

PEC 1

Ximena Gabriela Galván Villavicencio

Análisis del conjunto de dattos cachexia.

Tabla de Contenidos

1. Resumen
2. Objetivos del Estudio
 1. Objetivo principal
 2. Objetivos específicos
3. Materiales y Métodos
 1. Origen de los datos
 2. Estudios previos y selección de metabolitos para el análisis
 3. Procedimiento de Análisis
4. Resultados
5. Discusión y Conclusiones
6. Bibliografía
7. Anexos

Resumen

La caquexia es un síndrome de desnutrición asociado a enfermedades crónicas como el cáncer, insuficiencia cardíaca, insuficiencia renal y enfermedades autoinmunes. Se caracteriza por la pérdida progresiva de masa muscular y adiposa, lo que deteriora significativamente la calidad de vida del paciente (Nishikawa et al., 2021). Este estudio analizó datos metabolómicos de pacientes con caquexia y un grupo control, utilizando el contenedor SummarizedExperiment.

Se aplicó una prueba estadística t-test para comparar los niveles de creatina entre los dos grupos, encontrando diferencias significativas (valor $p < 0.05$). Para la visualización de los datos, se usaron gráficos generados con ggplot2, como boxplots y gráficos de dispersión, que destacaron variaciones en metabolitos como 3-metilhistidina, glucosa y aminoácidos de cadena ramificada (leucina, isoleucina, valina), ayudando a comprender mejor la distribución de los datos entre los grupos.

Objetivos del Estudio

Objetivo principal

El objetivo principal de este estudio es crear un objeto SummarizedExperiment a partir de los datos del conjunto de datos ‘human_cachexia.csv’ y realizar un análisis exploratorio para identificar diferencias metabolómicas significativas entre los pacientes con caquexia y los controles.”

Objetivo Especificos

- Crear un objeto SummarizedExperiment a partir de los datos.
- Explorar y analizar el conjunto de datos.

Materiales y Métodos

Origen de los datos

Los datos utilizados en este análisis provienen de GitHub [nutrimetabolomics/metaboData](#), específicamente del dataset titulado `human_cachexia.csv`, el cual contiene datos de diferentes metabolitos y aminoácidos, pertenecientes a muestras de pacientes con diagnóstico de caquexia y de un grupo de control.

Estudios previos y selección de metabolitos para el análisis

Estudios previos en modelos animales han revelado que la caquexia presenta un perfil metabólico dinámicamente cambiante, identificando la participación de cinco vías metabólicas clave. Estas incluyen niveles bajos de glucosa, elevación de cuerpos cetónicos, disminución de aminoácidos de cadena ramificada, aumento de aminoácidos neutros y altos niveles de 3-metilhistidina y creatina. Estos hallazgos proporcionan una comprensión más profunda de los mecanismos metabólicos involucrados en la caquexia (Yang et al., 2017) (Chapela & Martinuzzi, 2018). En base a estos estudios, se seleccionaron ciertos metabolitos para su análisis, específicamente 3-metilhistidina, creatinina, glucosa y los aminoácidos que forman parte de la cadena ramificada que son: leucina, isoleucina y valina.

Procedimiento de Análisis

El procedimiento de análisis se llevó a cabo en los siguientes pasos: Instalación y carga de bibliotecas necesarias como el paquete `readr`, que permite la lectura de archivos CSV.

```
library(readr)
```

Se utilizó la función `read_csv()` para cargar el archivo CSV en un objeto llamado `cachexia_data1`, El conjunto de datos `human_cachexia.csv` consta de 77 filas y 65 columnas.

```
cachexia_data1 <- read_csv("C:/Users/ADMINISTRATOR/Desktop/Bioinformatica/Datos omicos/human_cachexia.csv")
```

Exploramos los datos: Realizamos un resumen con `summary()` para obtener una visión general de las variables, y tiene la siguiente estructura: Identificador de cada paciente (tipo `chr`). Muscle loss: Estado de pérdida muscular, categorizado como “cachexic” o “control” (tipo `chr`). Metabolitos: 63 columnas correspondientes a diferentes metabolitos, representados como `num`

```
summary(cachexia_data1)
```

```
## Muscle.loss      X1.6.Anhydro.beta.D.glucose X1.Methylnicotinamide
## Length:77       Min.      : 4.71                Min.      : 6.42
## Class :character 1st Qu.: 28.79                1st Qu.: 15.80
## Mode  :character Median : 45.60                Median : 36.60
##                Mean   :105.63                Mean   : 71.57
##                3rd Qu.:141.17                3rd Qu.: 73.70
```

