

PEC1 Anàlisi de dades òmiques

Àlex Vidal Romera

2024-11-04

Índex de contingut

Abstract	3
Objectius de l'estudi	4
Materials i mètodes	5
Obtenció de les dades	5
Eines informàtiques i bioinformàtiques	5
Procediment general d'anàlisi	5
Resultats	7
Discussió, limitacions i conclusions	15

Abstract

El conjunt de dades escollit mostra les diferents concentracions de metabòlits obtinguts en pacients amb caquèxia i pacients controls.

La caquèxia, tal com defineixi el National Cancer Institute (NCI) és la pèrdua de pes corporal i massa muscular que produeix debilitat. Aquesta afecció a vegades es presenta en pacients amb càncer, SIDA o altres malalties cròniques. L'estudi dels diferents metabòlits en malalts de caquèxia pot ajudar a la identificació de biomarcadors específics de la malaltia així com a la identificació de diferents processos biològics associats. Alhora, el resultat d'una anàlisi exhaustiva de les dades pot obrir possibles portes a la investigació de tractaments relacionats amb els processos associats a la manca o excés de certs metabòlits.

Per a la realització de l'estudi, s'ha procedit a la descàrrega del conjunt de dades del repositori de GitHub proporcionat. Seguidament, s'han importat les dades a un arxiu d'R Markdown amb l'objectiu de processar-les. Amb aquestes dades s'ha creat un contenidor de tipus `SummarizedExperiment` i s'han analitzat les dades.

Els resultats obtinguts ens han permés complir els objectius establerts a l'inici de la prova; S'ha treballat amb diferents eines utilitzades en l'anàlisi de dades òmiques, s'ha creat un contenidor de tipus *SummarizedExperiment* i s'ha pogut realitzar una exploració del conjunt de dades per a proporcionar una visió general. Finalment s'ha elaborat un informe del procés realitzat.

Objectius de l'estudi

L'objectiu principal d'aquesta prova d'avaluació continuada (PAC) és la culminació del procés d'introducció a les òmiques mitjançant el repàs i l'ampliació del contingut impartit durant el curs. Per a realitzar-ho, s'ha dividit la metodologia en diferents objectius:

- Familiaritzar-se amb les tecnologies òmiques.
- Treballar amb diferents eines utilitzades en l'anàlisi de dades òmiques com Bioconductor, expressionSets i GitHub.
- Crear un contenidor de tipus *SummarizedExperiment* que contingui les dades i metadades
- Realitzar una exploració del conjunt de dades seleccionat per a proporcionar una visió general d'aquest.
- Elaborar un informe descrivint el procés realitzat a la PAC.

Materials i mètodes

Obtenció de les dades

Les dades obtingudes per a la realització de l'exercici van ser obtingudes del l'enunciat.

Un cop dins el repositori, s'ha d'accedir a la carpeta Datasets i, un cop allà, seleccionar la carpeta “2024-Cachexia”. Aquesta carpeta conté un fitxer descriptiu de les dades *description.md* i el conjunt de dades utilitzat per a la prova. Al següent enllaç es pot consultar i descarregar directament el conjunt de dades *human_cachexia.csv*.

Tal com es pot consultar al fitxer descriptiu, el conjunt de dades original es pot descarregar del següent enllaç: https://rest.xialab.ca/api/download/metaboanalyst/human_cachexia.csv.

Eines informàtiques i bioinformàtiques

L'exercici s'ha realitzat amb el llenguatge de programació R, utilitzat per a computació estadística i la visualització de dades. Per a treballar més fàcilment amb aquest llenguatge, s'ha utilitzat l'entorn de desenvolupament integrat (IDE) Rstudio, tot i que qualsevol altre IDE és perfectament funcional. Rstudio proporciona una fàcil visualització i execució del codi alhora que permet la creació d'un document exportable que mostri els resultats de l'execució i/o el codi utilitzat.

