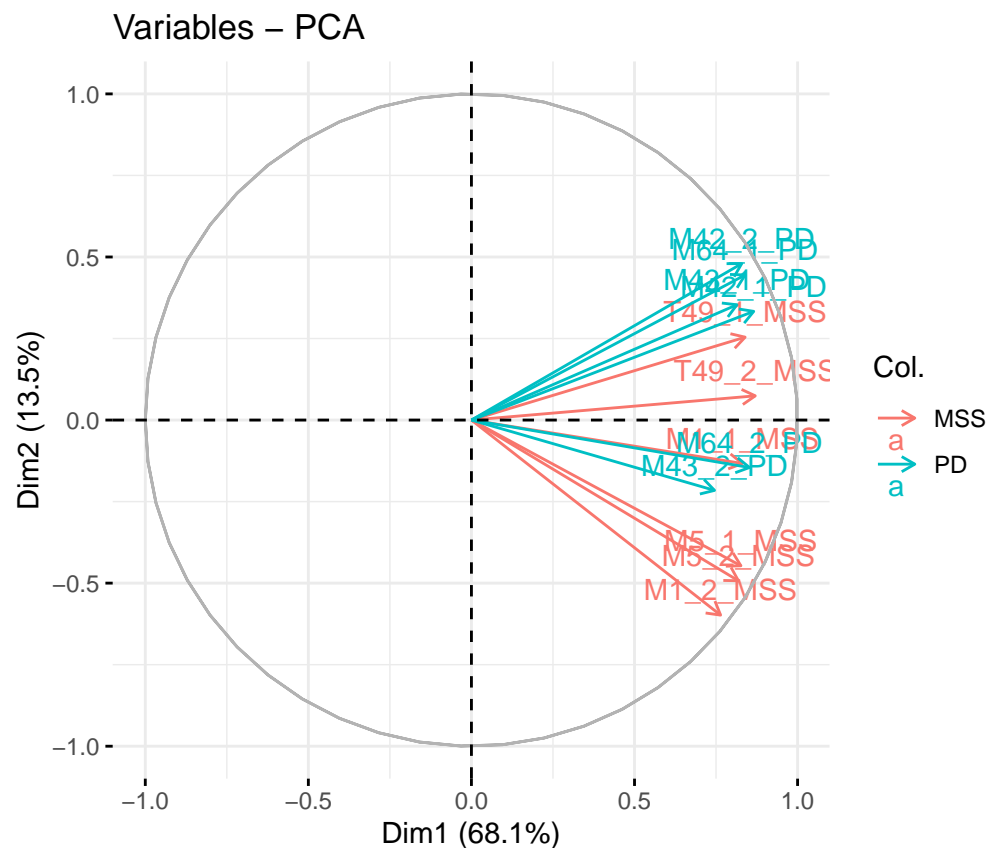
```
## Importance of components:
##              PC1      PC2      PC3      PC4      PC5      PC6      PC7
## Standard deviation  2.859 1.2725 1.07971 0.62493 0.47360 0.44078 0.34584
## Proportion of Variance 0.681 0.1349 0.09715 0.03255 0.01869 0.01619 0.00997
## Cumulative Proportion 0.681 0.8159 0.91310 0.94564 0.96433 0.98052 0.99049
##              PC8      PC9      PC10     PC11     PC12
```

```
## Standard deviation      0.2593 0.18457 0.08757 0.05426 0.04655
## Proportion of Variance 0.0056 0.00284 0.00064 0.00025 0.00018
## Cumulative Proportion  0.9961 0.99894 0.99957 0.99982 1.00000
```

Podem veure que s'han generat els 12 components principals (PC1-PC12) que corresponen al nombre de variables de les nostres dades. Cada component explica un percentatge de la variància total del dataset. Podem veure, en la proporció acumulada que els dos primers datasets expliquen un 81.59% de la variància del dataset. És a dir, més de tres quarts de la informació del dataset pot ser representat amb aquests dos components principals i aquest suposen una representació acurada de les dades.

Podem representar gràficament aquest PCA:

```
#Representem gràficament el PCA
fviz_pca_var(pca, col.var = colData(se)$Phenotype) #Indiquem que el color correspongui al fenotip
```



Aquest gràfic revela informació similar a la proporcionada anteriorment. Generalment, hi ha diferències entre fenotips MSS i PD a excepció de la T49 MSS i algunes repeticions de mostres PD (M62_2 i M43_2), que també podríem observar en el heatmap.

Donada a la diferència de la mostra T49, resultaria interessant saber si hi ha alguna diferència d'aquest xenograft respecte a la resta de PDX de fenotip MSS que pugui explicar la seva similitud al fenotip PD.

