

# Análisis de datos ómicos

Joaquín Villegas

2025-03-27

## Tabla de contenidos

El estudio que he utilizado consiste en un análisis fosfoproteómico cuyos resultados es un set de abundancias normalizadas de señales de espectrometría de masas para más de 1000 fosfopéptidos. A continuación muestro un resumen del contenido del dataset y la metadata una vez preprocesados.

```
library(readxl)
dataset <- read_excel("dataset1.xlsx")
metadata_muestras <- read_excel("metadatumuestra.xlsx")
metadata_carac <- read_excel("metadatacarac.xlsx")
library(readr)
dataset <- as.data.frame(dataset)
rownames(dataset) <- make.unique(dataset[,1]) #Ajustamos la dimensión de los datos
#para manipularlos posteriormente
dataset <- dataset[, -1]
```

```
dim(dataset)
```

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

```
names(dataset)
```

```
## [1] "M1_1" "M1_2" "M5_1" "M5_2" "T49_1" "T49_2" "M42_1" "M42_2" "M43_1"
## [10] "M43_2" "M64_1" "M64_2"
```

```
head(dataset, 5)
```

```
##
## LYPELSQYMGLSLNEEEIR[2] Phospho| [9] Oxidation      24.29438  44475.964
## VDKVIQAQTAFSANPANPAILSEASAPIPHDGNLYPR[35] Phospho    0.00000  43138.904
## VIQAQTAFSANPANPAILSEASAPIPHDGNLYPR[32] Phospho    3412.60332 172143.040
## HADAEMTGYVVTR[6] Oxidation| [9] Phospho    220431.17880 145656.887
## HADAEMTGYVVTR[9] Phospho    18254.77813   8529.755
##
## M5_1      M5_2
## LYPELSQYMGLSLNEEEIR[2] Phospho| [9] Oxidation      0.000   6269.141
## VDKVIQAQTAFSANPANPAILSEASAPIPHDGNLYPR[35] Phospho   2102.056  50355.051
## VIQAQTAFSANPANPAILSEASAPIPHDGNLYPR[32] Phospho   77323.019 307637.429
## HADAEMTGYVVTR[6] Oxidation| [9] Phospho   104287.815  75887.365
## HADAEMTGYVVTR[9] Phospho    35955.901  44102.316
##
## T49_1     T49_2
## LYPELSQYMGLSLNEEEIR[2] Phospho| [9] Oxidation   1135.8169 21933.90
## VDKVIQAQTAFSANPANPAILSEASAPIPHDGNLYPR[35] Phospho   248.9275  3239.16
## VIQAQTAFSANPANPAILSEASAPIPHDGNLYPR[32] Phospho   98442.2773 192982.37
## HADAEMTGYVVTR[6] Oxidation| [9] Phospho   773377.4981 481165.54
## HADAEMTGYVVTR[9] Phospho    57145.1682  34638.01
```

```
##
## LYPELSQYMGLSLNEEEIR[2] Phospho|[9] Oxidation      M42_1      M42_2
## VDKVIQAQTAFSANPANPAILSEASAPIPHDGNLYPR[35] Phospho 1315.904      0.00
## VIQAQTAFSANPANPAILSEASAPIPHDGNLYPR[32] Phospho 24851.344 16547.95
## HADAEMTGYVVTR[6] Oxidation|[9] Phospho 1027196.292 1163747.38
## HADAEMTGYVVTR[9] Phospho 21231.256 49499.70
##
## LYPELSQYMGLSLNEEEIR[2] Phospho|[9] Oxidation      M43_1      M43_2
## VDKVIQAQTAFSANPANPAILSEASAPIPHDGNLYPR[35] Phospho 772.9056 2136.746
## VIQAQTAFSANPANPAILSEASAPIPHDGNLYPR[32] Phospho 0.0000 0.000
## HADAEMTGYVVTR[6] Oxidation|[9] Phospho 5565.2821 0.000
## HADAEMTGYVVTR[9] Phospho 4080239.1820 4885818.113
## HADAEMTGYVVTR[9] Phospho 666107.0448 379313.615
##
## LYPELSQYMGLSLNEEEIR[2] Phospho|[9] Oxidation      M64_1      M64_2
## VDKVIQAQTAFSANPANPAILSEASAPIPHDGNLYPR[35] Phospho 1820.724 1727.9098
## VIQAQTAFSANPANPAILSEASAPIPHDGNLYPR[32] Phospho 0.000 892.3565
## HADAEMTGYVVTR[6] Oxidation|[9] Phospho 5901.9577
## HADAEMTGYVVTR[9] Phospho 3093786.793 2759104.5440
## HADAEMTGYVVTR[9] Phospho 255792.117 579765.0018
```

```
dim(metadata_muestras)
```

```
## [1] 12 4
```

```
names(metadata_muestras)
```

```
## [1] "Sample...1" "Sample...2" "Individual" "Phenotype"
```

```
head(metadata_muestras, 5)
```

```
## # A tibble: 5 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
```

```
dim(metadata_carac)
```

```
## [1] 1438 6
```

```
names(metadata_carac)
```

```
## [1] "SequenceModifications" "Accession"      "Description"
## [4] "Score"                "CLASS"          "PHOSPHO"
```

```
head(metadata_carac, 5)
```

```
## # A tibble: 5 x 6
##   SequenceModifications Accession Description Score CLASS PHOSPHO
##   <chr>                <chr>      <chr>      <dbl> <chr> <chr>
## 1 LYPELSQYMGLSLNEEEIR[2] Phospho|[9] ~ 000560 Syntenin-1~ 48.1 H Y
## 2 VDKVIQAQTAFSANPANPAILSEASAPIPHDGNLY~ 000560 Syntenin-1~ 67.0 H Y
## 3 VIQAQTAFSANPANPAILSEASAPIPHDGNLYPR[~ 000560 Syntenin-1~ 77.7 H Y
## 4 HADAEMTGYVVTR[6] Oxidation|[9] Phos~ 015264 Mitogen-ac~ 44.9 H Y
## 5 HADAEMTGYVVTR[9] Phospho 015264 Mitogen-ac~ 67.4 H Y
```

