

# Anàlisi de dades òmiques: PAC1

Sara Sánchez Sanclemente

Novembre, 2024

## Contents

Introducció i objectius . . . . .	2
Importació i lectura de les dades . . . . .	3
Creació de l'objecte <i>SummarizedExperiment</i> . . . . .	6
Exploració de les dades . . . . .	7
Anàlisi exploratòria de les dades . . . . .	7
Exploració del disseny d'estudi . . . . .	16
Resultats i conclusions . . . . .	25

## Introducció i objectius

L'objectiu d'aquest projecte és realitzar una anàlisi de dades metabolòmiques mitjançant un conjunt de dades obtingut d'un repositori en línia, en aquest cas *GitHub*. Un cop descarregades les dades, crearem un contenidor de tipus *SummarizedExperiment* que inclogui tant les dades com les metadades associades al conjunt, com ara informació sobre les files i les columnes. A continuació, portarem a terme una exploració exhaustiva del dataset per obtenir una visió general dels resultats.

Per a fer aquest estudi, s'ha triat la base de dades *human\_achexia.csv*, del repositori de dades *GitHub* que és un recurs reconegut i àmpliament utilitzat en diversos tutorials de MetaboAnalyst. La base de dades ha superat la comprovació de coherència, amb la particularitat que les mostres no estan emparellades i s'han identificat un total de 2 grups entre les mostres. A més, tots els valors continguts en les dades són numèrics, i no s'han detectat valors perduts.

## Importació i lectura de les dades

En aquesta secció, descriuré el procés d'importació i lectura del conjunt de dades seleccionat. Utilitzaré diverses funcions de R per descarregar les dades des del repositori escollit, assegurant-me que es mantinguin els formats adequats. Això permetrà preparar les dades per a l'exploració i l'anàlisi posterior, així com revisar les metadades associades.

En primer lloc cridem el paquet SummarizedExperiment i si no està instal·lat, l'instal·lem.

```
if (!requireNamespace("SummarizedExperiment", quietly = TRUE)) {  
  install.packages("BiocManager")  
  BiocManager::install("SummarizedExperiment")  
}  
library(SummarizedExperiment)
```

Carreguem les dades seleccionades, mostrem una primera visualització i un resum de les variables:

```
dades <- read.csv("C:/Users/saruc/Desktop/master bioest i bionifo/dades_omiques/PAC1/pec1/human_cachexia.csv")  
head(dades)
```

```
## Patient.ID Muscle.loss X1.6.Anhydro.beta.D.glucose X1.Methylnicotinamide  
## 1 PIF_178 cachexic 40.85 65.37  
## 2 PIF_087 cachexic 62.18 340.36  
## 3 PIF_090 cachexic 270.43 64.72  
## 4 NETL_005_V1 cachexic 154.47 52.98  
## 5 PIF_115 cachexic 22.20 73.70  
## 6 PIF_110 cachexic 212.72 31.82  
## X2.Aminobutyrate X2.Hydroxyisobutyrate X2.Oxoglutarate X3.Aminoisobutyrate  
## 1 18.73 26.05 71.52 1480.30  
## 2 24.29 41.68 67.36 116.75  
## 3 12.18 65.37 23.81 14.30  
## 4 172.43 74.44 1199.91 555.57  
## 5 15.64 83.93 33.12 29.67  
## 6 18.36 80.64 47.94 17.46  
## X3.Hydroxybutyrate X3.Hydroxyisovalerate X3.Indoxylsulfate  
## 1 56.83 10.07 566.80  
## 2 43.82 79.84 368.71  
## 3 5.64 23.34 665.14  
## 4 175.91 25.03 411.58  
## 5 76.71 69.41 165.67  
## 6 31.82 35.16 183.09  
## X4.Hydroxyphenylacetate Acetate Acetone Adipate Alanine Asparagine Betaine  
## 1 120.30 126.47 9.49 38.09 314.19 159.17 109.95  
## 2 432.68 212.72 11.82 327.01 871.31 157.59 244.69  
## 3 292.95 314.19 4.44 131.63 464.05 89.12 116.75  
## 4 214.86 37.34 206.44 144.03 589.93 273.14 278.66  
## 5 97.51 407.48 44.26 15.03 1118.79 42.52 391.51  
## 6 132.95 81.45 14.44 25.28 237.46 157.59 66.69  
## Carnitine Citrate Creatine Creatinine Dimethylamine Ethanolamine Formate  
## 1 265.07 3714.50 196.37 16481.60 632.70 645.48 441.42  
## 2 120.30 2617.57 212.72 15835.35 607.89 487.85 252.14  
## 3 25.03 862.64 221.41 24587.66 735.10 407.48 249.64  
## 4 200.34 13629.61 85.63 20952.22 1064.22 820.57 468.72  
## 5 84.77 854.06 105.64 6768.26 242.26 365.04 114.43
```

## 6	40.04	1958.63	200.34	15677.78	614.00	459.44	314.19	
##	Fucose	Fumarate	Glucose	Glutamine	Glycine	Glycolate	Guanidoacetate	Hippurate
## 1	336.97	7.69	395.44	871.31	2038.56	685.40	154.47	4582.50
## 2	198.34	18.92	8690.62	601.85	1107.65	651.97	109.95	1737.15
## 3	186.79	7.10	1352.89	301.87	620.17	141.17	183.09	4315.64
## 4	407.48	96.54	862.64	1685.81	5064.45	70.81	102.51	757.48
## 5	26.05	19.69	6836.29	432.68	395.44	26.58	52.98	1152.86
## 6	123.97	5.05	512.86	298.87	482.99	428.38	57.97	3568.85
##	Histidine	Hypoxanthine	Isoleucine	Lactate	Leucine	Lysine	Methylamine	
## 1	925.19	97.51	5.58	106.70	42.10	146.94	52.46	
## 2	845.56	82.27	8.17	368.71	77.48	284.29	23.57	
## 3	284.29	114.43	9.30	749.95	31.50	97.51	18.73	
## 4	1043.15	223.63	37.71	368.71	103.54	290.03	48.91	
## 5	327.01	66.69	40.04	3640.95	101.49	122.73	27.94	
## 6	459.44	62.80	8.17	113.30	28.79	120.30	36.97	
##	Methylguanidine	N.N.Dimethylglycine	O.Acetylcarnitine	Pantothenate				
## 1	9.97	23.34	52.98	25.79				
## 2	7.69	87.36	50.40	186.79				
## 3	4.66	24.53	5.58	145.47				
## 4	141.17	40.04	254.68	42.52				
## 5	5.31	46.06	45.60	74.44				
## 6	43.38	24.29	13.46	35.52				
##	Pyroglutamate	Pyruvate	Quinolinatate	Serine	Succinate	Sucrose	Tartrate	Taurine
## 1	437.03	21.12	165.67	284.29	154.47	45.15	97.51	1919.85
## 2	437.03	36.97	72.97	391.51	244.69	459.44	32.79	1261.43
## 3	713.37	29.37	192.48	295.89	142.59	160.77	16.28	4272.69
## 4	566.80	64.07	86.49	1248.88	144.03	111.05	837.15	1525.38
## 5	184.93	12.30	38.09	206.44	68.72	75.19	4.53	468.72
## 6	432.68	32.79	112.17	387.61	33.45	336.97	24.05	2059.05
##	Threonine	Trigonelline	Trimethylamine.N.oxide	Tryptophan	Tyrosine	Uracil		
## 1	184.93	943.88	2121.76	259.82	290.03	111.05		
## 2	198.34	208.51	639.06	83.10	167.34	46.99		
## 3	109.95	192.48	1152.86	82.27	60.34	31.50		
## 4	376.15	992.27	1450.99	235.10	323.76	30.57		
## 5	64.07	86.49	172.43	103.54	142.59	44.26		
## 6	105.64	862.64	880.07	239.85	127.74	29.67		
##	Valine	Xylose	cis.Aconitate	myo.Inositol	trans.Aconitate	pi.Methylhistidine		
## 1	86.49	72.24	237.46	135.64	51.94	157.59		
## 2	109.95	192.48	333.62	376.15	217.02	307.97		
## 3	59.15	2164.62	330.30	86.49	58.56	145.47		
## 4	102.51	125.21	1863.11	247.15	75.94	249.64		
## 5	160.77	186.79	101.49	749.95	98.49	84.77		
## 6	36.97	89.12	287.15	129.02	121.51	399.41		
##	tau.Methylhistidine							
## 1	160.77							
## 2	130.32							
## 3	83.93							
## 4	254.68							
## 5	79.84							
## 6	68.72							

*#summary(dades)*

La base de dades conté 77 files (observacions) i 65 columnes (variables). Les columnes inclouen una identifi-

cació de pacient (Patient ID), la condició de pèrdua muscular (Muscle loss), i diverses variables quantitatives sobre metabòlits.