Podem analitzar les dades mitjançant clústers, a continuació farem servir el sistema K-means, àmpliament popular en l'anàlisi de clústers. Aquest mètode, però, requereix que s'especifiqui el nombre de clústers que s'han d'extreure de forma manual, per tal de determinar-ho fem el gràfic de suma de quadrats dins dels grups o within sum of squares error (WSS).

```

# Fem la visualització del decreixement de la suma de quadrats dins dels grups
# Fem el càlcul de la suma de quadrats dins de grups
wss <- (nrow(data_normalized)-1)*sum(apply(data_normalized,2,var))
for (i in 2:15) wss[i] <- sum(kmeans(data_normalized,
  centers=i)$withinss)

```

```

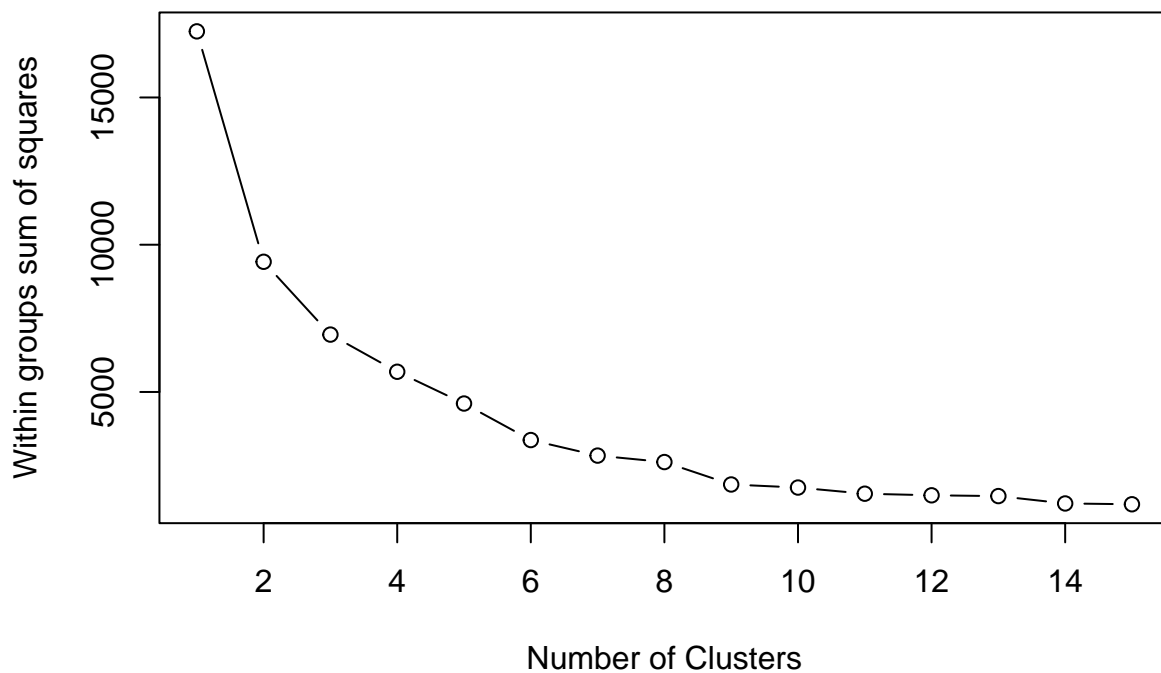
## Warning: did not converge in 10 iterations
## Warning: did not converge in 10 iterations

```

```

plot(1:15, wss, type="b", xlab="Number of Clusters",
  ylab="Within groups sum of squares")

```



```

library(tibble)
# Determinem el nombre de clústers
n_clusters <- 10

# Inicialitzem el total dins del wss
wss <- numeric(n_clusters)

set.seed(324)

# Mirem de 1 a 10 possibles clústers
for (i in 1:10) {t
  km.out <- kmeans(data_normalized, centers = i, nstart = 20)