##		Max. :685.40		Max. :1032.77
##	X2.Aminobutyrate	X2.Hydroxyisobutyrate	X2.Oxoglutarate	X3.Aminoisobutyrate
##	Min. : 1.28	Min. : 4.85	Min. : 5.53	Min. : 2.61
##	1st Qu.: 5.26	1st Qu.:15.80	1st Qu.: 22.42	1st Qu.: 11.70
##	Median : 10.49	Median :32.46	Median : 55.15	Median : 22.65
##	Mean : 18.16	Mean :37.25	Mean : 145.09	Mean : 76.76
##	3rd Qu.: 19.49	3rd Qu.:54.60	3rd Qu.: 92.76	3rd Qu.: 56.26
##	Max. :172.43	Max. :93.69	Max. :2465.13	Max. :1480.30
##	X3.Hydroxybutyrate	X3.Hydroxyisovalerate	X3.Indoxylsulfate	
##	Min. : 1.70	Min. : 0.92	Min. : 27.66	
##	1st Qu.: 5.99	1st Qu.: 5.26	1st Qu.: 82.27	
##	Median : 11.70	Median : 12.55	Median : 144.03	
##	Mean : 21.72	Mean : 21.65	Mean : 218.88	
##	3rd Qu.: 29.96	3rd Qu.: 30.27	3rd Qu.: 333.62	
##	Max. :175.91	Max. :164.02	Max. :1043.15	
##	X4.Hydroxyphenylacetate	Acetate	Acetone	Adipate
##	Min. : 15.49	Min. : 3.49	Min. : 2.29	Min. : 1.55
##	1st Qu.: 41.68	1st Qu.: 16.28	1st Qu.: 4.95	1st Qu.: 6.11
##	Median : 70.11	Median : 39.65	Median : 7.10	Median : 10.18
##	Mean :112.02	Mean : 66.14	Mean : 11.43	Mean : 24.76
##	3rd Qu.:145.47	3rd Qu.: 86.49	3rd Qu.: 10.49	3rd Qu.: 19.11
##	Max. :796.32	Max. :411.58	Max. :206.44	Max. :327.01
##	Alanine	Asparagine	Betaine	Carnitine
##	Min. : 16.78	Min. : 6.69	Min. : 2.29	Min. : 2.18
##	1st Qu.: 78.26	1st Qu.: 20.49	1st Qu.: 28.79	1st Qu.: 14.44
##	Median : 194.42	Median : 42.10	Median : 64.72	Median : 23.81
##	Mean : 273.56	Mean : 62.28	Mean : 90.32	Mean : 52.09
##	3rd Qu.: 399.41	3rd Qu.: 89.12	3rd Qu.:127.74	3rd Qu.: 60.95
##	Max. :1312.91	Max. :273.14	Max. :391.51	Max. :487.85
##	Citrate	Creatine	Creatinine	Dimethylamine
##	Min. : 59.74	Min. : 2.75	Min. : 1002	Min. : 41.26
##	1st Qu.: 788.40	1st Qu.: 17.64	1st Qu.: 3498	1st Qu.: 142.59
##	Median : 1790.05	Median : 44.26	Median : 7631	Median : 304.90
##	Mean : 2235.35	Mean : 126.83	Mean : 8734	Mean : 358.17
##	3rd Qu.: 3071.74	3rd Qu.: 117.92	3rd Qu.:12333	3rd Qu.: 454.86
##	Max. :13629.61	Max. :1863.11	Max. :33860	Max. :1556.20
##	Ethanolamine	Formate	Fucose	Fumarate
##	Min. : 16.12	Min. : 6.42	Min. : 5.70	Min. : 0.79
##	1st Qu.: 86.49	1st Qu.: 53.52	1st Qu.: 29.37	1st Qu.: 2.23
##	Median : 204.38	Median : 95.58	Median : 61.56	Median : 4.10
##	Mean : 276.26	Mean : 147.40	Mean : 88.67	Mean : 8.44
##	3rd Qu.: 407.48	3rd Qu.: 167.34	3rd Qu.:123.97	3rd Qu.: 7.85
##	Max. :1436.55	Max. :1480.30	Max. :407.48	Max. :96.54
##	Glucose	Glutamine	Glycine	Glycolate
##	Min. : 26.84	Min. : 23.34	Min. : 38.09	Min. : 5.42
##	1st Qu.: 80.64	1st Qu.: 113.30	1st Qu.: 262.43	1st Qu.: 50.91
##	Median : 210.61	Median : 225.88	Median : 528.48	Median :130.32
##	Mean : 559.85	Mean : 306.87	Mean : 880.72	Mean :187.99
##	3rd Qu.: 407.48	3rd Qu.: 445.86	3rd Qu.:1096.63	3rd Qu.:267.74
##	Max. :8690.62	Max. :1685.81	Max. :5064.45	Max. :720.54
##	Guanidoacetate	Hippurate	Histidine	Hypoxanthine
##	Min. : 7.03	Min. : 92.76	Min. : 14.15	Min. : 3.78
##	1st Qu.: 33.78	1st Qu.: 492.75	1st Qu.: 66.69	1st Qu.: 20.70
##	Median : 64.72	Median : 1224.15	Median : 174.16	Median : 40.04