## Creació de l'objecte *SummarizedExperiment*

Ara, per crear el contenidor *SummarizedExperiment*, seguirem aquests passos:

- Dades: Guardarem les variables quantitatives (metabòlits) com a matriu de dades.
- Metadades: Inclourem informació sobre cada pacient i sobre cada columna del dadeset.

```
#Extraiem la matriu de dades metabòliques
mat<- as.matrix(dades[, 3:ncol(dades)])
#comencen a la tercera columna (id i muscle.loss no ho son)

# Definim les dades de les files (metadades de mostres)
files <- data.frame(PatientID = dades$Patient.ID, MuscleLoss = dades$Muscle.loss)

# Creem l'objecte SummarizedExperiment (se)
se <- SummarizedExperiment(
  assays = list(matriu = mat),
  rowData = files,
  colData = DataFrame(Metabolit = colnames(mat))
)

se

## class: SummarizedExperiment
## dim: 77 63
## metadata(0):
## assays(1): matriu
## rownames: NULL
## rowData names(2): PatientID MuscleLoss
## colnames(63): X1.6.Anhydro.beta.D.glucose X1.Methylnicotinamide ...
##   pi.Methylhistidine tau.Methylhistidine
## colData names(1): Metabolit

save(se, file = "objecte_se.Rda")
save(mat,file = "metadades.Rmd")
```

L'objecte *SummarizedExperiment* conté dades de 77 mostres i 63 metabòlits en una matriu que emmagatzema les mesures de concentració de cada metabòlit per mostra. L'objecte no inclou metadades generals, però sí que disposa d'informació específica sobre les files i les columnes. Concretament, les files (mostres) tenen dues metadades (rowData): PatientID, que identifica el pacient, i MuscleLoss, que indica si el pacient presenta caquèxia. Les columnes representen 63 noms de metabòlits agrupats en una única variable (colData) anomenada Metabolit.

## Exploració de les dades

En aquesta secció, realitzaré una anàlisi exploratòria del conjunt de dades fent servir l'objecte *se* creat abans, on presentaré un resum de mesures estadístiques i histogrames per oferir una visualització gràfica clara de les dades. També exploraré el disseny de l'estudi, que es divideix en dos grups segons la condició de pèrdua muscular: el grup control i el grup cachexic. A través de comparacions de boxplots per a cada variable, podrem observar les diferències entre els dos grups, la qual cosa ens permetrà obtenir conclusions sobre l'impacte de la caquexia en les mesures metabòliques.

### Anàlisi exploratòria de les dades

S'ha calculat un resum estadístic per a cada variable, obtenint valors com la mitjana, la mediana, el rang interquartílic i els valors mínim i màxim.

```
#Anàlisi estadística
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
```

Podem observar com alguns metabòlits presenten concentracions especialment elevades en certs casos (com el lactat i l'hipurat), cosa que podria suggerir un augment associat amb l'estat caquètic.

Per altra banda, s'han generat histogrames per a cada metabòlit, proporcionant una representació gràfica de la distribució de les concentracions

```
#Anàlisi gràfic (histogrames)
library(reshape2)
library(gridExtra)
library(ggplot2)

# Convertim la matriu a un data frame i ho formatejem
exprs_df <- as.data.frame(assay(se))
melted_df <- melt(exprs_df)

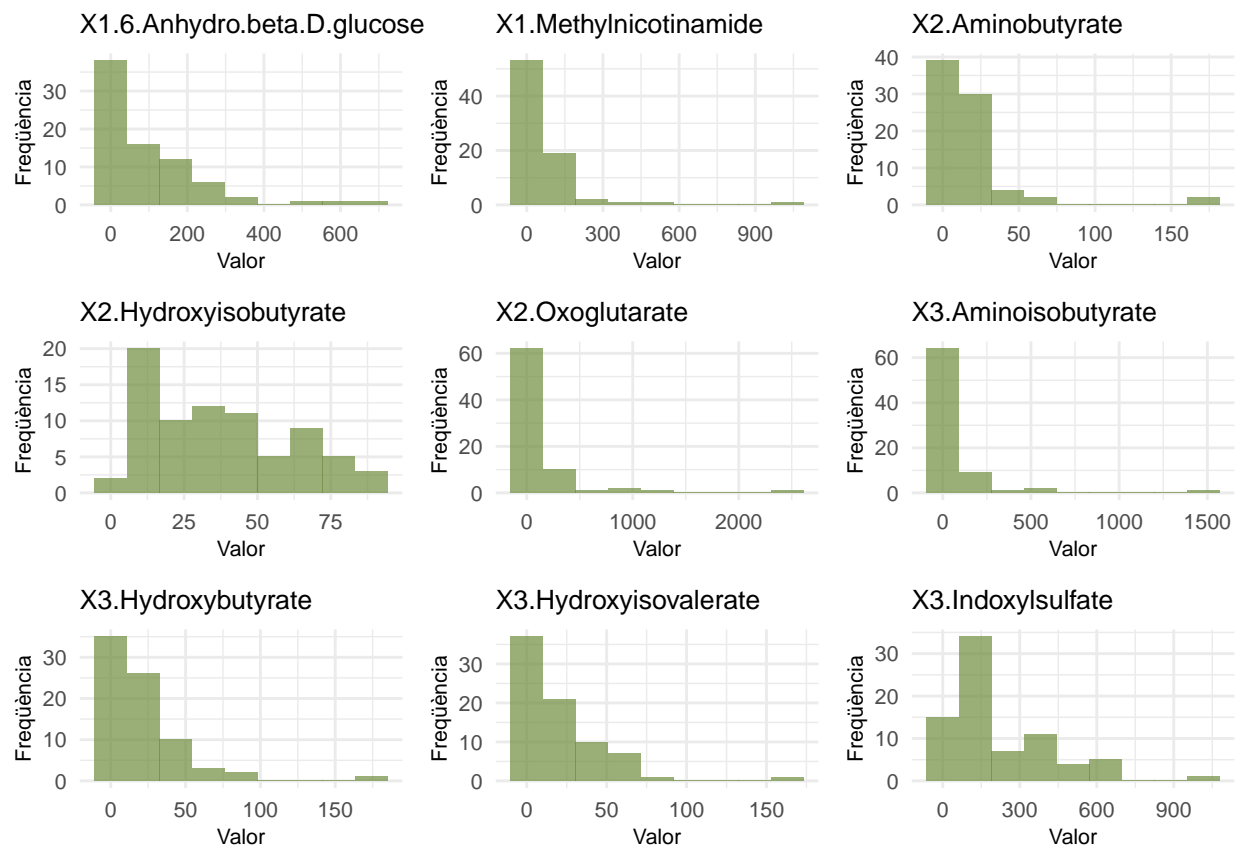
# Creem els histogrames
plots <- lapply(unique(melted_df$variable), function(var) {
  ggplot(melted_df[melted_df$variable == var, ], aes(x = value)) +
    geom_histogram(bins = 9, fill="darkolivegreen4", alpha = 0.7) +
    labs(title = var, x = "Valor", y = "Freqüència") +
    theme_minimal() +
    theme(
      plot.title = element_text(size = 10),
      axis.title = element_text(size = 8),
```