```

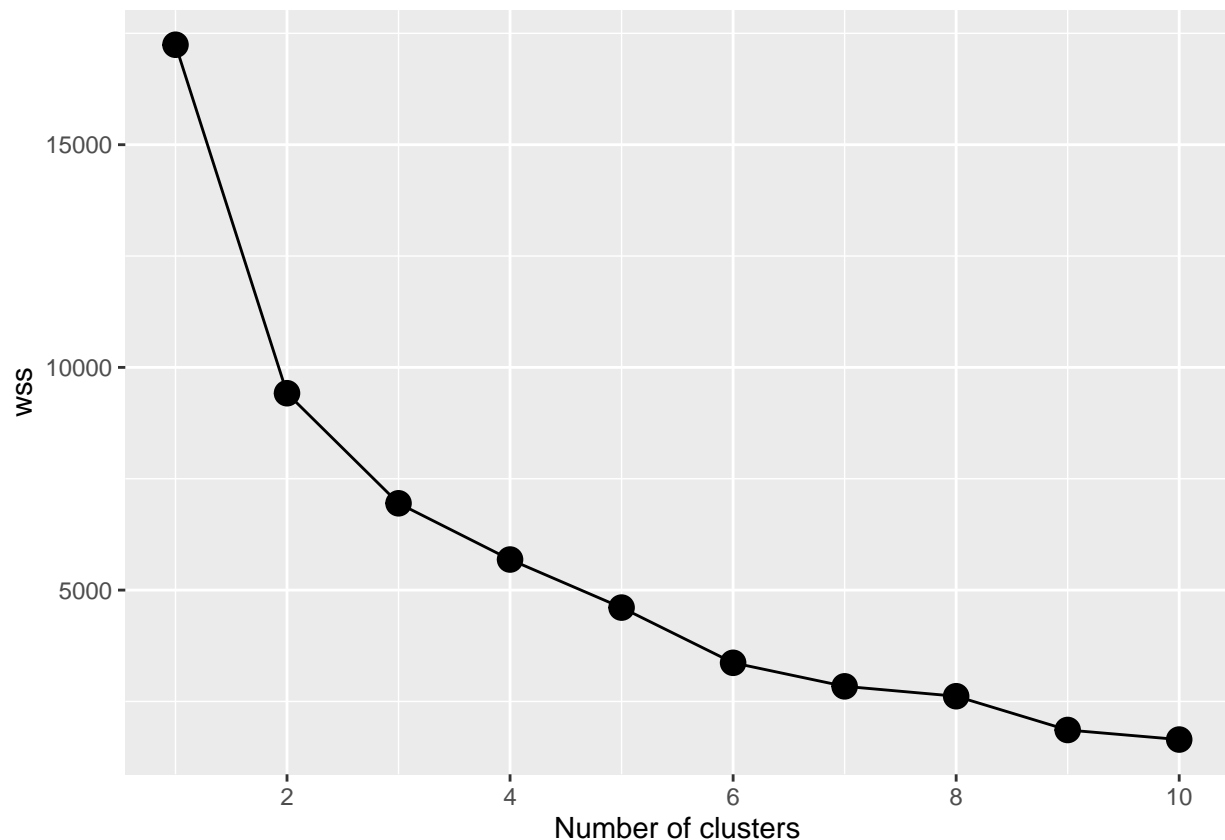
```
wss[i] <- km.out$tot.withinss
}
```

```
## Warning: did not converge in 10 iterations
```

```
# Generem el gràfic:
```

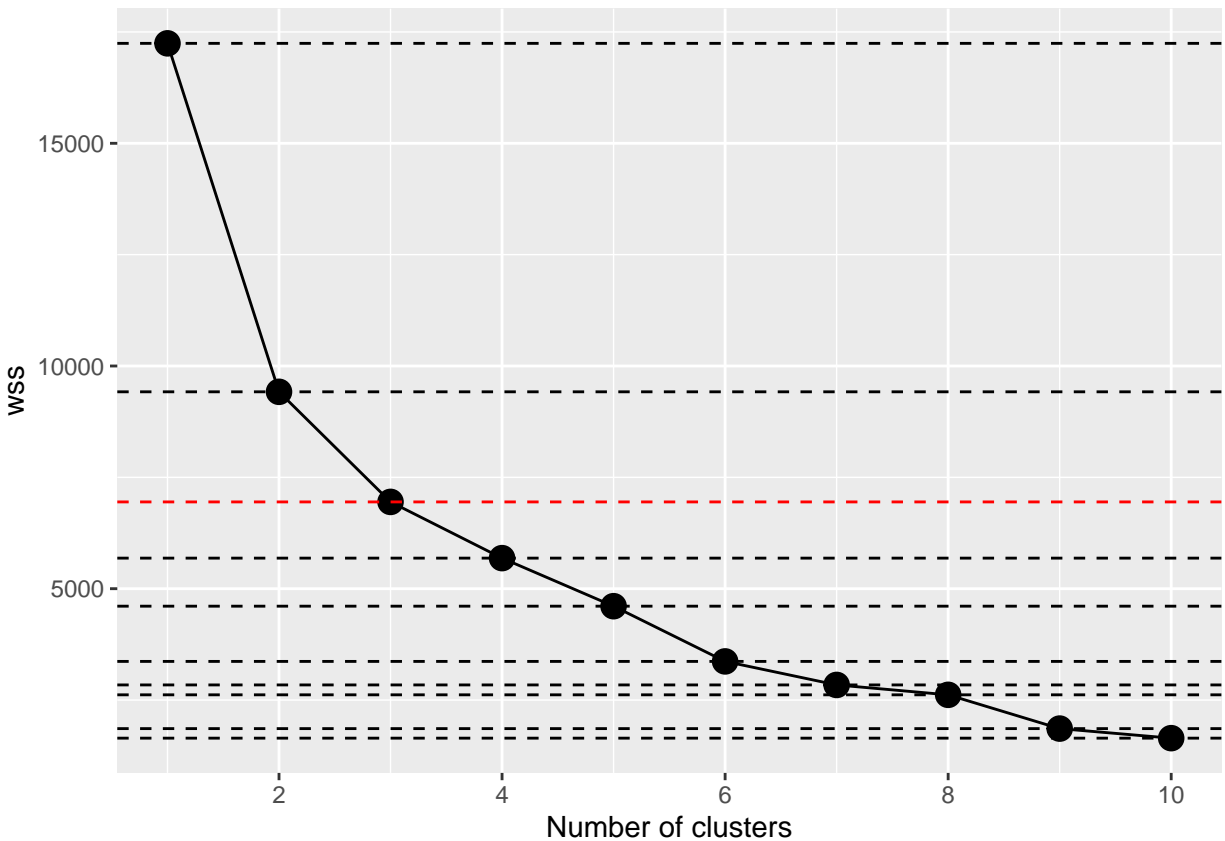
```
wss_df <- tibble(clusters = 1:10, wss = wss)
```

```
scee_plot <- ggplot(wss_df, aes(x = clusters, y = wss, group = 1)) +
  geom_point(size = 4) +
  geom_line() +
  scale_x_continuous(breaks = c(2, 4, 6, 8, 10)) +
  xlab('Number of clusters')
scee_plot
```