## Abstract

## Objetivos

El objetivo general de este análisis es familiarizarse con el uso de SummarizedExperiment, el workflow general del análisis de los datos ómicos, así como el uso de GitHub y Git como herramienta de control de versiones. El objetivo conceptual que atañe al dataset escogido es la búsqueda de alguna diferencia significativa entre los fosfopéptidos que permita la diferenciación de dos grupos tumorales.

## Métodos

Para facilitar la manipulación de los datos y metadatos del dataset se ha creado un objeto de clase SummarizedExperiment, extensión de ExpressionSet, de la siguiente forma:

```
library(SummarizedExperiment)
data_expset <- SummarizedExperiment(assays=list(counts=dataset),
colData=metadata_muestras, rowData=metadata_carac)
data_expset

## class: SummarizedExperiment
## dim: 1438 12
## metadata(0):
## assays(1): counts
## rownames(1438): LYPELSQYMGLSLNEEEIR[2] Phospho| [9] Oxidation
##   VDKVVIQAQTAFSANPANPAILSEASAPIPHDGNLYPR[35] Phospho ...
##   YQDEVFGGFVTEPQEESEEEVEEPEER[17] Phospho YSPSQNSPIHHIPSRR[1]
##   Phospho| [7] Phospho
## rowData names(6): SequenceModifications Accession ... CLASS PHOSPHO
## colnames(12): M1_1 M1_2 ... M64_1 M64_2
## colData names(4): Sample...1 Sample...2 Individual Phenotype
```

## Resultados

## Discusión

## Conclusiones

## Referencias

[https://github.com/joakovm/Villegas\\_Martinez\\_Joaquin\\_PEC1.git](https://github.com/joakovm/Villegas_Martinez_Joaquin_PEC1.git)