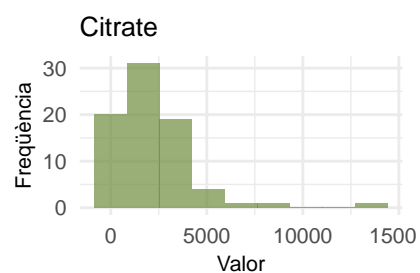
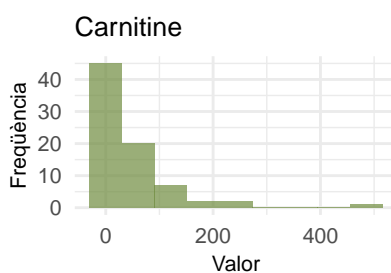
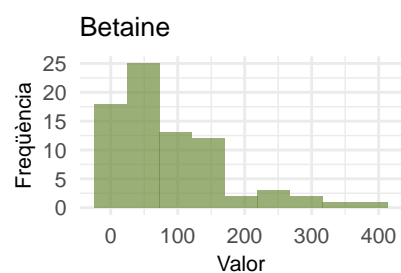
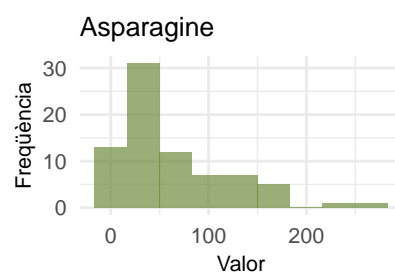
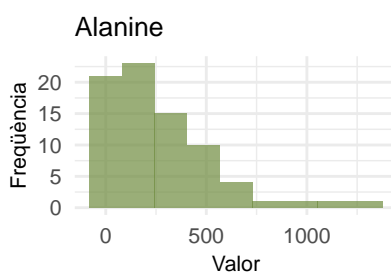
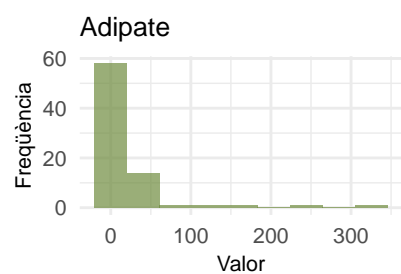
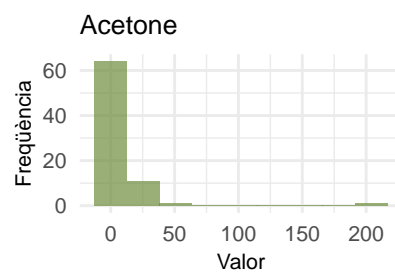
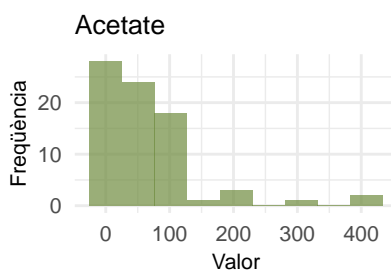
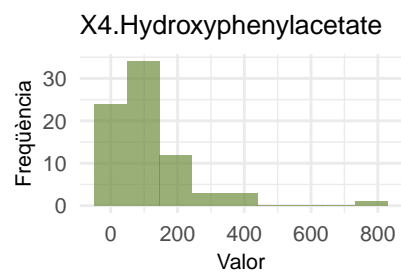
```

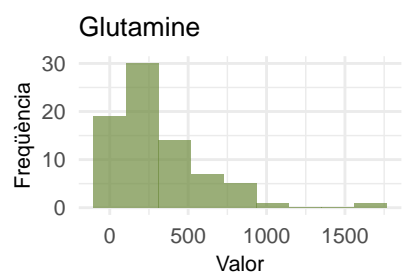
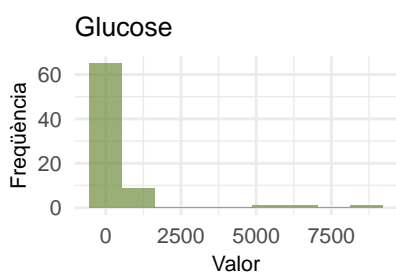
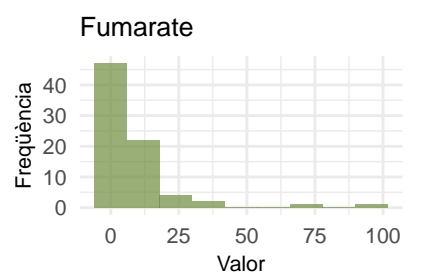
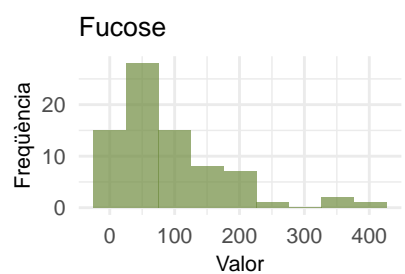
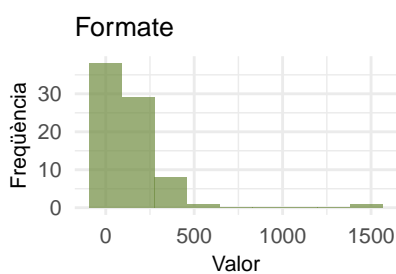
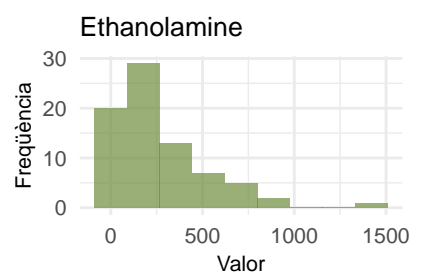
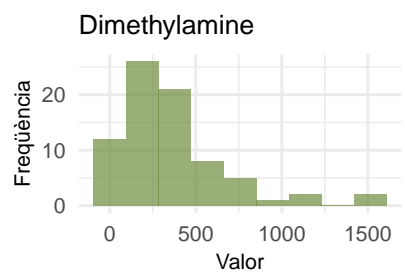
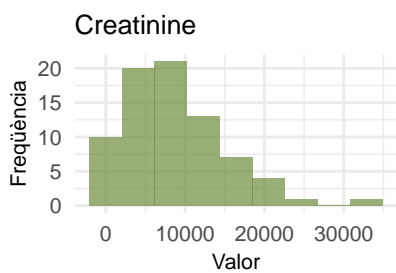