Aquest gràfic ens ajuda a determinar el nombre de clústers apropiats. Podem veure que a partir el colze de la gràfica es troba a 3, a partir d'aquest valor el decreixement és molt més pausat, si marquem diferents línies en cada punt la visualització d'aquesta diferència resulta més evident:

```
scee_plot +
  geom_hline(
    yintercept = wss,
    linetype = 'dashed',
    col = c(rep('#000000', 2), '#FF0000', rep('#000000', 7))
  )
```



Font: Pàgina web Datacamp: Tutorial K-Means clusterin in R. Obtingut en l'enllaç: <https://www.datacamp.com/tutorial/k-means-clustering-r> (Consultat per última vegada el 6 de novembre de 2024)

Si fem el anàlisi de clústers, dividint per 3 clústers diferents:

```
fit <- kmeans(data_normalized, 3) # Especifiquem el nombre de clústers a 3
# Obtenim les mitges dels clústers
aggregate(data_normalized, by=list(fit$cluster), FUN=mean)
```

```
##   Group.1  M1_1_MSS  M1_2_MSS  M5_1_MSS  M5_2_MSS  T49_1_MSS  T49_2_MSS
## 1      1 -0.1214932 -0.09405126 -0.1174616 -0.1126726 -0.1188879 -0.1259818
## 2      2  3.5797210  2.19047757  3.2487664  2.9400771  3.4453185  3.6972735
## 3      3  8.6397067  9.88199191  9.5199211 10.1010709  8.7714547  9.0397452
##      M42_1_PD  M42_2_PD  M43_1_PD  M43_2_PD  M64_1_PD  M64_2_PD
## 1 -0.1072008 -0.1047481 -0.09556473 -0.07675705 -0.09557378 -0.08772582
## 2  2.9225501  2.9794550  2.11438494  1.30279161  2.02538794  1.43894104
## 3  8.9216207  8.0367577 10.65339219 10.73183163 11.14498647 12.54056176
```