## Mean : 86.37	Mean : 2286.84	Mean : 292.64	Mean : 61.10
## 3rd Qu.:108.85	3rd Qu.: 2921.93	3rd Qu.: 419.89	3rd Qu.: 83.93
## Max. :561.16	Max. :19341.34	Max. :1863.11	Max. :265.07
## Isoleucine	Lactate	Leucine	Lysine
## Min. : 1.790	Min. : 7.32	Min. : 2.51	Min. : 10.49
## 1st Qu.: 3.900	1st Qu.: 35.52	1st Qu.: 9.12	1st Qu.: 30.27
## Median : 7.170	Median : 81.45	Median : 19.11	Median : 69.41
## Mean : 8.709	Mean : 158.46	Mean : 24.36	Mean :108.79
## 3rd Qu.:11.250	3rd Qu.: 139.77	3rd Qu.: 31.19	3rd Qu.:121.51
## Max. :40.040	Max. :3640.95	Max. :103.54	Max. :788.40
## Methylamine	Methylguanidine	N.N.Dimethylglycine	O.Acetylcarnitine
## Min. : 1.51	Min. : 1.70	Min. : 0.79	Min. : 1.23
## 1st Qu.: 5.26	1st Qu.: 4.26	1st Qu.: 7.03	1st Qu.: 3.94
## Median :14.73	Median : 7.85	Median : 21.98	Median : 11.47
## Mean :17.38	Mean : 15.32	Mean : 26.35	Mean : 19.73
## 3rd Qu.:24.05	3rd Qu.: 19.30	3rd Qu.: 40.04	3rd Qu.: 20.91
## Max. :52.46	Max. :141.17	Max. :120.30	Max. :254.68
## Pantothenate	Pyroglutamate	Pyruvate	Quinolinate
## Min. : 2.59	Min. : 21.33	Min. : 0.90	Min. : 5.21
## 1st Qu.: 11.13	1st Qu.: 68.72	1st Qu.: 4.85	1st Qu.: 26.58
## Median : 22.65	Median : 157.59	Median : 13.46	Median : 51.42
## Mean : 44.88	Mean : 211.45	Mean : 21.29	Mean : 66.44
## 3rd Qu.: 41.26	3rd Qu.: 301.87	3rd Qu.: 29.08	3rd Qu.: 87.36
## Max. :692.29	Max. :1064.22	Max. :184.93	Max. :259.82
## Serine	Succinate	Sucrose	Tartrate
## Min. : 16.12	Min. : 1.72	Min. : 6.49	Min. : 2.20
## 1st Qu.: 83.10	1st Qu.: 8.58	1st Qu.: 19.30	1st Qu.: 6.89
## Median : 142.59	Median : 30.88	Median : 40.85	Median : 12.94
## Mean : 197.69	Mean : 60.23	Mean : 113.23	Mean : 40.00
## 3rd Qu.: 270.43	3rd Qu.: 74.44	3rd Qu.: 94.63	3rd Qu.: 25.79
## Max. :1248.88	Max. :589.93	Max. :2079.74	Max. :837.15
## Taurine	Threonine	Trigonelline	Trimethylamine.N.oxide
## Min. : 17.81	Min. : 8.25	Min. : 10.07	Min. : 55.7
## 1st Qu.: 99.48	1st Qu.: 31.82	1st Qu.: 53.52	1st Qu.: 175.9
## Median : 249.64	Median : 64.07	Median : 114.43	Median : 383.8
## Mean : 525.12	Mean : 95.36	Mean : 270.44	Mean : 652.2
## 3rd Qu.: 665.14	3rd Qu.:137.00	3rd Qu.: 340.36	3rd Qu.: 735.1
## Max. :4272.69	Max. :450.34	Max. :2252.96	Max. :5486.2
## Tryptophan	Tyrosine	Uracil	Valine
## Min. : 8.67	Min. : 4.22	Min. : 3.10	Min. : 4.10
## 1st Qu.: 21.33	1st Qu.: 23.57	1st Qu.: 11.94	1st Qu.: 12.18
## Median : 46.99	Median : 60.34	Median : 27.39	Median : 33.12
## Mean : 66.24	Mean : 81.76	Mean : 35.56	Mean : 35.67
## 3rd Qu.: 96.54	3rd Qu.:113.30	3rd Qu.: 44.26	3rd Qu.: 50.40
## Max. :259.82	Max. :539.15	Max. :179.47	Max. :160.77
## Xylose	cis.Aconitate	myo.Inositol	trans.Aconitate
## Min. : 10.07	Min. : 12.94	Min. : 11.59	Min. : 4.90
## 1st Qu.: 29.96	1st Qu.: 36.23	1st Qu.: 30.27	1st Qu.: 12.43
## Median : 50.40	Median : 129.02	Median : 78.26	Median : 26.84
## Mean : 100.93	Mean : 204.22	Mean :135.40	Mean : 40.63
## 3rd Qu.: 89.12	3rd Qu.: 254.68	3rd Qu.:167.34	3rd Qu.: 57.40
## Max. :2164.62	Max. :1863.11	Max. :854.06	Max. :217.02
## pi.Methylhistidine	tau.Methylhistidine		
## Min. : 11.36	Min. : 8.00		

```
## 1st Qu.: 67.36      1st Qu.: 27.39
## Median : 162.39    Median : 68.72
## Mean   : 370.29    Mean   : 89.69
## 3rd Qu.: 387.61    3rd Qu.:130.32
## Max.   :2697.28    Max.    :317.35
```