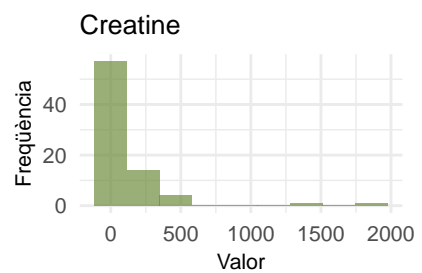
    axis.text = element_text(size = 8)
  )
})

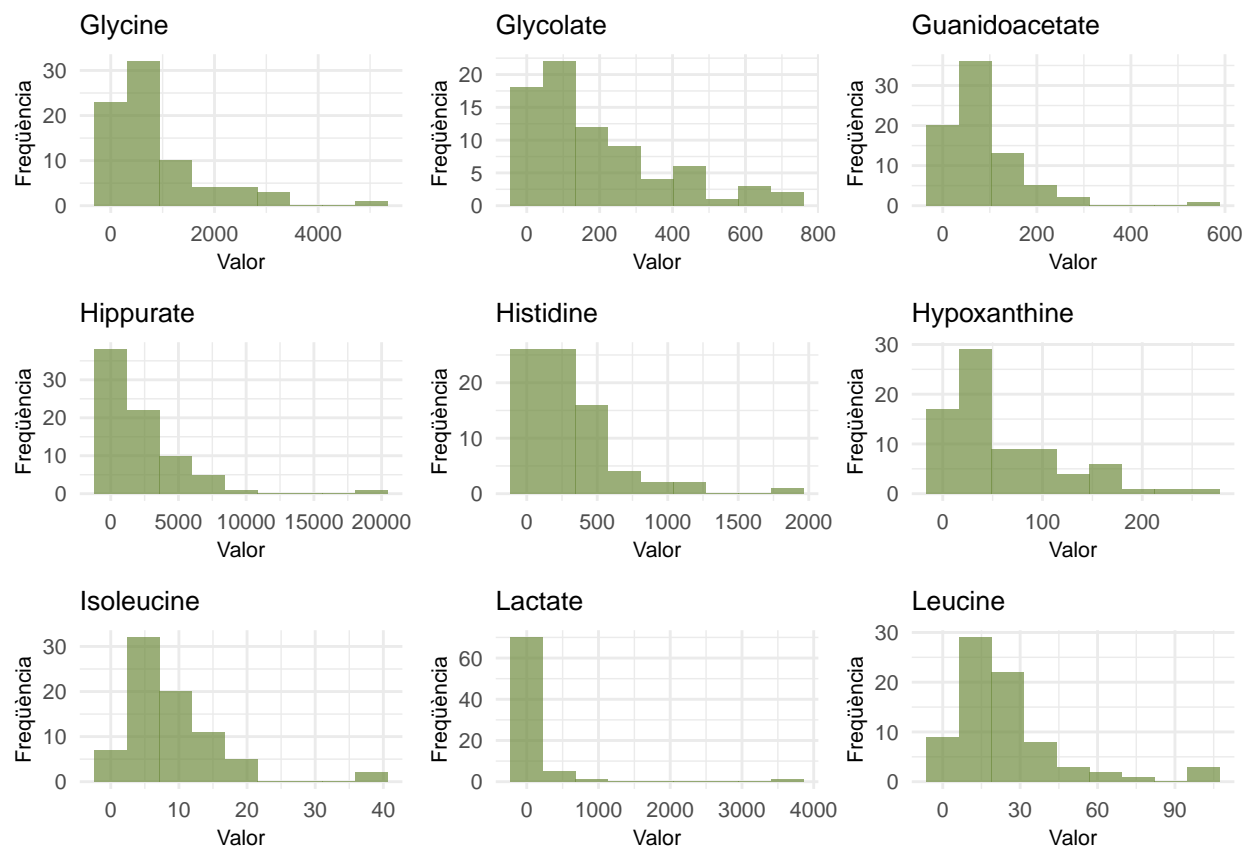
# Mostrem els grafics en grups de 9 (per a que es vegin be i no ocupin molt)
for (i in seq(1, length(plots), by = 9)) {
  grid.arrange(grobs = plots[i:min(i + 8, length(plots))], ncol = 3)
}

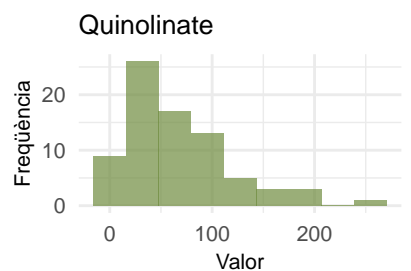
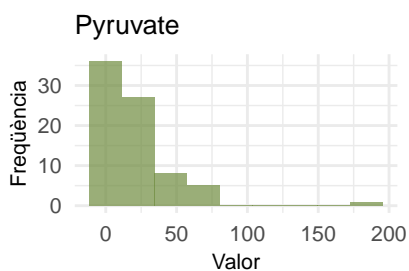
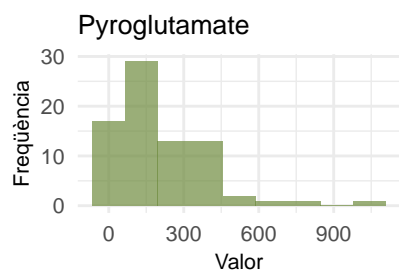
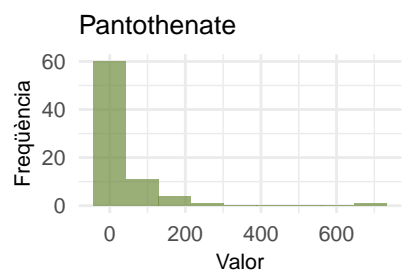
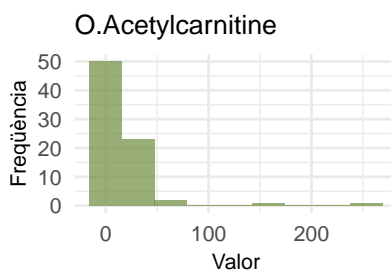
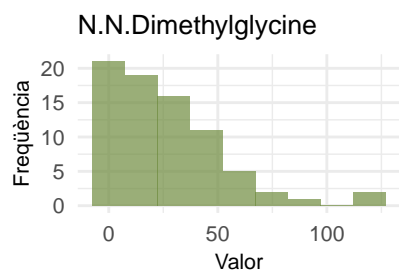
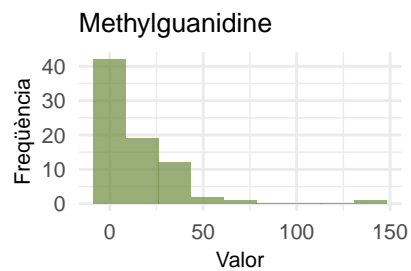
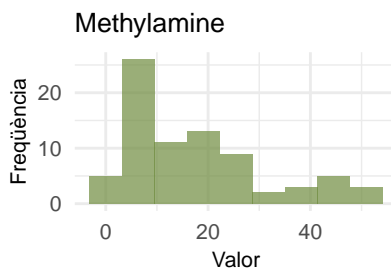
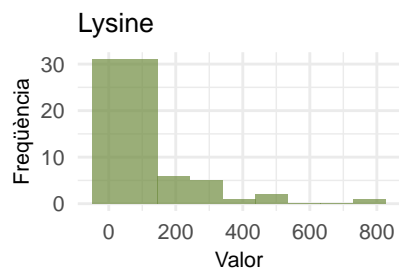
```

