

PEC 1 Análisis de datos omicos.

Eva Díaz Hernández

Descarga de datos:

Siguiendo las instrucciones de la PEC accedí al repositorio de github: <https://github.com/nutrimetabolomics/metaboData/> donde accedí a la carpeta “Datasets” y posteriormente a “[2018-Phosphoproteomics](#)” descargando así el archivo XLSX “TIO2+PTYR-human-MSS+MSIvsPD” del cual témenos la siguiente información a partir de “description.md”:

The accompanying dataset has been obtained from a phosphoproteomics experiment that was performed to analyze (3 + 3) PDX models of two different subtypes using Phosphopeptide enriched samples. 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 columns can be omitted.

Manejo en R:

Estos datos se importaron en R usando la función `read_excel()` del cual tuvimos en cuenta las columnas *SequenceModification* y aquellas que hacían referencia a las distintas medidas convirtiendo todo ello en una matriz que nos permitiera crear un contenedor del tipo `SummarizedExperiment`.

Para ello definimos las muestras, grupos y replicas y creamos un dataframe con los metadatos como se indica en la figura 1

```

```{r}
Creamos la matriz
abundance_matrix <- as.matrix(data_mod)
```

```{r}
Definimos los nombres de muestra, grupos y replicados
sample_names <- colnames(abundance_matrix)
groups <- c(rep("MSS", 6), rep("PD", 6)) # seis de cada grupo
replicates <- rep(1:2, times = 6) # Dos replicas

Creamos un DataFrame con los metadatos
col_metadata <- DataFrame(
 sample = sample_names,
 Group = groups,
 Replicate = replicates)

row_metadata <- data %>%
 select(SequenceModifications, Accession, Description, Score, CLASS, PHOSPHO) %>%
 as.data.frame()

row_metadata <- DataFrame(row_metadata)
rownames(row_metadata) <- row_metadata$SequenceModifications
```

```{r}
Creamos el objeto SummarizedExperiment
se <- SummarizedExperiment(
 assays = list(counts = abundance_matrix),
 rowData = row_metadata,
 colData = col_metadata
)
```

```

Figura 1: Creación el contenedor tipo SummarizedExperiment

Tras crear el contenedor estudiamos la distribución de los datos, tanto con la función `summary()` como con un diagrama de cajas con la función `boxplot()` (Figura 2). Donde podemos ver que los datos se distribuyen de manera sesgada con outliers extremos indicando la necesidad de transformar estos datos al igual puede indicar que las muestras pueden ser muy heterogéneas, pudiendo indicar la necesidad de investigar si existen subgrupos o factores no considerados que estén afectando a la variabilidad.

De la misma manera realizamos una matriz de correlación con el método de Pearson y se representó con la función `ggplot()` para estudiar cómo se correlacionan entre si (Figura 3), pudiendo así observar que M5 tiene gran correlación entre replicas aunque otras como M2 no presentan una gran correlación entre sus réplicas.

Posteriormente, se calculó el coeficiente de variación de los fosopeptidos y se realizó un histograma (Figura 4) donde pudimos observar que en la mayoría la variación es de entre 0,5 y 1 aunque hay algunos con variaciones mayores, lo que indica una variabilidad moderada en relación con sus medias, habiendo que tener en cuenta esto a la hora de interpretar los resultados y estabilidad de los datos.