Per a la creació del contenidor tipus SummarizedExperiment ha sigut necessària la instal·lació de Bioconductor. Bioconductor és un programari utilitzat per a l'anàlisi de dades biològiques i conté paquets dedicats a diferents tasques relacionades amb el camp de la bioinformàtica. El paquet de Bioconductor utilitzat en l'exercici ha sigut SummarizedExperiment. Aquest paquet permet l'organització de les dades de tal forma que sigui més senzill el tractament de les dades així com la realització d'anàlisis posteriors amb aquestes. SummarizedExperiment divideix les dades en tres parts: - Assays: Una matriu que conté les dades (mostres) obtingudes de manera experimental. - Row Data: Conté les característiques de cada fila, en el nostre cas l'identificador del pacient així com al grup al qual pertany, caquèxia o control. - Column Data: Conté informació sobre les mostres. En el nostre cas descriu els noms dels metabòlits analitzats.

Finalment, per a la reposició de les dades i la presentació de la prova, s'ha utilitzat l'eina de control de versions GitHub. Aquesta plataforma web permet emmagatzemar tota mena de projectes al núvol, això permet un control dinàmic dels canvis realitzats en aquest.

Procediment general d'anàlisi

A Rstudio, s'ha definit el directori de treball i s'ha llegit el CSV. S'ha definit la primera columna com la columna d'identificadors de fila. Seguidament, s'ha observat l'estructura de les dades i s'ha comprovat quins possibles valors poden tenir les variables categòriques del conjunt de dades. L'única columna que conté variables categòriques és la columna Muscle.loss i conté només els valors “Cachexia” per les dades dels malalts i “Control” per a les dades de pacients control.

Per a la realització del contenidor SummarizedExperiment s'ha convertit en matriu les dades pertinents a les concentracions de metabòlits i s'han creat dos DataFrames referents a RowData i ColumnData. El DataFrame

de RowData conté “PatientId” i “MuscleLoss” mentre que el DataFrame de ColumnData conté els noms dels metabòlits com “Metabolites”. Amb aquestes dades s’ha pogut crear l’objecte SummarizedExperiment com “se”. S’ha afegit com a metadada el nom del projecte així com una breu descripció d’aquest.

Per acabar, s’ha realitzat una exploració del conjunt de dades a través de l’objecte “se” per a proporcionar una visió general. Aquesta exploració conté les dimensions del conjunt de dades, la quantitat de pacients a cada grup experimental, el nombre de valors nuls, una descripció estadística de les dades de cada metabòlit (mitja, mitjana, quartils, etc.) i una representació gràfica dels resultats.

Resultats

Per tal de complir amb els objectius plantejats a la prova s'ha realitzat un pre-anàlisi del conjunt de dades. Aquest pre-anàlisi ens ajuda a definir l'estructura de les dades i altres característiques d'aquestes.

En llegir l'arxiu CSV s'ha establert la primera columna com a columna d'identificador de fila.