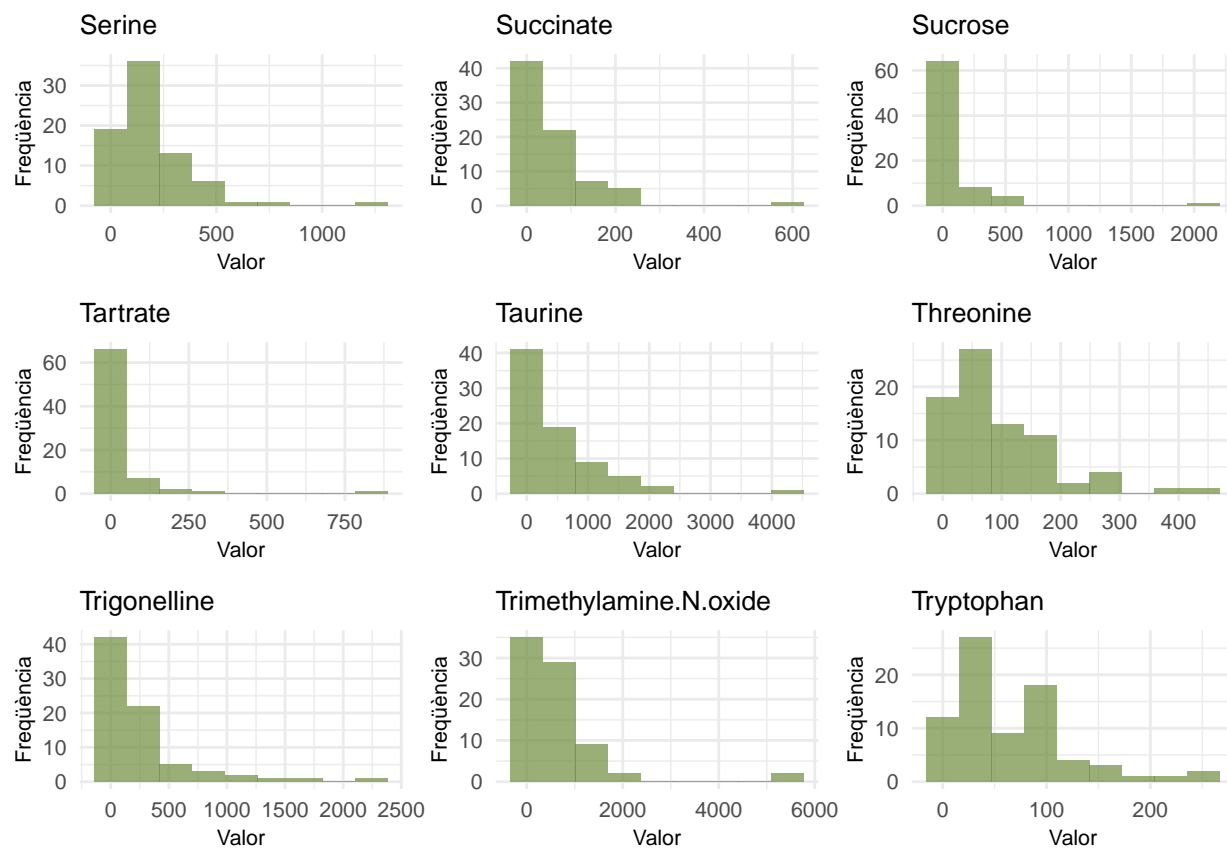


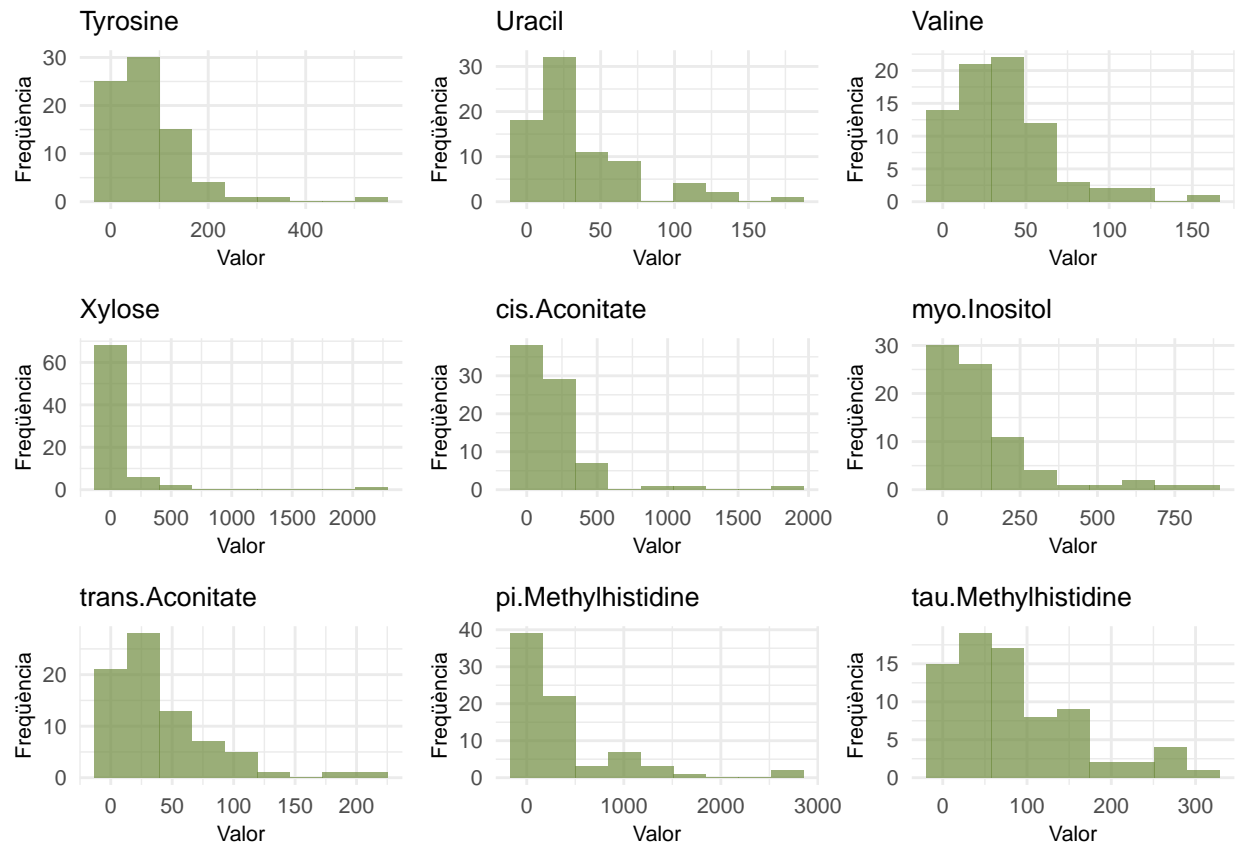












La majoria de metabòlits mostren distribucions esbiaixades, amb una acumulació de valors en l'extrem inferior i alguns valors extremadament alts. Això és habitual en dades metabolòmiques.

### Exploració del disseny d'estudi

Aquest estudi segueix un disseny comparatiu entre dos grups: un grup de pacients caquètics i un grup control. Aquest tipus d'estudi és útil per identificar possibles diferències metabolòmiques associades amb la caquèxia, ja que permet comparar directament els perfils metabòlics entre individus amb aquesta condició i individus sense alteracions metabolòmiques relacionades.