Para el análisis de los datos de caquexia, se utilizó la clase SummarizedExperiment, que es parte de Bioconductor. Para instalar y cargar esta biblioteca en R, primero fue necesario instalar BiocManager y luego la biblioteca SummarizedExperiment, con el siguiente código:

```
library(SummarizedExperiment)
```

```
## Cargando paquete requerido: MatrixGenerics
```

```
## Cargando paquete requerido: matrixStats
```

```
##
```

```
## Adjuntando el paquete: 'MatrixGenerics'
```

```
## The following objects are masked from 'package:matrixStats':
```

```
##
```

```
## colAlls, colAnyNAs, colAnys, colAvgPerRowSet, colCollapse,
## colCounts, colCummaxs, colCummins, colCumprods, colCumsums,
## colDiffs, colIQRDiffs, colIQRs, colLogSumExps, colMadDiffs,
## colMads, colMaxs, colMeans2, colMedians, colMins, colOrderStats,
## colProds, colQuantiles, colRanges, colRanks, colSdDiffs, colSds,
## colSums2, colTabulates, colVarDiffs, colVars, colWeightedMads,
## colWeightedMeans, colWeightedMedians, colWeightedSds,
## colWeightedVars, rowAlls, rowAnyNAs, rowAnys, rowAvgPerColSet,
## rowCollapse, rowCounts, rowCummaxs, rowCummins, rowCumprods,
## rowCumsums, rowDiffs, rowIQRDiffs, rowIQRs, rowLogSumExps,
## rowMadDiffs, rowMads, rowMaxs, rowMeans2, rowMedians, rowMins,
## rowOrderStats, rowProds, rowQuantiles, rowRanges, rowRanks,
## rowSdDiffs, rowSds, rowSums2, rowTabulates, rowVarDiffs, rowVars,
## rowWeightedMads, rowWeightedMeans, rowWeightedMedians,
## rowWeightedSds, rowWeightedVars
```

```
## Cargando paquete requerido: GenomicRanges
```

```
## Cargando paquete requerido: stats4
```

```
## Cargando paquete requerido: BiocGenerics
```

```
##
```

```
## Adjuntando el paquete: 'BiocGenerics'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
## IQR, mad, sd, var, xtabs
```

```

## The following objects are masked from 'package:base':
##
##   anyDuplicated, aperm, append, as.data.frame, basename, cbind,
##   colnames, dirname, do.call, duplicated, eval, evalq, Filter, Find,
##   get, grep, grepl, intersect, is.unsorted, lapply, Map, mapply,
##   match, mget, order, paste, pmax, pmax.int, pmin, pmin.int,
##   Position, rank, rbind, Reduce, rownames, sapply, saveRDS, setdiff,
##   table, tapply, union, unique, unsplit, which.max, which.min

## Cargando paquete requerido: S4Vectors

##
## Adjuntando el paquete: 'S4Vectors'

## The following object is masked from 'package:utils':
##
##   findMatches

## The following objects are masked from 'package:base':
##
##   expand.grid, I, unname

## Cargando paquete requerido: IRanges

##
## Adjuntando el paquete: 'IRanges'

## The following object is masked from 'package:grDevices':
##
##   windows

## Cargando paquete requerido: GenomeInfoDb

## Cargando paquete requerido: Biobase

## Welcome to Bioconductor
##
##   Vignettes contain introductory material; view with
##   'browseVignettes()'. To cite Bioconductor, see
##   'citation("Biobase)"', and for packages 'citation("pkgname)".

##
## Adjuntando el paquete: 'Biobase'

## The following object is masked from 'package:MatrixGenerics':
##
##   rowMedians

## The following objects are masked from 'package:matrixStats':
##
##   anyMissing, rowMedians