Observem l'estructura de les dades.

```
str(data)
```

```
## 'data.frame':  77 obs. of  64 variables:
## $ Muscle.loss           : chr  "cachexic" "cachexic" "cachexic" "cachexic" ...
## $ X1.6.Anhydro.beta.D.glucose: num  40.9 62.2 270.4 154.5 22.2 ...
## $ X1.Methylnicotinamide   : num  65.4 340.4 64.7 53 73.7 ...
## $ X2.Aminobutyrate        : num  18.7 24.3 12.2 172.4 15.6 ...
## $ X2.Hydroxyisobutyrate   : num  26.1 41.7 65.4 74.4 83.9 ...
## $ X2.Oxoglutarate         : num  71.5 67.4 23.8 1199.9 33.1 ...
## $ X3.Aminoisobutyrate     : num  1480.3 116.8 14.3 555.6 29.7 ...
## $ X3.Hydroxybutyrate      : num  56.83 43.82 5.64 175.91 76.71 ...
## $ X3.Hydroxyisovalerate   : num  10.1 79.8 23.3 25 69.4 ...
## $ X3.Indoxylsulfate       : num  567 369 665 412 166 ...
## $ X4.Hydroxyphenylacetate : num  120.3 432.7 292.9 214.9 97.5 ...
## $ Acetate                 : num  126.5 212.7 314.2 37.3 407.5 ...
## $ Acetone                  : num  9.49 11.82 4.44 206.44 44.26 ...
## $ Adipate                  : num  38.1 327 131.6 144 15 ...
## $ Alanine                  : num  314 871 464 590 1119 ...
## $ Asparagine               : num  159.2 157.6 89.1 273.1 42.5 ...
## $ Betaine                  : num  110 245 117 279 392 ...
## $ Carnitine                : num  265.1 120.3 25 200.3 84.8 ...
## $ Citrate                  : num  3714 2618 863 13630 854 ...
## $ Creatine                 : num  196.4 212.7 221.4 85.6 105.6 ...
## $ Creatinine               : num  16482 15835 24588 20952 6768 ...
## $ Dimethylamine            : num  633 608 735 1064 242 ...
## $ Ethanolamine             : num  645 488 407 821 365 ...
## $ Formate                  : num  441 252 250 469 114 ...
## $ Fucose                   : num  337 198.3 186.8 407.5 26.1 ...
## $ Fumarate                 : num  7.69 18.92 7.1 96.54 19.69 ...
## $ Glucose                   : num  395 8691 1353 863 6836 ...
## $ Glutamine                : num  871 602 302 1686 433 ...
## $ Glycine                  : num  2039 1108 620 5064 395 ...
## $ Glycolate                : num  685.4 652 141.2 70.8 26.6 ...
## $ Guanidoacetate           : num  154 110 183 103 53 ...
## $ Hippurate                : num  4582 1737 4316 757 1153 ...
```

```
## $ Histidine      : num  925 846 284 1043 327 ...
## $ Hypoxanthine   : num  97.5 82.3 114.4 223.6 66.7 ...
## $ Isoleucine     : num  5.58 8.17 9.3 37.71 40.04 ...
## $ Lactate        : num  107 369 750 369 3641 ...
## $ Leucine        : num  42.1 77.5 31.5 103.5 101.5 ...
## $ Lysine         : num  146.9 284.3 97.5 290 122.7 ...
## $ Methylamine    : num  52.5 23.6 18.7 48.9 27.9 ...
## $ Methylguanidine : num  9.97 7.69 4.66 141.17 5.31 ...
## $ N.N.Dimethylglycine : num  23.3 87.4 24.5 40 46.1 ...
## $ O.Acetylcarnitine : num  52.98 50.4 5.58 254.68 45.6 ...
## $ Pantothenate   : num  25.8 186.8 145.5 42.5 74.4 ...
## $ Pyroglutamate  : num  437 437 713 567 185 ...
## $ Pyruvate       : num  21.1 37 29.4 64.1 12.3 ...
## $ Quinolate      : num  165.7 73 192.5 86.5 38.1 ...
## $ Serine         : num  284 392 296 1249 206 ...
## $ Succinate      : num  154.5 244.7 142.6 144 68.7 ...
## $ Sucrose        : num  45.1 459.4 160.8 111 75.2 ...
## $ Tartrate       : num  97.51 32.79 16.28 837.15 4.53 ...
## $ Taurine        : num  1920 1261 4273 1525 469 ...
## $ Threonine      : num  184.9 198.3 110 376.1 64.1 ...
## $ Trigonelline   : num  943.9 208.5 192.5 992.3 86.5 ...
## $ Trimethylamine.N.oxide : num  2122 639 1153 1451 172 ...
## $ Tryptophan     : num  259.8 83.1 82.3 235.1 103.5 ...
## $ Tyrosine       : num  290 167.3 60.3 323.8 142.6 ...
## $ Uracil         : num  111 47 31.5 30.6 44.3 ...
## $ Valine         : num  86.5 110 59.1 102.5 160.8 ...
## $ Xylose         : num  72.2 192.5 2164.6 125.2 186.8 ...
## $ cis.Aconitate   : num  237 334 330 1863 101 ...
## $ myo.Inositol    : num  135.6 376.1 86.5 247.2 750 ...
## $ trans.Aconitate : num  51.9 217 58.6 75.9 98.5 ...
## $ pi.Methylhistidine : num  157.6 308 145.5 249.6 84.8 ...
## $ tau.Methylhistidine : num  160.8 130.3 83.9 254.7 79.8 ...
```

```
# Totes les variables excepte les dues primeres columnes són numèriques.
```

```
# Comprovem quins valors pot prendre la columna "Muscle.loss".
```

```
unique(data$Muscle.loss)
```

```
## [1] "cachexic" "control"
```

```
# Els valors de Muscle.loss poden ser "cachexic" o "control".
```

```
# Convertim la columna en factor per a que R pugui treballar millor les dades categòriques.
```

```
data$Muscle.loss <- as.factor(data$Muscle.loss)
```