Per explorar les diferències entre grups, s'han creat diagrames de caixes (boxplots) per a cada metabòlit, classificats per grup (control i caquètic).

```
library(dplyr)
library(ggplot2)
library(gridExtra)

# Taula de distribució de la variable MuscleLoss
table(rowData(se)$MuscleLoss)
```

```
##
## cachexic control
##      47      30
```



```

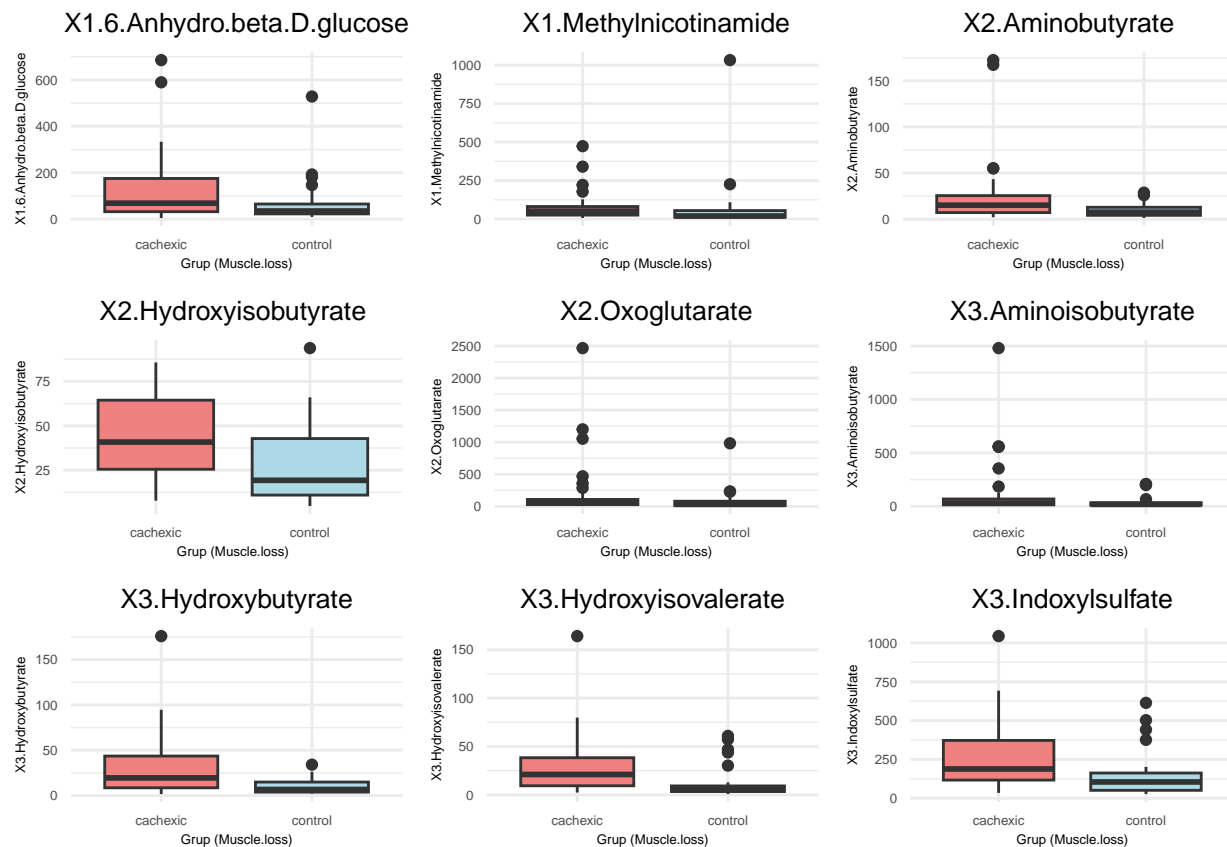
variables_metaboliques <-
  colnames(dades)[!colnames(dades) %in% c("Patient.ID", "Muscle.loss")]

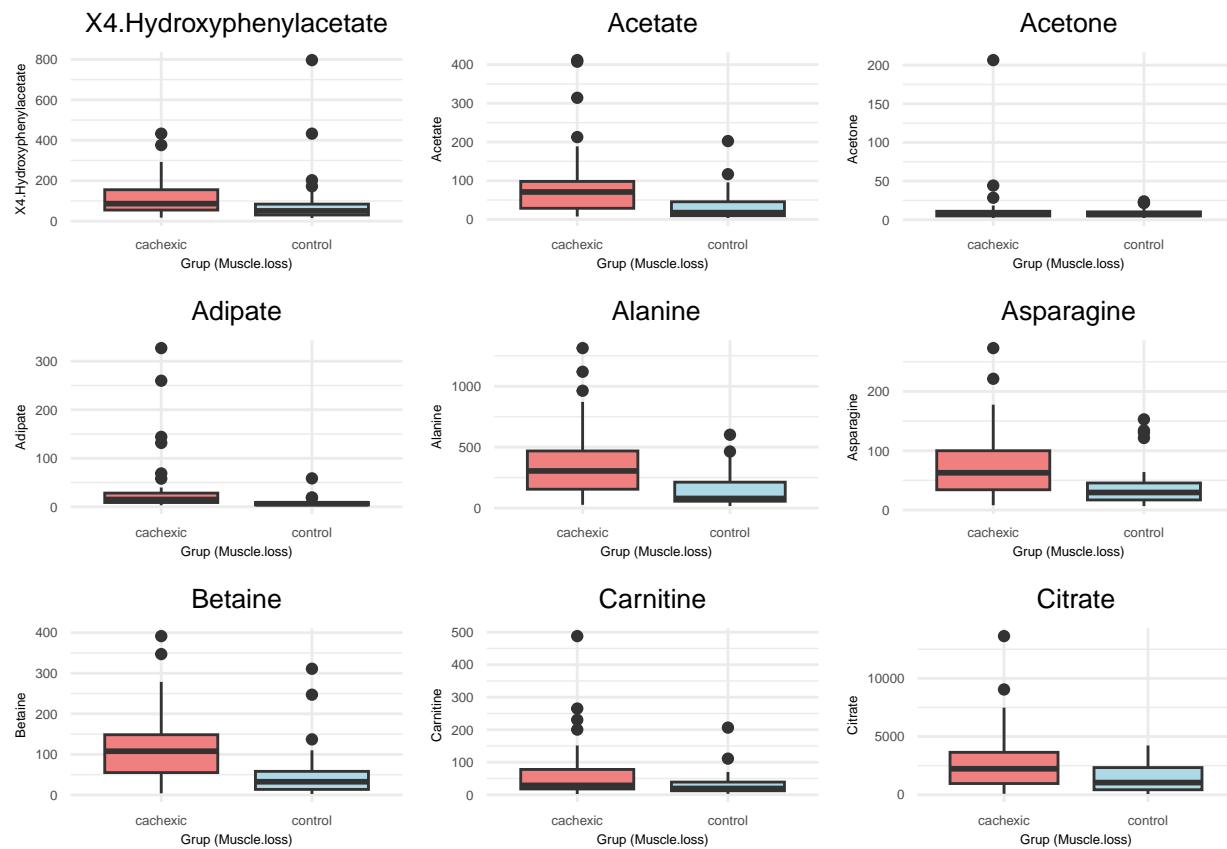
plots <- list() #llista grafic

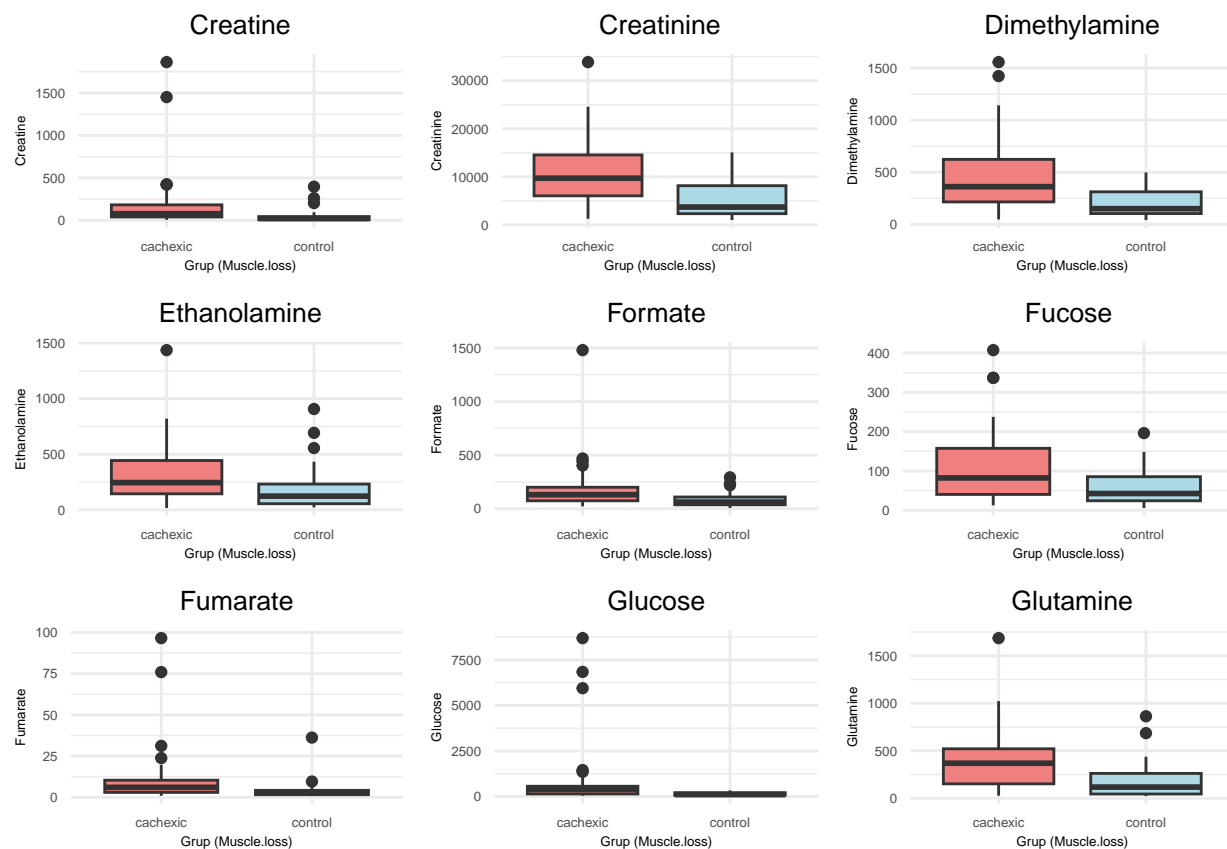
for (variable in variables_metaboliques) {
  graf <- ggplot(dades, aes(x = Muscle.loss, y = .data[[variable]], fill=Muscle.loss)) +
    geom_boxplot() +
    labs(title = paste( variable ),