```

Eliminamos el prefijo “X” que a veces se agrega a los nombres de las columnas cuando se cargan los datos desde un archivo CSV. Así se asegura que los nombres de las columnas sean más claros.

```
colnames(cachexia_data1) <- gsub("^X", "", colnames(cachexia_data1))
```

Se transforma la columna Muscle.loss en un factor, definiendo dos niveles: “cachexic” y “control”

```
cachexia_data1$Muscle.loss <- factor(cachexia_data1$Muscle.loss, levels = c("cachexic", "control"))
```

Fue necesario realizar una transposición (t()) para ajustar la estructura de los datos

```
expression_data_transposed <- t(cachexia_data1[, -which(names(cachexia_data1) == "Muscle.loss")])
```

Se creó el contenedor donde assays: los datos de expresión con los metabolitos colData: los metadatos que incluye únicamente la variable Muscle.loss

```
se <- SummarizedExperiment(
  assays = list(counts = as.matrix(expression_data_transposed)),
  colData = DataFrame(Muscle.loss = cachexia_data1$Muscle.loss)
)
```

```
se
```

```
## class: SummarizedExperiment
## dim: 63 77
## metadata(0):
## assays(1): counts
## rownames(63): 1.6.Anhydro.beta.D.glucose 1.Methylnicotinamide ...
## pi.Methylhistidine tau.Methylhistidine
## rowData names(0):
## colnames(77): PIF_178 PIF_087 ... NETL_003_V1 NETL_003_V2
## colData names(1): Muscle.loss
```

Exploramos los datos del contenedor:

```
head(assay(se))
```

```
##               PIF_178 PIF_087 PIF_090 NETL_005_V1 PIF_115 PIF_110
## 1.6.Anhydro.beta.D.glucose 40.85  62.18 270.43    154.47  22.20 212.72
## 1.Methylnicotinamide      65.37 340.36  64.72     52.98  73.70  31.82
## 2.Aminobutyrate          18.73  24.29  12.18    172.43  15.64  18.36
## 2.Hydroxyisobutyrate      26.05  41.68  65.37     74.44  83.93  80.64
## 2.Oxoglutarate           71.52  67.36  23.81   1199.91  33.12  47.94
## 3.Aminoisobutyrate       1480.30 116.75  14.30    555.57  29.67  17.46
##               NETL_019_V1 NETCR_014_V1 NETCR_014_V2 PIF_154
## 1.6.Anhydro.beta.D.glucose 151.41      31.50      51.42 117.92
## 1.Methylnicotinamide       36.60       6.82      30.27  52.46
## 2.Aminobutyrate            8.67       4.18       7.54 19.49
## 2.Hydroxyisobutyrate       42.52      12.94      34.81 72.24
## 2.Oxoglutarate            223.63      25.03      80.64 73.70
```