Les dades han estat pre-analitzades i preparades per a la creació de l'objecte *SummarizedExperiment*. Per tal d'assolir aquest objectiu s'ha organitzat la informació en una matriu de dades i dos DataFrames que

corresponen a les metadades de fila i columna.

```
# Matriu de dades.
assays_data <- as.matrix(data[, -1])

# Es crea DataFrame de Bioconductor per treballar de manera més eficient.

# Dataframe de metadades de fila
row_data <- DataFrame(
  PatiendId = rownames(data),
  MuscleLoss = data$Muscle.loss
)

# Dataframe de metadades de columna
col_data <- DataFrame(
  Metabolites = colnames(assays_data)
)

# Es crea l'objecte SummarizedExperiment i es mostra el resultat
se <- SummarizedExperiment(assays = SimpleList(counts = assays_data), rowData = row_data, colData = col_data)
print(se)
```

```
## class: SummarizedExperiment
## dim: 77 63
## metadata(0):
## assays(1): counts
## rownames(77): PIF_178 PIF_087 ... NETL_003_V1 NETL_003_V2
## rowData names(2): PatiendId MuscleLoss
## colnames(63): X1.6.Anhydro.beta.D.glucose X1.Methylnicotinamide ...
##   pi.Methylhistidine tau.Methylhistidine
## colData names(1): Metabolites
```

Dels resultats anteriors podem afirmar que s'ha creat l'objecte *SummarizedExperiment* correctament.

A més a més, afegim metadades com el nom del projecte o la descripció del conjunt de dades.

```
metadata(se) <-list(
  project = "PEC1 Analisis de datos omicos",
  dataset_description = "Concentraciones de ciertos metabolitos en pacientes con caquexia y pacientes con ..."
)

metadata(se)

## $project
## [1] "PEC1 Analisis de datos omicos"
##
## $dataset_description
```

```
## [1] "Concentraciones de ciertos metabolitos en pacientes con caquexia y pacientes control"
# Guardem el fitxer tal com es requereix a la prova
save(se, file = "SummarizedExperiment_data.Rda")
```