         x = "Grup (Muscle.loss)",
         y = variable) +
    theme_minimal() +
    theme(legend.position = "none")+
    scale_fill_manual(values = c("control" = "lightblue", "cachexic" = "lightcoral")) +
    theme(plot.title = element_text(hjust = 0.5, size = 10), #ajustar tamany lletra
          axis.title.x = element_text(size = 5),
          axis.title.y = element_text(size = 5),
          axis.text.x = element_text(size = 5),
          axis.text.y = element_text(size = 5))
  plots[[variable]] <- graf #guadem els grafics a la llista
}

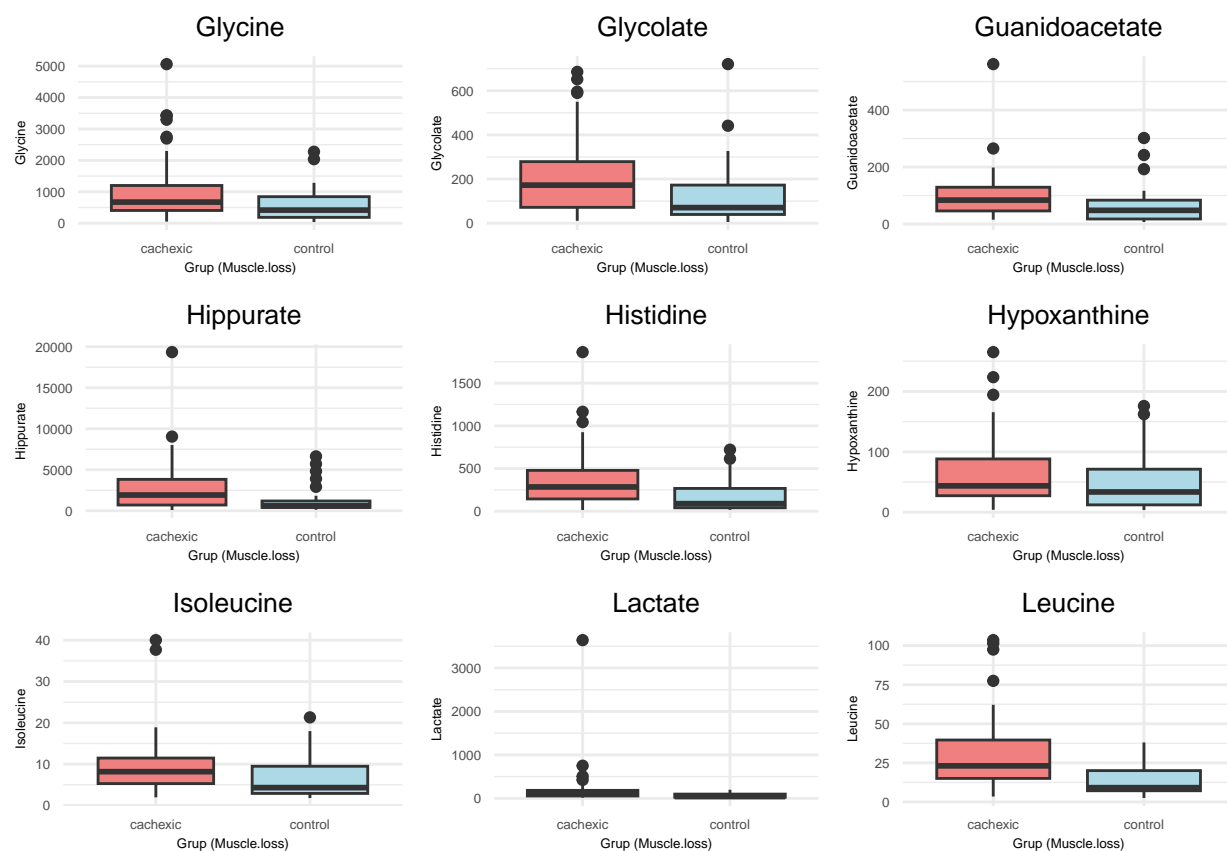
# mostrem els gràfics en grups de 9
for (i in seq(1, length(plots), by = 9)) {
  grid.arrange(grobs = plots[i:min(i + 8, length(plots))], ncol = 3)
}

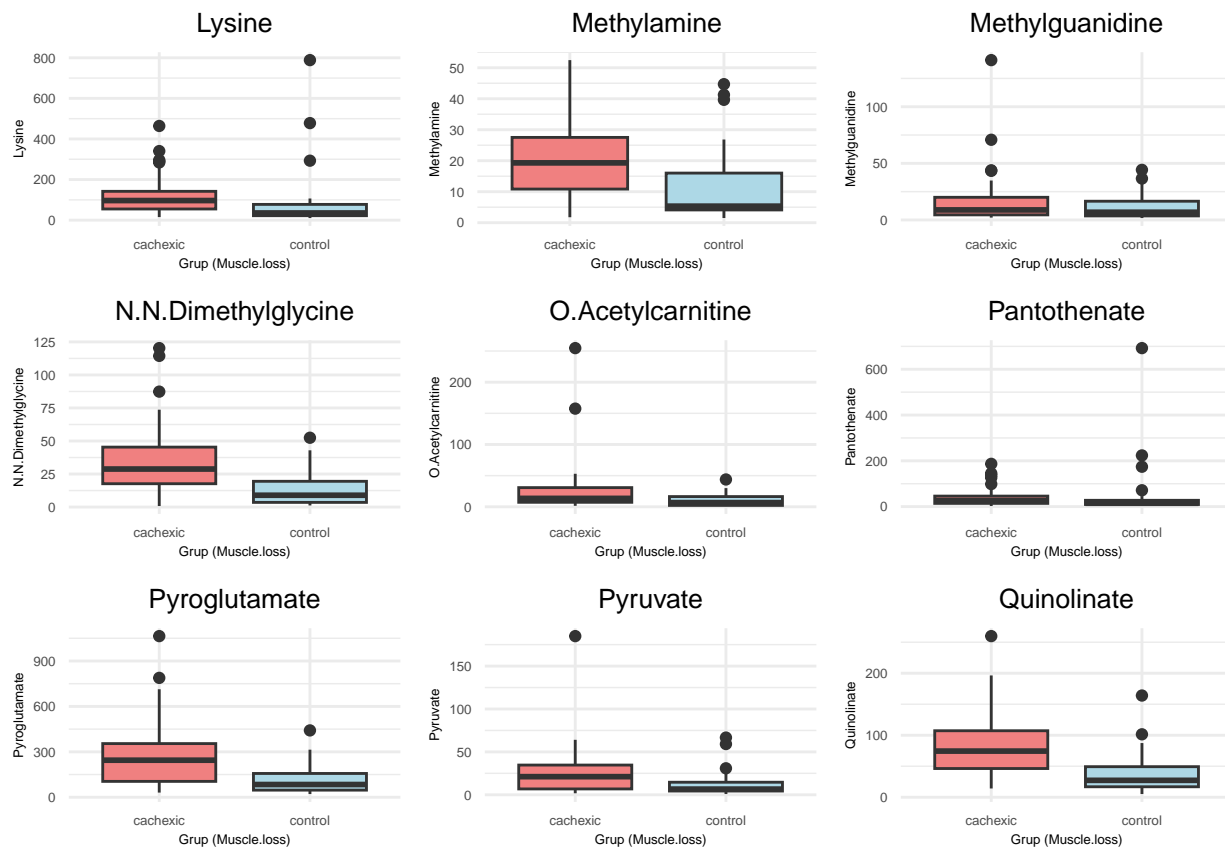
```