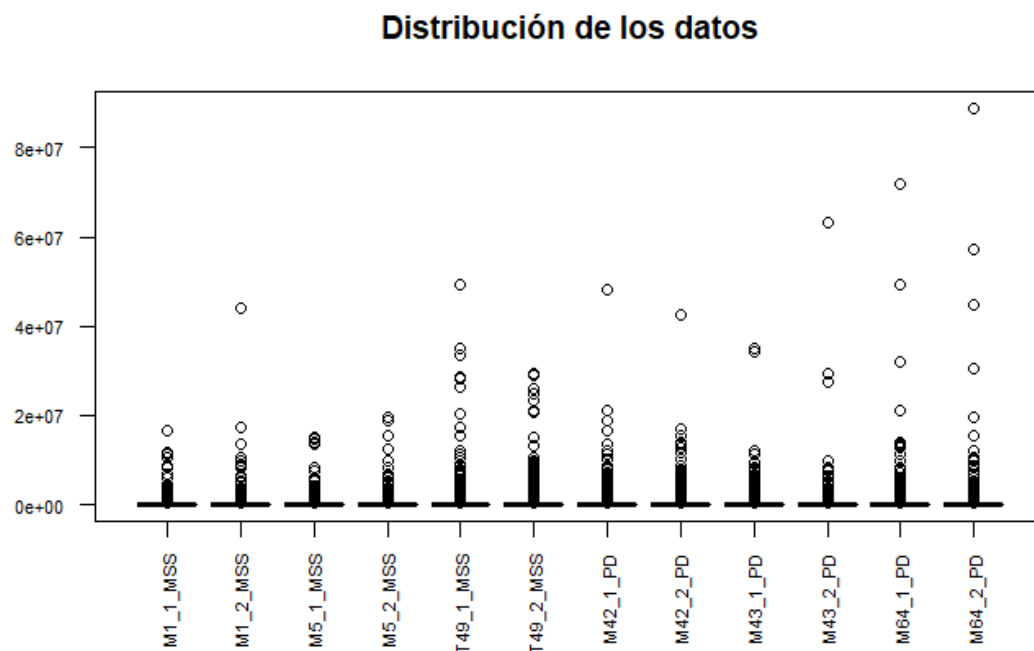


Figura 2: Diagrama de cajas para la distribución de los datos

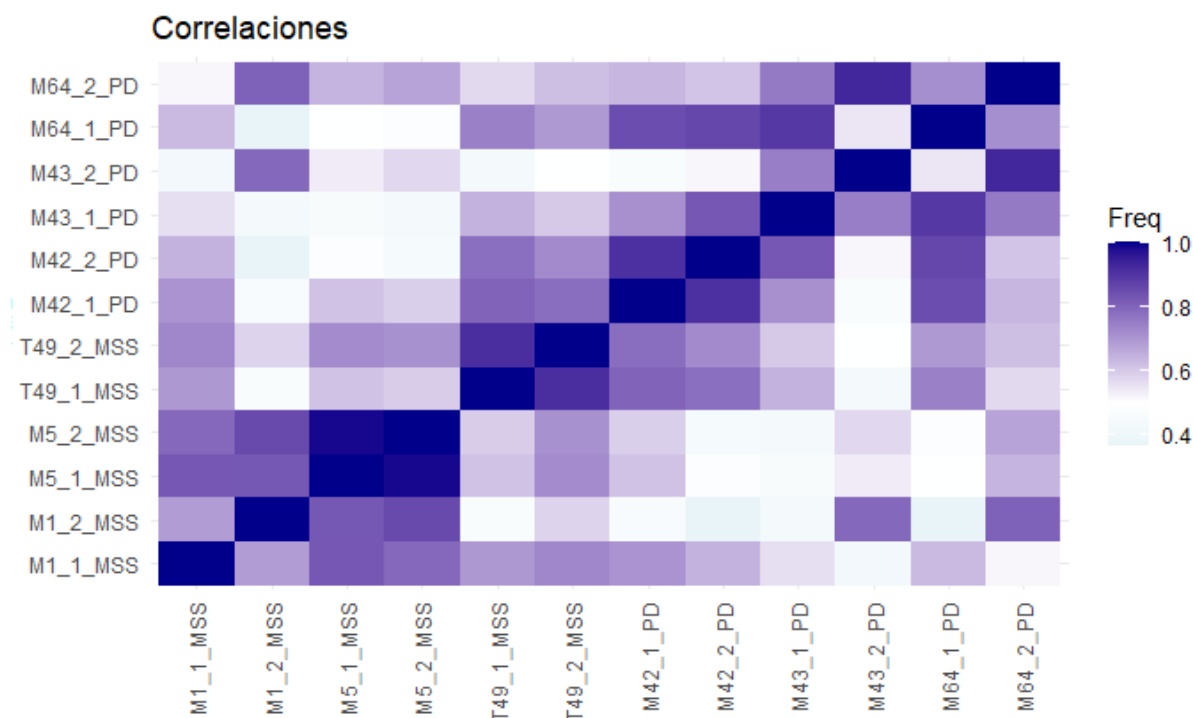


Figura 3: Estudio de correlaciones

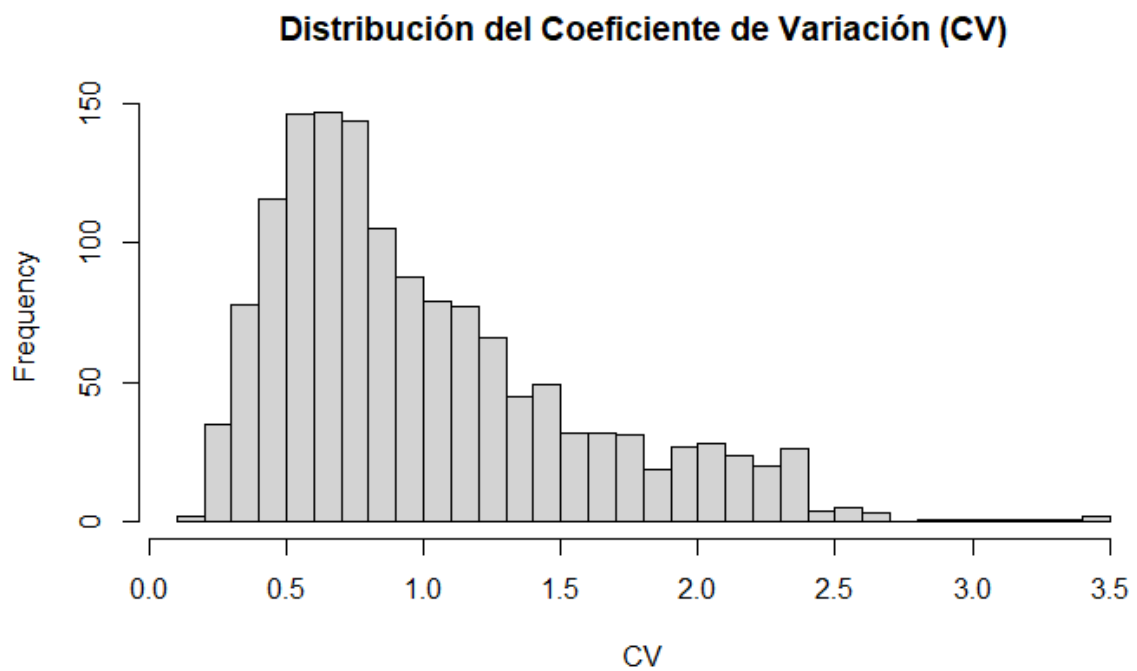


Figura 4: Histograma de los coeficientes de variación