Per acabar, s'explora l'objecte *se* creat anteriorment. Comprovem les dimensions del conjunt de dades, la quantitat de pacients a cada grup experimental, el nombre de valors nuls, una descripció estadística de les dades de cada metabòlit (mitja, mitjana, quartils, etc.) i una representació gràfica dels resultats.

```
# Dimensions del conjunt de dades
dim(se)
```

```
## [1] 77 63
```

```
# DataFrame dels metabòlits
colData(se)
```

```
## DataFrame with 63 rows and 1 column
##                                     Metabolites
##                                     <character>
## X1.6.Anhydro.beta.D.glucose X1.6.Anhydro.beta.D...
## X1.Methylnicotinamide       X1.Methylnicotinamide
## X2.Aminobutyrate            X2.Aminobutyrate
## X2.Hydroxyisobutyrate       X2.Hydroxyisobutyrate
## X2.Oxoglutarate             X2.Oxoglutarate
## ...                         ...
## cis.Aconitate               cis.Aconitate
## myo.Inositol                myo.Inositol
## trans.Aconitate             trans.Aconitate
## pi.Methylhistidine          pi.Methylhistidine
## tau.Methylhistidine         tau.Methylhistidine
```

```
#DataFrame dels pacients
rowData(se)
```

```
## DataFrame with 77 rows and 2 columns
##           PatiendId MuscleLoss
##           <character> <factor>
## PIF_178      PIF_178  cachexic
## PIF_087      PIF_087  cachexic
## PIF_090      PIF_090  cachexic
## NETL_005_V1  NETL_005_V1 cachexic
## PIF_115      PIF_115  cachexic
## ...         ...      ...
## NETCR_019_V2 NETCR_019_V2 control
## NETL_012_V1  NETL_012_V1 control
## NETL_012_V2  NETL_012_V2 control
## NETL_003_V1  NETL_003_V1 control
```

```
## NETL_003_V2    NETL_003_V2    control

# Nombre de pacients a cada grup experimental
table(rowData(se)$MuscleLoss)

##
## cachexic    control
##          47          30

# Nombre de valors nuls
sum(is.na(assay(se)))

## [1] 0

# Descripció estadística de les dades de cada metabòlit
summary(assay(se))

## X1.6.Anhydro.beta.D.glucose X1.Methylnicotinamide X2.Aminobutyrate
## Min.      : 4.71           Min.      : 6.42           Min.      : 1.28
## 1st Qu.: 28.79           1st Qu.: 15.80           1st Qu.: 5.26
## Median : 45.60           Median : 36.60           Median : 10.49
## Mean    :105.63           Mean    : 71.57           Mean    : 18.16
## 3rd Qu.:141.17           3rd Qu.: 73.70           3rd Qu.: 19.49
## Max.     :685.40           Max.     :1032.77         Max.     :172.43
## X2.Hydroxyisobutyrate X2.Oxoglutarate X3.Aminoisobutyrate X3.Hydroxybutyrate
## Min.      : 4.85           Min.      : 5.53           Min.      : 2.61           Min.      : 1.70
## 1st Qu.:15.80           1st Qu.: 22.42           1st Qu.: 11.70           1st Qu.: 5.99
## Median :32.46           Median : 55.15           Median : 22.65           Median : 11.70
## Mean    :37.25           Mean    :145.09           Mean    : 76.76           Mean    : 21.72
## 3rd Qu.:54.60           3rd Qu.: 92.76           3rd Qu.: 56.26           3rd Qu.: 29.96
## Max.     :93.69           Max.     :2465.13         Max.     :1480.30         Max.     :175.91
## X3.Hydroxyisovalerate X3.Indoxylsulfate X4.Hydroxyphenylacetate
## Min.      : 0.92           Min.      : 27.66           Min.      : 15.49
## 1st Qu.: 5.26           1st Qu.: 82.27           1st Qu.: 41.68
## Median : 12.55           Median : 144.03           Median : 70.11
## Mean    : 21.65           Mean    : 218.88           Mean    :112.02
## 3rd Qu.: 30.27           3rd Qu.: 333.62           3rd Qu.:145.47
## Max.     :164.02           Max.     :1043.15         Max.     :796.32
## Acetate           Acetone           Adipate           Alanine
## Min.      : 3.49           Min.      : 2.29           Min.      : 1.55           Min.      : 16.78
## 1st Qu.: 16.28           1st Qu.: 4.95           1st Qu.: 6.11           1st Qu.: 78.26
## Median : 39.65           Median : 7.10           Median : 10.18           Median : 194.42
## Mean    : 66.14           Mean    : 11.43           Mean    : 24.76           Mean    : 273.56
## 3rd Qu.: 86.49           3rd Qu.: 10.49           3rd Qu.: 19.11           3rd Qu.: 399.41
## Max.     :411.58           Max.     :206.44           Max.     :327.01           Max.     :1312.91
## Asparagine           Betaine           Carnitine           Citrate
```