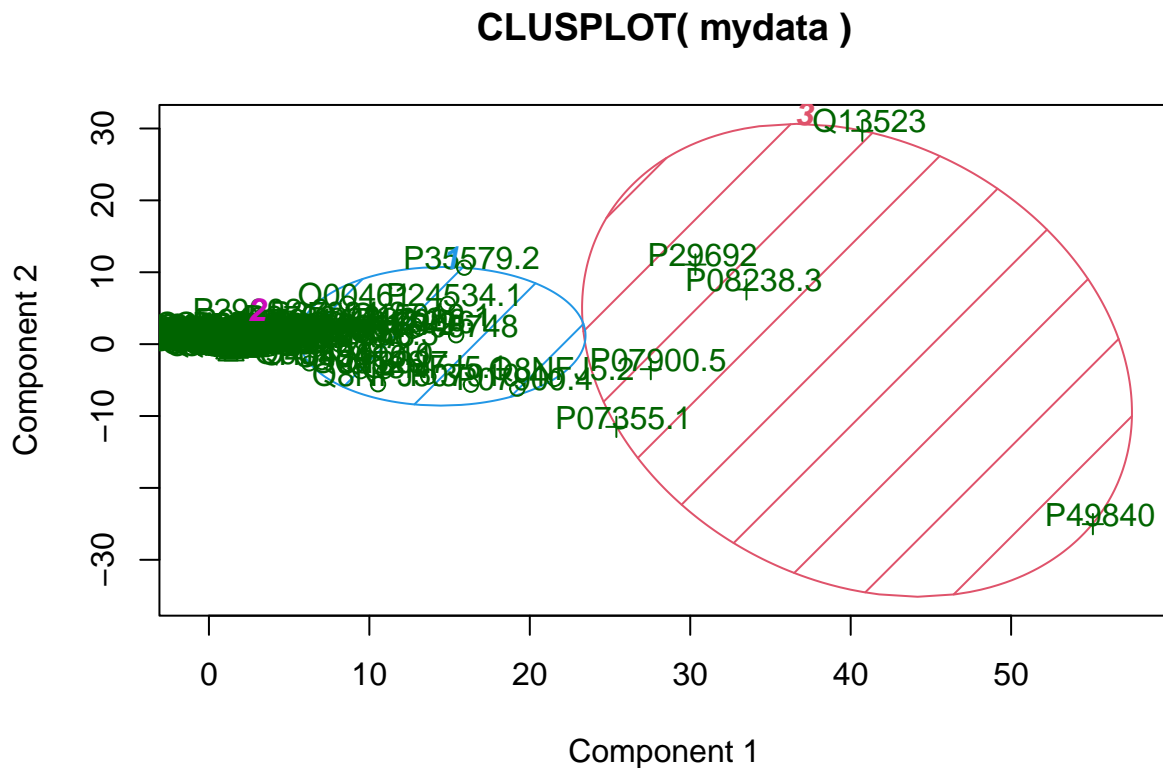
```
# Ho introduïm a un dataframe denominat mydata
mydata <- data.frame(data_normalized, fit$cluster)
```

Podem visualitzar aquesta informació gràficament de la manera següent:

```
# K-Means Clustering amb 3 clústers
fit <- kmeans(mydata, 3)
```

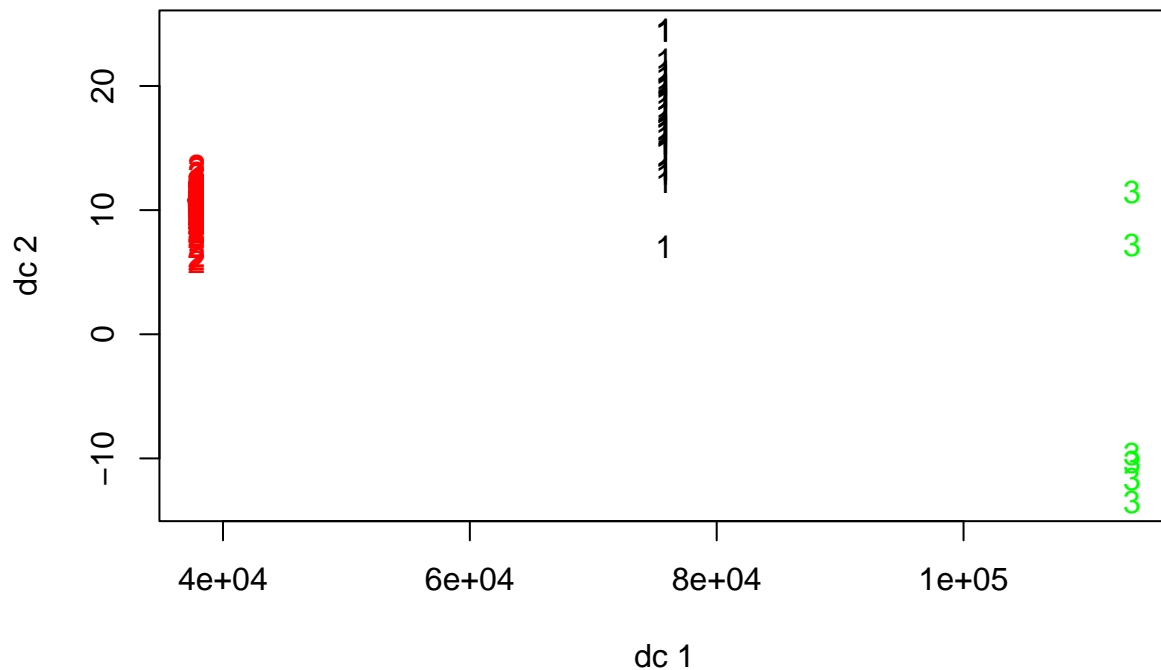
```
# Cluster Plot contra el 1r i el 2n PC:
library(cluster)

clusplot(mydata, fit$cluster, color=TRUE, shade=TRUE,
  labels=2, lines=0)
```