## 3.Aminoisobutyrate	56.26	8.67	17.99	57.97	
##	NETL_022_V1	NETL_022_V2	NETL_008_V1	PIF_146	PIF_119
## 1.6.Anhydro.beta.D.glucose	20.70	127.74	59.74	89.12	23.57
## 1.Methylnicotinamide	221.41	177.68	50.91	32.79	6.89
## 2.Aminobutyrate	15.18	12.68	6.82	10.38	2.12
## 2.Hydroxyisobutyrate	28.79	15.03	46.06	32.14	7.85
## 2.Oxoglutarate	357.81	68.03	111.05	32.46	8.33
## 3.Aminoisobutyrate	93.69	105.64	8.08	43.38	2.97
##	PIF_099	PIF_162	PIF_160	PIF_113	PIF_143
## 1.6.Anhydro.beta.D.glucose	41.26	589.93	112.17	167.34	183.09
## 1.Methylnicotinamide	8.67	21.98	25.28	19.89	90.92
## 2.Aminobutyrate	2.56	15.18	15.49	13.46	8.94
## 2.Hydroxyisobutyrate	7.85	46.06	47.94	31.19	64.07
## 2.Oxoglutarate	6.89	32.79	28.79	47.94	20.49
## 3.Aminoisobutyrate	6.36	31.82	16.12	79.04	18.73
##	NETCR_007_V2	PIF_137	PIF_100	NETL_004_V1	PIF_094
## 1.6.Anhydro.beta.D.glucose	34.81	333.62	32.46	4.71	68.72
## 1.Methylnicotinamide	95.58	35.87	9.68	11.13	13.87
## 2.Aminobutyrate	23.57	7.92	3.90	43.38	12.18
## 2.Hydroxyisobutyrate	68.03	54.60	11.02	30.88	25.03
## 2.Oxoglutarate	287.15	20.49	170.72	104.58	28.22
## 3.Aminoisobutyrate	104.58	63.43	2.97	54.05	72.97
##	PIF_132	PIF_163	NETCR_003_V1	NETL_028_V1	NETL_028_V2
## 1.6.Anhydro.beta.D.glucose	214.86	304.90	37.71	45.60	34.12
## 1.Methylnicotinamide	127.74	25.79	10.80	473.43	92.76
## 2.Aminobutyrate	31.50	27.11	5.00	16.28	8.25
## 2.Hydroxyisobutyrate	33.78	40.45	8.25	63.43	16.61
## 2.Oxoglutarate	88.23	70.81	11.70	221.41	55.15
## 3.Aminoisobutyrate	64.07	126.47	8.41	15.49	3.39
##	NETCR_013_V1	NETL_020_V1	NETL_020_V2	PIF_192	
## 1.6.Anhydro.beta.D.glucose	107.77	13.33	27.94	141.17	
## 1.Methylnicotinamide	16.61	50.91	80.64	68.03	
## 2.Aminobutyrate	26.84	2.92	15.80	40.85	
## 2.Hydroxyisobutyrate	32.46	40.85	64.72	12.81	
## 2.Oxoglutarate	62.80	46.99	88.23	26.05	
## 3.Aminoisobutyrate	29.67	22.42	11.70	21.76	
##	NETCR_012_V1	NETCR_012_V2	PIF_089	NETCR_002_V1	
## 1.6.Anhydro.beta.D.glucose	14.01	244.69	123.97	141.17	
## 1.Methylnicotinamide	46.06	116.75	81.45	28.50	
## 2.Aminobutyrate	29.08	40.04	55.15	20.29	
## 2.Hydroxyisobutyrate	24.53	61.56	70.81	14.30	
## 2.Oxoglutarate	64.07	174.16	92.76	97.51	
## 3.Aminoisobutyrate	13.07	53.52	561.16	8.41	
##	PIF_179	PIF_114	NETCR_006_V1	PIF_141	NETCR_025_V1
## 1.6.Anhydro.beta.D.glucose	35.16	685.40	278.66	15.80	29.96
## 1.Methylnicotinamide	26.58	36.23	40.45	23.57	96.54
## 2.Aminobutyrate	5.21	32.46	55.15	17.99	6.55
## 2.Hydroxyisobutyrate	30.27	85.63	51.42	37.34	65.37
## 2.Oxoglutarate	7.39	25.03	74.44	21.33	1053.63
## 3.Aminoisobutyrate	8.41	184.93	354.25	26.84	14.15
##	NETCR_025_V2	NETCR_016_V1	PIF_116	PIF_191	PIF_164
## 1.6.Anhydro.beta.D.glucose	16.95	292.95	29.67	18.92	127.74
## 1.Methylnicotinamide	114.43	57.97	70.11	24.53	1032.77
## 2.Aminobutyrate	2.53	167.34	5.58	3.29	8.58