##	Min.	: 6.69	Min.	: 2.29	Min.	: 2.18	Min.	: 59.74
##	1st Qu.:	20.49	1st Qu.:	28.79	1st Qu.:	14.44	1st Qu.:	788.40
##	Median :	42.10	Median :	64.72	Median :	23.81	Median :	1790.05
##	Mean :	62.28	Mean :	90.32	Mean :	52.09	Mean :	2235.35
##	3rd Qu.:	89.12	3rd Qu.:	127.74	3rd Qu.:	60.95	3rd Qu.:	3071.74
##	Max.	:273.14	Max.	:391.51	Max.	:487.85	Max.	:13629.61
##	Creatine		Creatinine		Dimethylamine		Ethanolamine	
##	Min.	: 2.75	Min.	: 1002	Min.	: 41.26	Min.	: 16.12
##	1st Qu.:	17.64	1st Qu.:	3498	1st Qu.:	142.59	1st Qu.:	86.49
##	Median :	44.26	Median :	7631	Median :	304.90	Median :	204.38
##	Mean :	126.83	Mean :	8734	Mean :	358.17	Mean :	276.26
##	3rd Qu.:	117.92	3rd Qu.:	12333	3rd Qu.:	454.86	3rd Qu.:	407.48
##	Max.	:1863.11	Max.	:33860	Max.	:1556.20	Max.	:1436.55
##	Formate		Fucose		Fumarate		Glucose	
##	Min.	: 6.42	Min.	: 5.70	Min.	: 0.79	Min.	: 26.84
##	1st Qu.:	53.52	1st Qu.:	29.37	1st Qu.:	2.23	1st Qu.:	80.64
##	Median :	95.58	Median :	61.56	Median :	4.10	Median :	210.61
##	Mean :	147.40	Mean :	88.67	Mean :	8.44	Mean :	559.85
##	3rd Qu.:	167.34	3rd Qu.:	123.97	3rd Qu.:	7.85	3rd Qu.:	407.48
##	Max.	:1480.30	Max.	:407.48	Max.	:96.54	Max.	:8690.62
##	Glutamine		Glycine		Glycolate		Guanidoacetate	
##	Min.	: 23.34	Min.	: 38.09	Min.	: 5.42	Min.	: 7.03
##	1st Qu.:	113.30	1st Qu.:	262.43	1st Qu.:	50.91	1st Qu.:	33.78
##	Median :	225.88	Median :	528.48	Median :	130.32	Median :	64.72
##	Mean :	306.87	Mean :	880.72	Mean :	187.99	Mean :	86.37
##	3rd Qu.:	445.86	3rd Qu.:	1096.63	3rd Qu.:	267.74	3rd Qu.:	108.85
##	Max.	:1685.81	Max.	:5064.45	Max.	:720.54	Max.	:561.16
##	Hippurate		Histidine		Hypoxanthine		Isoleucine	
##	Min.	: 92.76	Min.	: 14.15	Min.	: 3.78	Min.	: 1.790
##	1st Qu.:	492.75	1st Qu.:	66.69	1st Qu.:	20.70	1st Qu.:	3.900
##	Median :	1224.15	Median :	174.16	Median :	40.04	Median :	7.170
##	Mean :	2286.84	Mean :	292.64	Mean :	61.10	Mean :	8.709
##	3rd Qu.:	2921.93	3rd Qu.:	419.89	3rd Qu.:	83.93	3rd Qu.:	11.250
##	Max.	:19341.34	Max.	:1863.11	Max.	:265.07	Max.	:40.040
##	Lactate		Leucine		Lysine		Methylamine	
##	Min.	: 7.32	Min.	: 2.51	Min.	: 10.49	Min.	: 1.51
##	1st Qu.:	35.52	1st Qu.:	9.12	1st Qu.:	30.27	1st Qu.:	5.26
##	Median :	81.45	Median :	19.11	Median :	69.41	Median :	14.73
##	Mean :	158.46	Mean :	24.36	Mean :	108.79	Mean :	17.38
##	3rd Qu.:	139.77	3rd Qu.:	31.19	3rd Qu.:	121.51	3rd Qu.:	24.05
##	Max.	:3640.95	Max.	:103.54	Max.	:788.40	Max.	:52.46
##	Methylguanidine		N.N.Dimethylglycine		O.Acetylcarnitine		Pantothenate	
##	Min.	: 1.70	Min.	: 0.79	Min.	: 1.23	Min.	: 2.59