Tras observar esto, se normalizaron los datos usando la función $\log_2()$ y vimos de nuevo el diagrama de cajas para ver como se distribuyen tras la transformación (Figura 5)

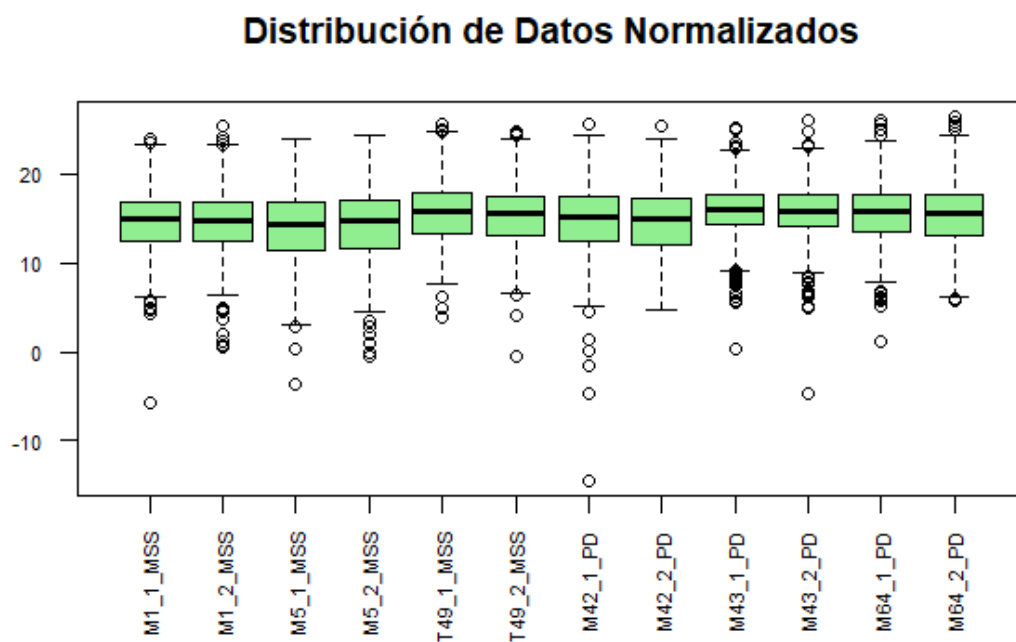


Figura 5: Diagrama de cajas de los datos normalizados

Observando en este caso una mejor distribución a pesar de que siguen habiendo muchos outliers.