## 2.Hydroxyisobutyrate	77.48	82.27	18.73	10.49	66.02
## 2.Oxoglutarate	2465.13	468.72	5.53	9.68	38.09
## 3.Aminoisobutyrate	19.49	53.52	2.61	26.84	66.69
##	NETL_013_V1	PIF_188	PIF_195	NETCR_015_V1	PIF_102
## 1.6.Anhydro.beta.D.glucose	34.81	65.37	15.18	70.81	25.28
## 1.Methylnicotinamide	12.30	24.05	94.63	75.94	101.49
## 2.Aminobutyrate	5.87	4.71	11.36	22.65	8.33
## 2.Hydroxyisobutyrate	15.18	15.80	8.17	60.95	59.15
## 2.Oxoglutarate	16.78	7.24	5.64	230.44	88.23
## 3.Aminoisobutyrate	11.25	3.13	5.99	53.52	22.65
##	NETL_010_V1	NETL_010_V2	NETL_001_V1	NETCR_015_V2	
## 1.6.Anhydro.beta.D.glucose	34.47	18.54	37.34	33.78	
## 1.Methylnicotinamide	12.81	8.41	55.15	53.52	
## 2.Aminobutyrate	3.78	3.78	7.39	18.17	
## 2.Hydroxyisobutyrate	8.33	4.85	36.23	46.53	
## 2.Oxoglutarate	14.30	8.08	75.94	81.45	
## 3.Aminoisobutyrate	24.29	22.87	9.87	44.70	
##	NETCR_005_V1	PIF_111	PIF_171	NETCR_008_V1	
## 1.6.Anhydro.beta.D.glucose	22.42	146.94	64.07	32.46	
## 1.Methylnicotinamide	55.15	10.07	6.42	14.01	
## 2.Aminobutyrate	20.70	6.30	28.79	2.97	
## 2.Hydroxyisobutyrate	38.47	27.94	18.92	5.16	
## 2.Oxoglutarate	164.02	24.05	85.63	8.08	
## 3.Aminoisobutyrate	206.44	14.88	31.82	5.99	
##	NETCR_008_V2	NETL_017_V1	NETL_017_V2	NETL_002_V1	
## 1.6.Anhydro.beta.D.glucose	113.30	22.20	46.53	192.48	
## 1.Methylnicotinamide	43.38	20.70	9.78	108.85	
## 2.Aminobutyrate	4.66	7.85	3.10	7.77	
## 2.Hydroxyisobutyrate	27.11	19.69	9.30	46.06	
## 2.Oxoglutarate	22.42	38.47	10.59	55.15	
## 3.Aminoisobutyrate	27.11	9.30	13.20	7.03	
##	NETL_002_V2	PIF_190	NETCR_009_V1	NETCR_009_V2	
## 1.6.Anhydro.beta.D.glucose	528.48	28.79	181.27	47.47	
## 1.Methylnicotinamide	225.88	9.21	48.42	7.69	
## 2.Aminobutyrate	13.46	5.53	8.94	4.06	
## 2.Hydroxyisobutyrate	93.69	17.64	51.94	9.30	
## 2.Oxoglutarate	230.44	14.44	982.40	65.37	
## 3.Aminoisobutyrate	10.80	15.49	198.34	50.40	
##	NETL_007_V1	PIF_112	NETCR_019_V2	NETL_012_V1	
## 1.6.Anhydro.beta.D.glucose	15.96	22.87	35.16	16.95	
## 1.Methylnicotinamide	16.12	10.38	52.46	15.80	
## 2.Aminobutyrate	1.93	1.28	13.87	10.49	
## 2.Hydroxyisobutyrate	15.80	5.58	44.26	22.42	
## 2.Oxoglutarate	25.28	8.50	99.48	62.80	
## 3.Aminoisobutyrate	13.46	13.74	208.51	10.91	
##	NETL_012_V2	NETL_003_V1	NETL_003_V2		
## 1.6.Anhydro.beta.D.glucose	9.39	37.71	38.47		
## 1.Methylnicotinamide	14.01	18.17	12.55		
## 2.Aminobutyrate	5.16	26.05	15.03		
## 2.Hydroxyisobutyrate	23.57	15.03	12.55		
## 2.Oxoglutarate	46.99	23.34	22.20		
## 3.Aminoisobutyrate	13.33	33.45	21.33		

```
head(colData(se))
```

```
## DataFrame with 6 rows and 1 column
##           Muscle.loss
##           <factor>
## PIF_178      cachexic
## PIF_087      cachexic
## PIF_090      cachexic
## NETL_005_V1  cachexic
## PIF_115      cachexic
## PIF_110      cachexic
```

Se observaron las primeras filas del conjunto de datos para entender su formato y los tipos de datos presentes y identificaron las principales columnas que diferencian a los grupos de estudio: la variable Muscle loss y los metabolitos.