These two components explain 81.49 % of the point variability.

```
library(fpc)
plotcluster(mydata, fit$cluster)
```

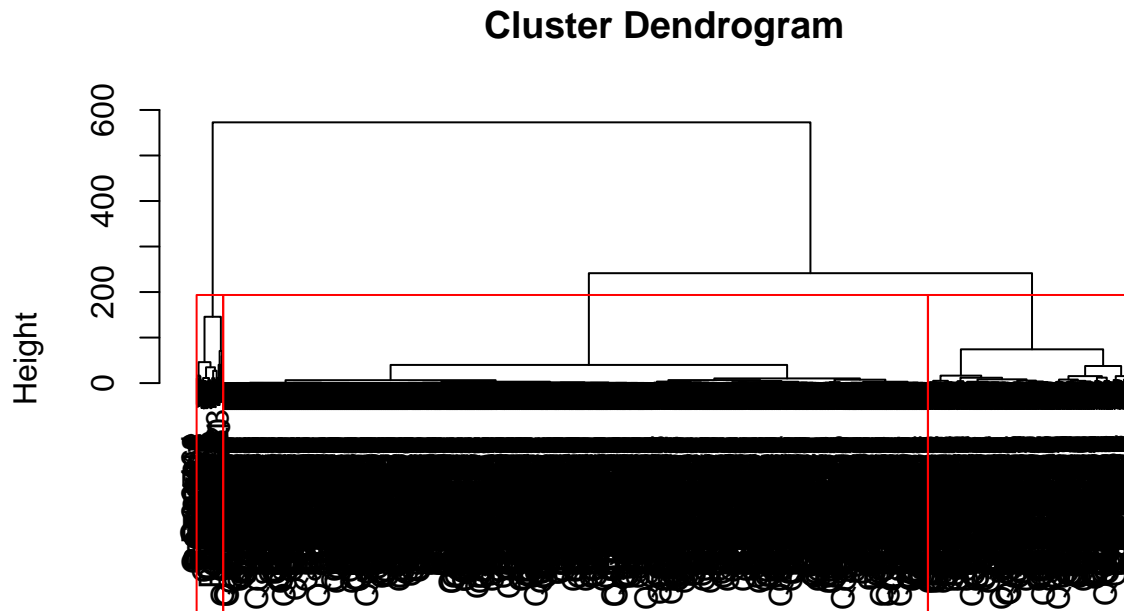



Podem veure que hi ha determinats pèptids que són agrupats de forma conjunta: Q13523, P29692, P08238.3, P07900.4, P07335.1, P4984. Podria resultar interessant estudiar la correlació d'aquests fosfopèptids, així com els que integren el 2n clúster.

A continuació podem visualitzar els clústers amb dendrogrames, per exemple aquí visualitzem un dendrograma jeràrquic generat amb el mètode euclidià.

```
# Clustering jeràrquic
d <- dist(mydata,
method = "euclidean")
fit <- hclust(d, method="ward.D")

plot(fit) # Visualitzar el dendrograma
groups <- cutree(fit, k=3) # Dividir l'arbre en 3 clústers
# Dibuixar en el dendrograma quadres vermells al voltant dels 3 clústers
rect.hclust(fit, k=3, border="red")
```



```
d
hclust (*, "ward.D")
```

El resultat dels clústers generats resulta similar, no obstant la interpretació amb aquesta visualització resulta menys pràctica que la dels clústers sobre el PCA.

4.2.4 Determinació d'accessions enriquides segons la condició

Per tal de determinar fosfopèptids que estan diferencialment enriquits utilitzarem el paquet 'limma':

```
library(BiocManager)
#install("limma")
library(limma)
```

```
##
## Adjuntando el paquete: 'limma'

## The following object is masked from 'package:BiocGenerics':
##
## plotMA
```

Per tal de utilitzar limma necessitem una matriu de disseny que descrigui les variables i una on es donin les comparacions.

Creem la matriu de les variables on establim a quin fenotip correspon cada mostra:

```
#Generem la matriu de disseny que descriu el fenotip de les mostres:
targets <- as.data.frame(targets)
groups <- as.factor(targets$Phenotype)
designMat <- model.matrix(~ -1 + groups)
show(designMat) #Visualitzem la matriu de disseny
```

```
##      groupsMSS groupsPD
## 1           1         0
## 2           1         0
## 3           1         0
## 4           1         0
## 5           1         0
## 6           1         0
## 7           0         1
## 8           0         1
## 9           0         1
## 10          0         1
## 11          0         1
## 12          0         1
## attr(,"assign")
## [1] 1 1
## attr(,"contrasts")
## attr(,"contrasts")$groups
## [1] "contr.treatment"
```