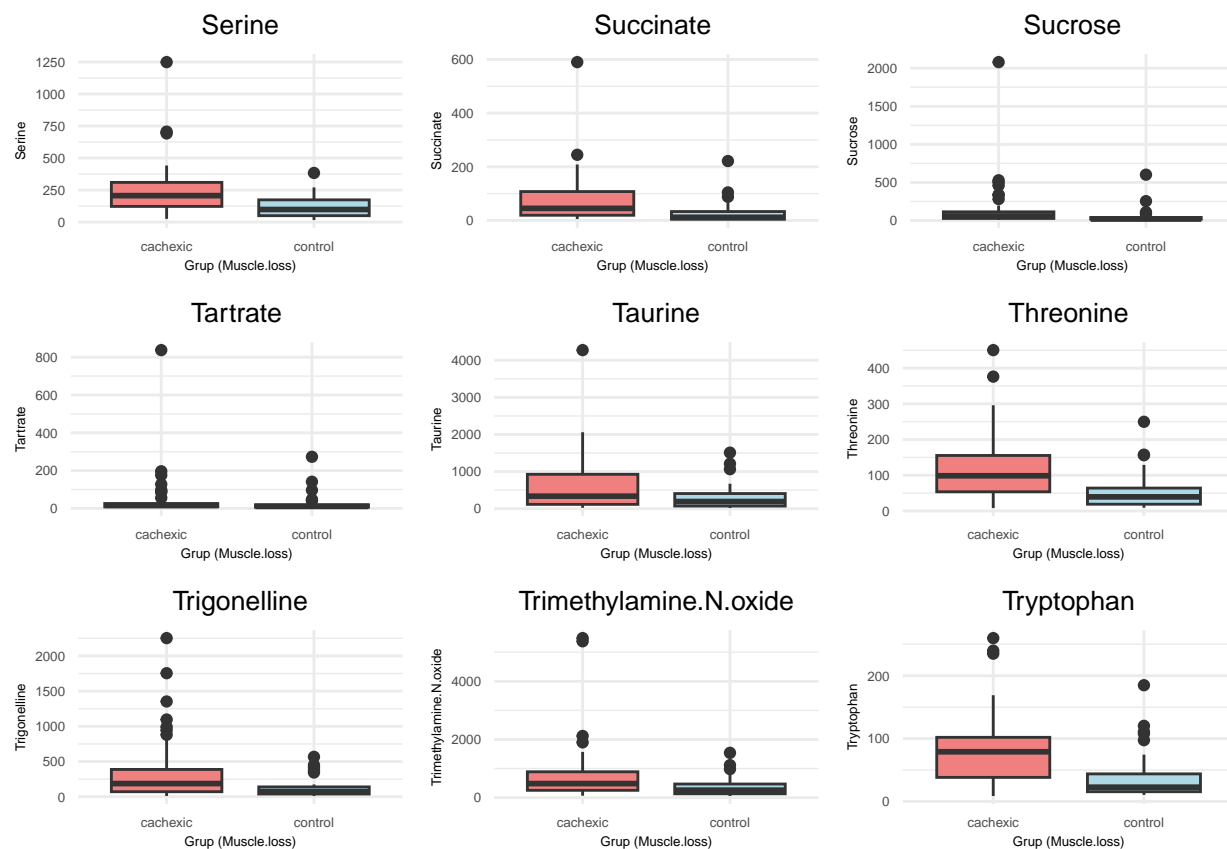


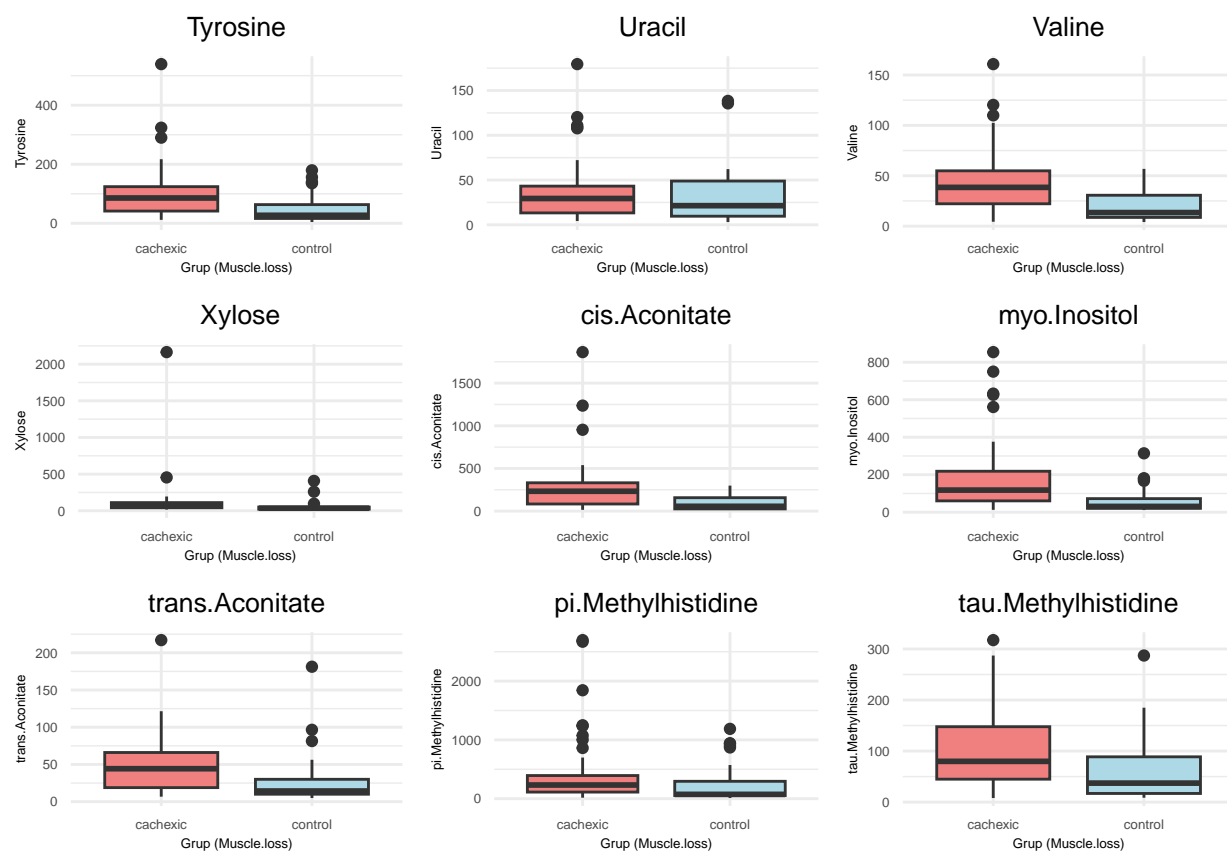












En molts casos, les medians dels metabòlits difereixen entre els dos grups, cosa que suggereix que l'estat caquètic pot estar associat amb canvis en el perfil metabòlic. Alguns metabòlits presenten una diferència estadísticament significativa entre grups, fet que podria indicar una relació amb la condició de caquècia.



## Resultats i conclusions

L'exploració de les dades i l'anàlisi dels boxplots revelen patrons significatius en la concentració de metabòlits entre el grup control i el grup caquètic. Els resultats clau inclouen:

1. S'ha observat un increment en la concentració de diversos metabòlits com el lactat i l'hipurat en el grup caquètic, patró que pot indicar una resposta metabòlica a la caquèxia.
2. Alguns metabòlits, com certs àcids orgànics, mostren una tendència a la baixa en el grup caquètic, la qual cosa podria suggerir una alteració en els processos catabòlics.
3. En diversos metabòlits s'observa una gran dispersió de valors, especialment en el grup caquètic, fet que podria estar relacionat amb l'heterogeneïtat de les respostes individuals a la condició de caquèxia.

Per tant, els resultats indiquen que la caquèxia podria estar associada amb canvis en el perfil metabòlic dels pacients. Aquests canvis podrien ajudar a trobar biomarcadors, és a dir, senyals en el cos que indiquen la presència o la progressió de la caquèxia. A més, estudiar aquestes diferències ens permet entendre millor com afecta la caquèxia al metabolisme dels pacients.

L'**URL** del repositori de Github que conté aquest informe, l'objecte contenidor amb les dades i les metadades en format binari de R (arxiu amb extensió `.Rda`), el codi R per a l'exploració de les dades, les dades en format text i les metadades sobre el dataset en un arxiu markdown és:

**<https://github.com/ssanchezsanc1/Sanchez-Sanclemente-Sara-PEC1.git>**