Para analizar los datos usamos un ajuste de modelo lineal usando la librería lina, obteniendo así los siguientes resultados (Figura 6):

```

      (Intercept)      GroupPD
(Intercept)  0.1666667 -0.1666667
GroupPD      -0.1666667  0.3333333
$qr
      (Intercept)      GroupPD
M1_1_MSS    -3.4641016 -1.7320508
M1_2_MSS      0.2886751  1.7320508
M5_1_MSS      0.2886751  0.2240092
M5_2_MSS      0.2886751  0.2240092
T49_1_MSS     0.2886751  0.2240092
T49_2_MSS     0.2886751  0.2240092
M42_1_PD      0.2886751 -0.3533410
M42_2_PD      0.2886751 -0.3533410
M43_1_PD      0.2886751 -0.3533410
M43_2_PD      0.2886751 -0.3533410
M64_1_PD      0.2886751 -0.3533410
M64_2_PD      0.2886751 -0.3533410
attr(,"assign")
[1] 0 1
attr(,"contrasts")
attr(,"contrasts")$Group
[1] "contr.treatment"

$graux
[1] 1.288675 1.224009

$pivot
[1] 1 2

$tol
[1] 1e-07

$rank
[1] 2

attr(,"class")
[1] "qr"

```

Figura 6: Resultados del modelo lineal

Donde se observa que la varianza del coeficiente (Intercept) es 0.1666667, y la de GroupPD es 0.3333333, habiendo una covarianza negativa entre ambos, indicando una relación inversa. También podemos ver que las muestras MSS muestran valores positivos en esta columna (excepto el primer valor que es el pivote de -3.464), mientras que las de PD son negativos.

Tol es un umbral para la precisión en el cálculo; 1e-07 indica una alta precisión.

Para ver los resultados de una manera mas grafica se realizo un histograma para ver como se distribuyen los P-valores (Figura 7) viendo que en muchos fosfopeptidos hay diferencias significativas entre grupos.

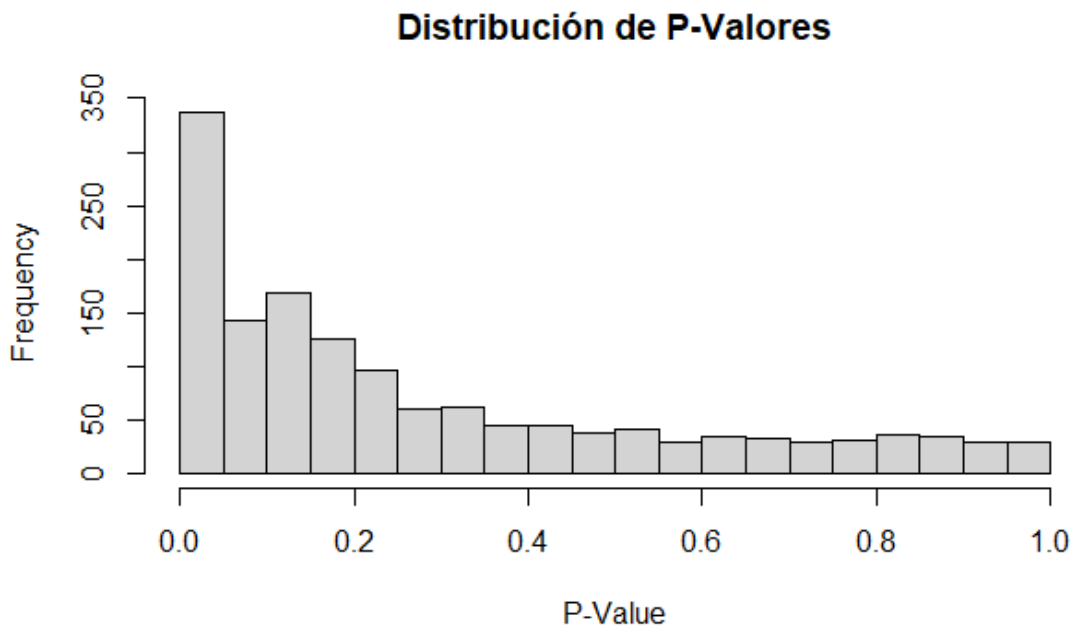


Figura 7: Histograma de p-valores

Para comprobar si el ajuste separo bien, se realizó un análisis de componentes principales (PCA) (Figura 8) viendo que ambos grupos están bien separados