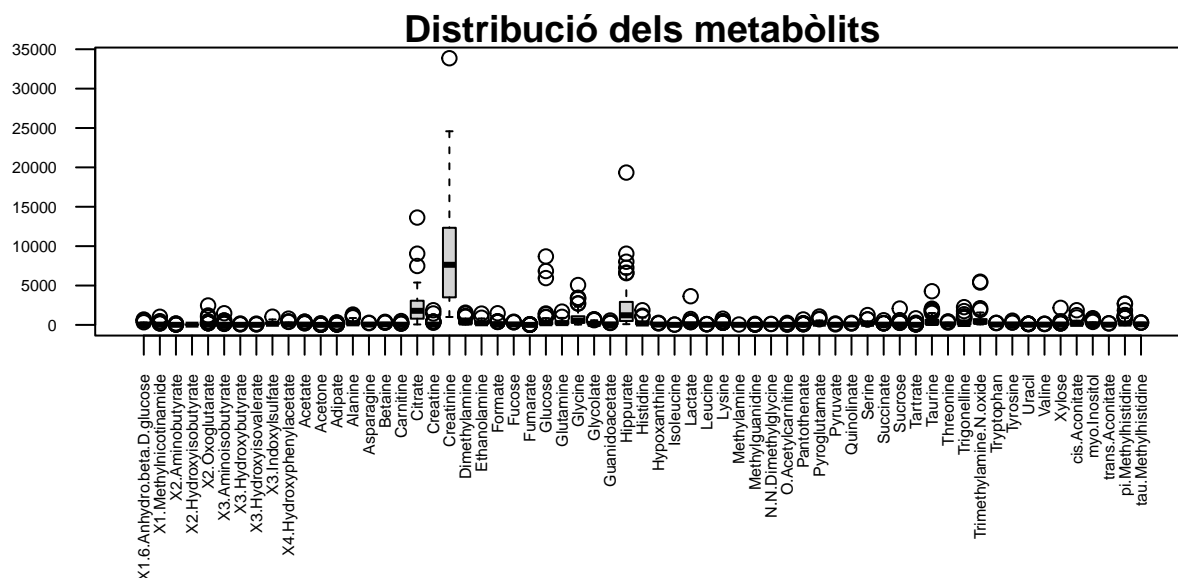
## 1st Qu.: 4.26	1st Qu.: 7.03	1st Qu.: 3.94	1st Qu.: 11.13
## Median : 7.85	Median : 21.98	Median : 11.47	Median : 22.65
## Mean : 15.32	Mean : 26.35	Mean : 19.73	Mean : 44.88
## 3rd Qu.: 19.30	3rd Qu.: 40.04	3rd Qu.: 20.91	3rd Qu.: 41.26
## Max. :141.17	Max. :120.30	Max. :254.68	Max. :692.29
## Pyroglutamate	Pyruvate	Quinolate	Serine
## Min. : 21.33	Min. : 0.90	Min. : 5.21	Min. : 16.12
## 1st Qu.: 68.72	1st Qu.: 4.85	1st Qu.: 26.58	1st Qu.: 83.10
## Median : 157.59	Median : 13.46	Median : 51.42	Median : 142.59
## Mean : 211.45	Mean : 21.29	Mean : 66.44	Mean : 197.69
## 3rd Qu.: 301.87	3rd Qu.: 29.08	3rd Qu.: 87.36	3rd Qu.: 270.43
## Max. :1064.22	Max. :184.93	Max. :259.82	Max. :1248.88
## Succinate	Sucrose	Tartrate	Taurine
## Min. : 1.72	Min. : 6.49	Min. : 2.20	Min. : 17.81
## 1st Qu.: 8.58	1st Qu.: 19.30	1st Qu.: 6.89	1st Qu.: 99.48
## Median : 30.88	Median : 40.85	Median : 12.94	Median : 249.64
## Mean : 60.23	Mean : 113.23	Mean : 40.00	Mean : 525.12
## 3rd Qu.: 74.44	3rd Qu.: 94.63	3rd Qu.: 25.79	3rd Qu.: 665.14
## Max. :589.93	Max. :2079.74	Max. :837.15	Max. :4272.69
## Threonine	Trigonelline	Trimethylamine.N.oxide	Tryptophan
## Min. : 8.25	Min. : 10.07	Min. : 55.7	Min. : 8.67
## 1st Qu.: 31.82	1st Qu.: 53.52	1st Qu.: 175.9	1st Qu.: 21.33
## Median : 64.07	Median : 114.43	Median : 383.8	Median : 46.99
## Mean : 95.36	Mean : 270.44	Mean : 652.2	Mean : 66.24
## 3rd Qu.:137.00	3rd Qu.: 340.36	3rd Qu.: 735.1	3rd Qu.: 96.54
## Max. :450.34	Max. :2252.96	Max. :5486.2	Max. :259.82
## Tyrosine	Uracil	Valine	Xylose
## Min. : 4.22	Min. : 3.10	Min. : 4.10	Min. : 10.07
## 1st Qu.: 23.57	1st Qu.: 11.94	1st Qu.: 12.18	1st Qu.: 29.96
## Median : 60.34	Median : 27.39	Median : 33.12	Median : 50.40
## Mean : 81.76	Mean : 35.56	Mean : 35.67	Mean : 100.93
## 3rd Qu.:113.30	3rd Qu.: 44.26	3rd Qu.: 50.40	3rd Qu.: 89.12
## Max. :539.15	Max. :179.47	Max. :160.77	Max. :2164.62
## cis.Aconitate	myo.Inositol	trans.Aconitate	pi.Methylhistidine
## Min. : 12.94	Min. : 11.59	Min. : 4.90	Min. : 11.36
## 1st Qu.: 36.23	1st Qu.: 30.27	1st Qu.: 12.43	1st Qu.: 67.36
## Median : 129.02	Median : 78.26	Median : 26.84	Median : 162.39
## Mean : 204.22	Mean :135.40	Mean : 40.63	Mean : 370.29
## 3rd Qu.: 254.68	3rd Qu.:167.34	3rd Qu.: 57.40	3rd Qu.: 387.61
## Max. :1863.11	Max. :854.06	Max. :217.02	Max. :2697.28
## tau.Methylhistidine			
## Min. : 8.00			
## 1st Qu.: 27.39			

```
## Median : 68.72
## Mean   : 89.69
## 3rd Qu.:130.32
## Max.    :317.35
```

```
# Representació gràfica de la distribució dels metabòlits
```

```
par(mar=c(14, 3, 1, 1))
```

```
boxplot(assay(se), main="Distribució dels metabòlits", las=2, cex.axis=0.5)
```



Discussió, limitacions i conclusions

En aquesta prova s'ha escollit el conjunt de dades *human_cachexia*, que recull la concentració de fins a seixanta-quatre metabòlits diferents en setanta-set pacients; trenta pacients control i quaranta-set pacients caquèxics.

La principal limitació de l'estudi és que no s'ha pogut determinar l'origen de les dades i, per tant, no s'ha pogut conèixer la manera en que s'han recollit les dades ni el tractament que han patit les mostres per tal d'obtenir-les. Això porta al fet que els resultats obtinguts no siguin rastrejables ni reproduïbles.

Tot i això, el treball amb el conjunt de dades ha estat possible així com l'exploració d'aquest, que ens ha permès assolir els objectius presentats a la prova.

Al següent repositori es pot trobar tota la informació sobre la prova realitzada: <https://github.com/ginkgob/Vidal-Romera-Alejandro-PEC1>.