Per tal d'interpretar les rèpliques tècniques de la mateixa mostra fem servir la funció `duplicateCorrelation` que ens dona la correlació mitja entre replicats:

```
correpr <- duplicateCorrelation(matrix, designMat, block=targets$Individual) #Determinem la correlació m
correpr$consensus.correlation
```

```
## [1] 0.8770936
```

Generem una matriu de contrast on veure les comparacions:

```
require(limma)
contMat <- makeContrasts(mainEff=groupsPD-groupsMSS, levels=designMat) #Generem la matriu
show(contMat) # La visualitzem
```

```
##           Contrasts
## Levels      mainEff
## groupsMSS      -1
## groupsPD       1
```

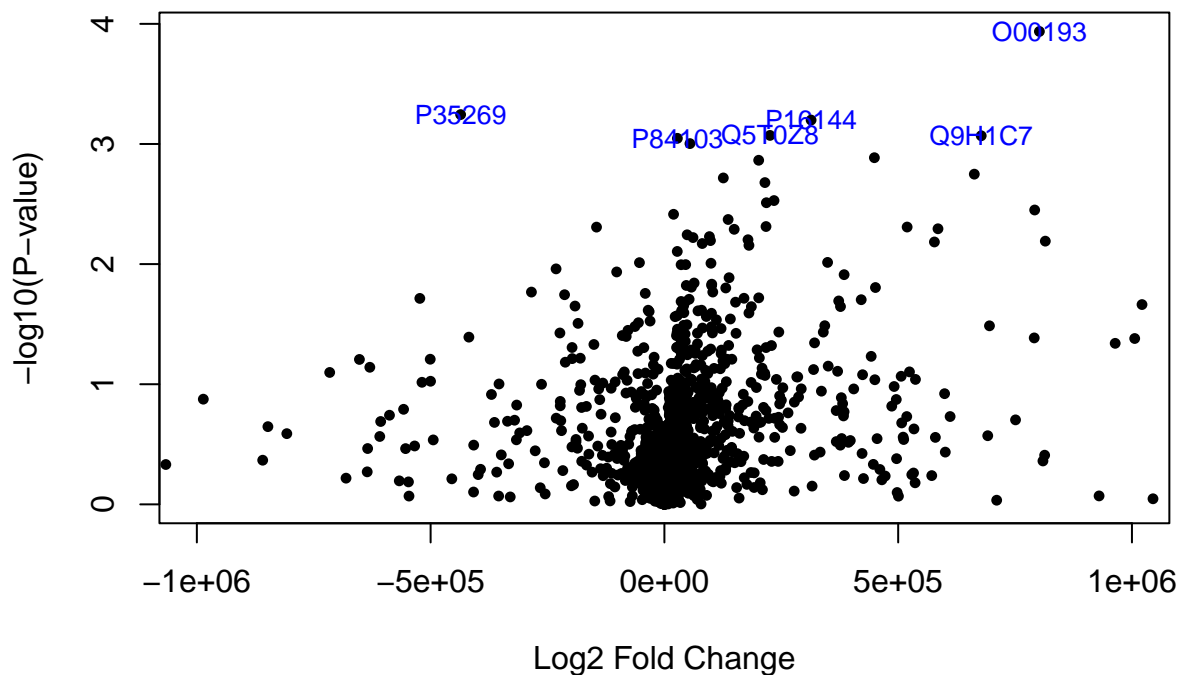
Ara que tenim les dos matrius podem comparar les abundàncies de pèptids per cada fenotip de mostra:

```
model.comparison <- lmFit(matrix, designMat, block=targets$Individual, correlation=correpr$consensus)
comparison <- contrasts.fit(model.comparison, contMat)
comparison <- eBayes(comparison)
results <- topTable(comparison, adjust="BH", number=nrow(matrix))
head(results) #Visualizem solament les primeres dades
```

##	ID	logFC	AveExpr	t	P.Value	adj.P.Val	B
## 1323	O00193	802048.92	641359.81	5.905578	0.0001157082	0.1663884	-4.59512
## 454	P35269	-435826.87	412752.57	-4.836639	0.0005694049	0.2040053	-4.59512
## 410	P16144	313748.10	270927.30	4.770056	0.0006322067	0.2040053	-4.59512
## 86	Q5T0Z8	225849.08	176503.85	4.584651	0.0008488023	0.2040053	-4.59512
## 226	Q9H1C7	677600.13	446716.73	4.580633	0.0008542838	0.2040053	-4.59512
## 146	P84103	27775.59	18777.87	4.550093	0.0008971889	0.2040053	-4.59512

Tal com podem veure en el resultat, obtenim els fosfopèptids ordenats de més diferencialment expressat a menys. Podem visualitzar aquest resultat en un volcano plot, on en l'eix vertical tindrem la diferència estadística i en l'horitzontal la diferència d'abundància entre fenotip MSS i PD, sent les mostres de la dreta els fosfopèptids enriquits en el fenotip PD mentre que en l'esquerra del zero tenim els enriquits en el fenotip MSS:

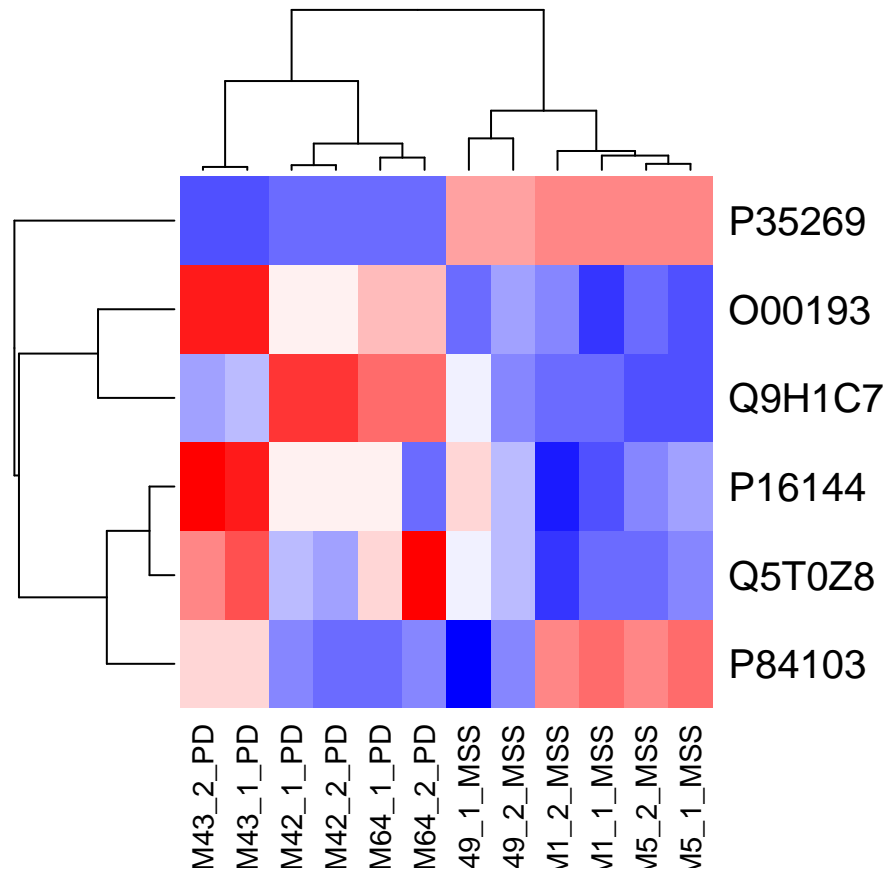
```
#Visualitzem en el volcano plot les 6 accessions amb un p-value més reduït
volcanoplot(comparison, highlight=6, names=rownames(matrix), cex=0.75,
            xlim=c(-1e+06,1e+06))
```



Si seleccionem aquests 6 pèptids i visualitzem la seva quantitat amb un heatmap veiem que efectivament la seva expressió és diferencial dependent del fenotip de la mostra:

```
#Fem una matriu que solament contingui els 6 pèptids seleccionats:
interestedacc<-c(1323,454, 410, 86, 226,146)
subaccesion<- matrix[interestedacc, , drop=FALSE]
norm.subacc<-scale(subaccesion) #Escalem les dades i en fem un heatmap:
```

```
col<- colorRampPalette(c("blue", "white", "red"))(20)
heatmap(x = norm.subacc, col=col)
```



Entre els pèptids seleccionats destaquem P35269 que es troba molt més en les mostres MSS respecte a les PD, i el pèptid O00193 es troba en el cas contrari, sent més upregulat en les mostres de fenotip PD.

5 DISCUSSIÓ, LIMITACIONS I CONCLUSIONS DE L'ESTUDI

L'estudi de diferències fosfopeptídiques entre mostres de PDX pot ser una forma interessant d'encarar patrons i vies de senyalització diferencialment expressades en diferents malalties. Aquest estudi, no obstant no ens permet discernir amb massa exactitud perfils clars entre les mostres. Majoritàriament, això és degut al fet que la qualitat de les dades és baixa, una evidència d'aquest fet és la gran variabilitat mostral i la presència de replicats tècnics amb perfils peptídics diferents entre ells. Per aquest motiu seria necessari considerar reanalitzar algunes dades, com és el cas de la mostra M1. No obstant així, hem pogut evidenciar que hi ha certa diferència entre els fenotips de PDX PD i MSS a nivell fosfopeptídic, especialment dos fosfopèptids amb número d'accessió P35269 i O00193.