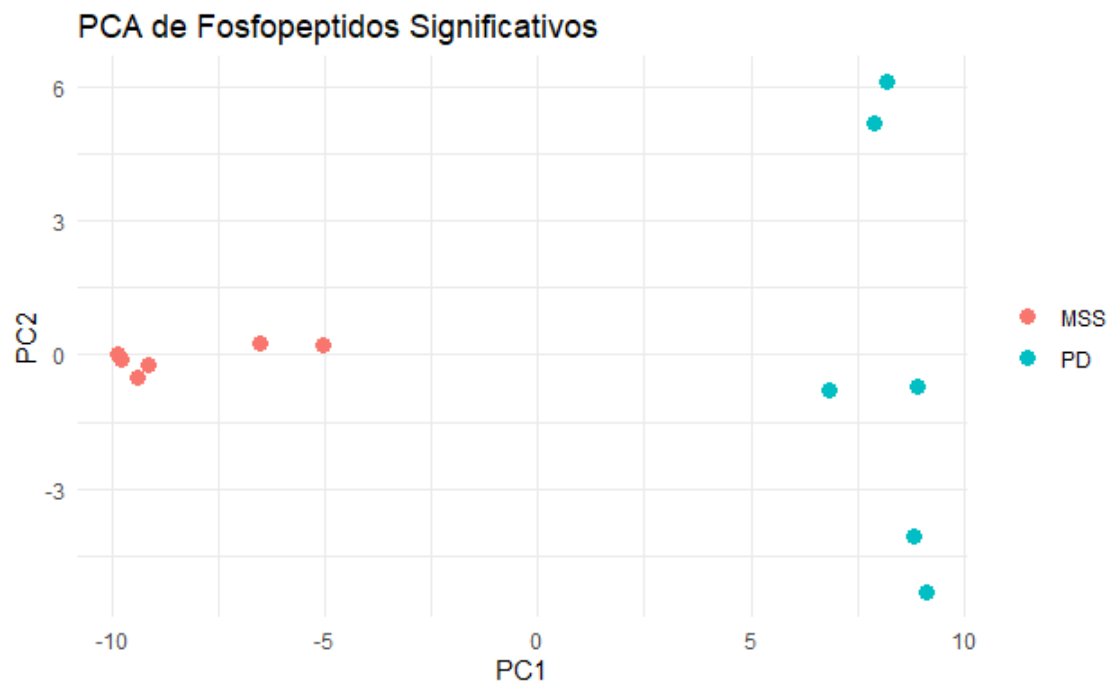


Figura 8: PCA para comprobar la separación de los grupos

Por ultimo se realizo una tabla que recogiera todos los fosfopeptidos que presentan diferencias significativas entre los grupos junto con el p-valor y el en que grupo su presencia es mayor (Tabla 1). Observando que en general el grupo MSS presenta mayor expresi3n de fosfopeptidos (en 90 de ellos, mientas PD solo tiene mayor expresi3n en 19)

Tabla 1: Resultados significativos del modelo

| | ID | MayorExpresion | adj.P.Val | |
|------|--|----------------|-----------|-------------|
| | | | | |
| 168 | GVGYETILK[4] Phospho | | MSS | 0.001259444 |
| 146 | AFGYYGPLR[4] Phospho | | MSS | 0.001259444 |
| 914 | KASPEPPDSAEGALK[3] Phospho | | MSS | 0.001259444 |
| 328 | KDPEDTGAEKSPTTSADLK[14] Phospho | | MSS | 0.001259444 |
| 454 | IHDLEDDLEMSSDASDASGEEGGRVPK[10] Oxidation [18] Phospho | | PD | 0.001469087 |
| 1238 | LDQPVSAPPSPR[10] Phospho | | MSS | 0.002007267 |
| 18 | PYQYPALTPEQK[4] Phospho | | MSS | 0.002561647 |
| 331 | GIPLATGDTSPEPELLPGAPLPPPK[9] Phospho | | MSS | 0.002901779 |
| 86 | HKAPGSADYGFAPAAGR[9] Phospho | | MSS | 0.004430436 |

| | ID | MayorExpresion | adj.P.Val | |
|----------|---|----------------|-----------------|--|
| | | | | |
| 112
2 | EDSGTFSLGK[3] Phospho | MSS | 0.00454446
8 | |
| 955 | NHSDSSTSESEVSSVSPLK[16] Phospho | MSS | 0.00461055
9 | |
| 871 | GLLAQGLRPESPPPAGPLLNGAPAGESPQPK[11] Phospho | MSS | 0.00461055
9 | |
| 571 | GKGGVTGSPEASISGSK[8] Phospho | MSS | 0.00461055
9 | |
| 575 | LPSGSGAASPTGSAVDIR[3] Phospho | MSS | 0.00461055
9 | |
| 935 | ASPSPQPSSQPLQIHR[4] Phospho | MSS | 0.00461055
9 | |
| 132
3 | SASPDDDLGSSNWEAADLGNEER[3] Phospho | MSS | 0.00504685
1 | |
| 143
5 | YLSFTPPEK[3] Phospho | MSS | 0.00504703
6 | |
| 213 | SFLDSGYR[7] Phospho | MSS | 0.00517586
5 | |
| 691 | EYVSNDAAQSDDEEKLQSQPTDTDGGR[10] Phospho | MSS | 0.00517586
5 | |
| 434 | DMESPTKLDVTLAK[2] Oxidation[4] Phospho | MSS | 0.00525798
0 | |
| 410 | LTAGVPDTPTR[8] Phospho | MSS | 0.00525798
0 | |

| | ID | MayorExpresion | adj.P.Val | |
|------|---|----------------|-------------|--|
| | | | | |
| 983 | SNSMLELAPK[3] Phospho[4] Oxidation | MSS | 0.005257980 | |
| 127 | AEDMYSAQSHQAATPPK[4] Oxidation[5] Phospho | MSS | 0.005257980 | |
| 613 | LDSSPSVSSTLAAK[4] Phospho | MSS | 0.005257980 | |
| 522 | TGQLEGAPAPGPAASPQTLDHSGATATGGASELK[15] Phospho | MSS | 0.005257980 | |
| 201 | NEENIYSVPHDSTQGK[7] Phospho | MSS | 0.005740956 | |
| 794 | LSVPTSDEEDEVPAKPR[6] Phospho | PD | 0.006537763 | |
| 209 | NVYYELNDVR[3] Phospho | MSS | 0.008367386 | |
| 1129 | ENPPVEDSSDEDDKR[9] Phospho | PD | 0.010085221 | |
| 1023 | TPPSTTVGSHSPPETPVLTR[11] Phospho | MSS | 0.011445579 | |
| 826 | SPSPAHLPDDPK[3] Phospho | MSS | 0.011603328 | |
| 431 | YGPADVEDTTGSGATDSKDDDDIDLFGSDDEEESEEAKR[28] Phospho | PD | 0.011638656 | |
| 446 | KPATPAEDDEDDDDIDLFGSDNEEEDKEAAQLR[19] Phospho | PD | 0.012042371 | |

| | ID | MayorExpresion | adj.P.Val | |
|----------|--------------------------------------|----------------|-----------------|--|
| 106
8 | KPVTVSPTTPTSPTEGEAS[12] Phospho | MSS | 0.01402555
3 | |
| 538 | RPEPSTLENITDDKYATVSSPSK[20] Phospho | MSS | 0.01434811
6 | |
| 569 | GGVTGSPEASISGSK[6] Phospho | MSS | 0.01544584
4 | |
| 116
9 | GSGTASDDEFENLR[6] Phospho | PD | 0.01590196
5 | |
| 226 | TTVYVVEDQR[4] Phospho | MSS | 0.01615946
9 | |
| 699 | ESTPSEEPSPPHNSK[9] Phospho | MSS | 0.01631565
9 | |
| 227 | TTVYVVEDQRR[4] Phospho | MSS | 0.01720631
5 | |
| 548 | SAESPTSPVTSETGSTFK[4] Phospho | MSS | 0.01761359
2 | |
| 131
2 | RSASPDDDLGSSNWEAADLGNEER[4] Phospho | MSS | 0.01766584
6 | |
| 452 | GEPAAAAAPEAGASPVEK[14] Phospho | MSS | 0.01766584
6 | |
| 94 | NSHTDNVSYEHSFNK[9] Phospho | MSS | 0.01804210
8 | |
| 979 | AEVPGATGGDSPHLQPAEPPGEPR[11] Phospho | MSS | 0.01848451
3 | |

| | ID | MayorExpresion | adj.P.Val | |
|------|---|----------------|-------------|--|
| 683 | CTLPEHESPSQDISDACEAESTER[1] Carbamidomethyl[8] Phospho[17] Carbamidomet hyl | MSS | 0.018484513 | |
| 1191 | IEDSEPHIPLIDDTDAEDDAPTKR[4] Phospho | MSS | 0.018696840 | |
| 577 | SSKASLGSLEGEAEAEASSPK[1] Phospho[2] Phospho | MSS | 0.018859726 | |
| 793 | KLSVPTSDEEDEVPAKPR[7] Phospho | PD | 0.021002641 | |
| 1397 | TQPDGTSVPGEPASPISQR[14] Phospho | MSS | 0.021002641 | |
| 953 | FTDKDQQPSGSEGEDDDAEAAALKK[9] Phospho | PD | 0.021002641 | |
| 402 | VADAKGDSESEEDLEVPVPSR[8] Phospho | PD | 0.021002641 | |
| 1436 | YNLDASEEEDSNK[6] Phospho | MSS | 0.021002641 | |
| 710 | ALELPLAASSIPRPTSPESHR[16] Phospho[18] Phospho | MSS | 0.021002641 | |
| 951 | SPGHMVILDQTK[1] Phospho[5] Oxidation | MSS | 0.021002641 | |
| 1438 | YSPSQNSPIHHPSRR[1] Phospho[7] Phospho | MSS | 0.021258717 | |
| 21 | AYTNFDAER[2] Phospho | MSS | 0.021674263 | |

| | ID | MayorExpresion | adj.P.Val | |
|------|---|----------------|-------------|--|
| | | | | |
| 130 | SSPEQSYQGDMYPTR[12] Phospho | MSS | 0.024819654 | |
| 457 | KGAGDGSDEEVDGKADGAEAKPAE[7] Phospho | PD | 0.024948620 | |
| 981 | SHSQASLAGPGVPDPSNR[3] Phospho | MSS | 0.025645202 | |
| 1025 | AHSPASTLPNSPGSTFER[3] Phospho | MSS | 0.026144807 | |
| 445 | LKTEKEPDATPPSPR[13] Phospho | PD | 0.026944370 | |
| 131 | SSPEQSYQGDMYPTR[7] Phospho [11] Oxidation | MSS | 0.027503218 | |
| 403 | LDIDSPBITAR[5] Phospho | MSS | 0.027503218 | |
| 304 | ESSPEKEAEEGCPEK[3] Phospho [12] Carbamidomethyl | MSS | 0.030251906 | |
| 1291 | RDSSSQLASTESDKPTTGR[3] Phospho | MSS | 0.031021745 | |
| 766 | IHQDSESGDELSSSTEQIR[7] Phospho | MSS | 0.031164388 | |
| 562 | EGMNPSYDEYADSDQHDAYLER[3] Oxidation [13] Phospho | PD | 0.031164388 | |
| 414 | SPVGSGAPQAAAPAPAAHVAGNPGGDAAPAATGTAAAASLATAAGSEDAEKK[1] Phospho | MSS | 0.031504192 | |

| | ID | MayorExpresion | adj.P.Val | |
|----------|---|----------------|-----------|-----------------|
| | | | | |
| 58 | FAGDKGYLTK[7] Phospho | | MSS | 0.03320952
2 |
| 142
4 | WAHDKFSGEEGEIEDDESGTENREEK[18] Phospho | | MSS | 0.03389076
9 |
| 584 | YYSPCEEHPAETNQNEGSESGTIR[3] Phospho [5] Carbamidomethyl | | MSS | 0.03392703
5 |
| 578 | SSKASLGSLGEAEAEASSPK[5] Phospho [8] Phospho | | MSS | 0.03659268
3 |
| 4 | HADAEMTGYVVTR[6] Oxidation [9] Phospho | | MSS | 0.03687474
2 |
| 66 | HIDRNYEPLKTQPK[6] Phospho | | PD | 0.03690287
5 |
| 22 | LSLEGDHSTPPSAYGSVK[14] Phospho | | MSS | 0.03771230
6 |
| 177 | IGEGTYGVVYK[10] Phospho | | MSS | 0.03771230
6 |
| 64 | VKGEYDMTVPK[5] Phospho [7] Oxidation | | MSS | 0.03771230
6 |
| 300 | EVAENQQNQSSDPREEKGSQPPPAESQSSLR[11] Phospho | | PD | 0.03858258
7 |
| 306 | VFDDESDEKEDEEYADEK[6] Phospho | | MSS | 0.03858258
7 |
| 453 | GNSRPGTPSAEGGSTSSTLR[7] Phospho | | MSS | 0.03858258
7 |

| | ID | MayorExpresion | adj.P.Val | |
|----------|---|----------------|-------------|--|
| | | | | |
| 110 | GSHFFPGNNVIYEK[12] Phospho | MSS | 0.038582587 | |
| 142
2 | WAHDKFSGEEGEIEDDESGTENR[7] Phospho | MSS | 0.038582587 | |
| 638 | SPAVATSTAAPPPSSPLPSK[16] Phospho | MSS | 0.038582587 | |
| 756 | RPPSPDVIVLSDNEQPSSPR[4] Phospho[[18] Phospho | MSS | 0.039392574 | |
| 117
9 | GVSMPNMLEPK[3] Phospho[[4] Oxidation[[7] Oxidation | MSS | 0.039392574 | |
| 559 | AIPVSPSAVEEDEDEDGHTVVATAR[5] Phospho | MSS | 0.039392574 | |
| 175 | HYGGLTGLNK[2] Phospho | MSS | 0.039392574 | |
| 666 | SEPVKEESSELEQPFAQDTSSVGPDR[21] Phospho | MSS | 0.039851882 | |
| 910 | TEAQDLCRASPEPPGPRESSR[7] Carbamidomethyl[[10] Phospho | PD | 0.039912300 | |
| 526 | RQVSASELHTSGILGPETLR[4] Phospho | PD | 0.039912300 | |
| 110
8 | DMDEPSPVPNVVEEVTLPK[2] Oxidation[[6] Phospho | MSS | 0.040033679 | |
| 141
2 | VGSLDNVGHLPAGGAVK[3] Phospho | MSS | 0.040722142 | |

| | ID | MayorExpresion | adj.P.Val | |
|----------|---|----------------|-----------------|--|
| | | | | |
| 172 | HTDDEMTGYVATR[6] Oxidation [9] Phospho | MSS | 0.04072214
2 | |
| 125
6 | LTFDSSFSPNTGKK[8] Phospho | PD | 0.04083577
0 | |
| 197 | LYDAYELK[5] Phospho | MSS | 0.04083577
0 | |
| 79 | YVLARP[1] Phospho | MSS | 0.04083577
0 | |
| 207 | NPGNQAAAYEHFETMK[8] Phospho [14] Oxidation | MSS | 0.04083577
0 | |
| 521 | KDPSGASNPSADSPLHR[13] Phospho | MSS | 0.04110040
8 | |
| 150 | APDFVIFYAPR[7] Phospho | MSS | 0.04236358
6 | |
| 442 | QKSDAEEDGGTVSQEEEDRKPK[13] Phospho | PD | 0.04383061
5 | |
| 345 | LDNTPASPPRSPAEPNDIPIAK[7] Phospho [11] Phospho | MSS | 0.04602823
9 | |
| 128
3 | QLSSGVSEIR[3] Phospho | MSS | 0.04602823
9 | |
| 134
3 | SLENETLNKEEDCHSPTSKPPKPDQPLK[13] Carbamidomethyl [18] Phospho | MSS | 0.04602823
9 | |
| 111
3 | DTSPDKGELVSDEEEDT[11] Phospho | MSS | 0.04882768
3 | |

| | ID | MayorExpresion | adj.P.Val |
|----------|---------------------------------------|----------------|-----------------|
| 116
5 | GPPQSPVFEGVYNNSR[5] Phospho | MSS | 0.04953036
5 |
| 681 | ALDSNSLENDLSDAPGREPGHFNPER[6] Phospho | MSS | 0.04953036
5 |
| 879 | RLEISPDSSPER[5] Phospho | PD | 0.04953036
5 |
| 85 | APGSADYGFAPAAGR[7] Phospho | MSS | 0.04953036
5 |

Showing 1 to 30 of 109 entries, 3 total columns

Reposición de los datos en Github:

Para hacer esto, desde nuestro perfil de github, se crea un nuevo repositorio pulsando en el símbolo + con el desplegable, se escribe el nombre del repositorio, en mi caso: “[Diaz-Hernandez-Eva-Pec1](https://github.com/EvaDiazH/Diaz-Hernandez-Eva-Pec1)” y desde el repositorio se le da a añadir archivos y puedes arrastrar dentro de una cajita los archivos que deseas o seleccionarlos

<https://github.com/EvaDiazH/Diaz-Hernandez-Eva-Pec1>.git

Referencias

Stanimirova, I., & Daszykowski, M. (2018). *Exploratory Analysis of Metabolomic Data. Comprehensive Analytical Chemistry*. doi:10.1016/bs.coac.2018.08.005