Se puede observar que de 70 pacientes, 30 de ellos son control y 37 pacientes con cachexia.

```
sum(cachexia_data1$Muscle.loss == "control")
```

```
## [1] 30
```

```
sum (cachexia_data1$Muscle.loss == "cachexic")
```

```
## [1] 47
```

Para la exploración y análisis de los datos, se utilizaron las siguientes bibliotecas: ggplot2: utilizada para la visualización y creación de gráficos. tidyr y dplyr: empleadas para manipular y transformar los datos. Se cargaron los datos de metabolitos y se separaron los grupos de pacientes en función de la variable Muscle.loss. Además se usó t-test para comparar los niveles de creatina entre los grupos de pérdida (cachexia y control), este test evalúa si existe una diferencia significativa entre las medias de los dos grupos. La hipótesis nula (H0) asumió que no había diferencia en las medias, mientras que la alternativa (H1) propuso que sí existía una diferencia. El análisis se realizó utilizando la función t.test()

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

```
##
## Adjuntando el paquete: 'dplyr'

## The following object is masked from 'package:Biobase':
##
##      combine

## The following objects are masked from 'package:GenomicRanges':
##
##      intersect, setdiff, union

## The following object is masked from 'package:GenomeInfoDb':
##
##      intersect
```

```
## The following objects are masked from 'package:IRanges':
##
## collapse, desc, intersect, setdiff, slice, union

## The following objects are masked from 'package:S4Vectors':
##
## first, intersect, rename, setdiff, setequal, union

## The following objects are masked from 'package:BiocGenerics':
##
## combine, intersect, setdiff, union

## The following object is masked from 'package:matrixStats':
##
## count

## The following objects are masked from 'package:stats':
##
## filter, lag

## The following objects are masked from 'package:base':
##
## intersect, setdiff, setequal, union
```

```
library(tidyr)
```

```
##
## Adjuntando el paquete: 'tidyr'

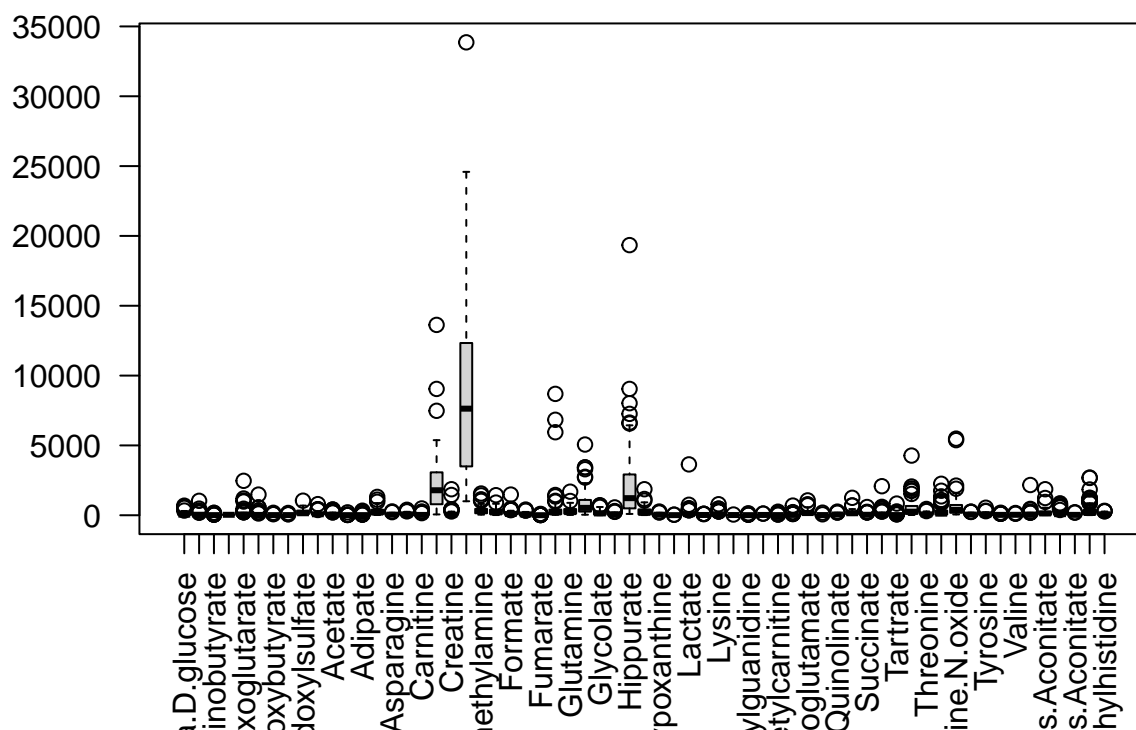
## The following object is masked from 'package:S4Vectors':
##
## expand
```

Resultados

Se observa una gran variabilidad en los niveles de los metabolitos entre las diferentes muestras. Algunos metabolitos presentan niveles muy altos en ciertas muestras, mientras que en otras son prácticamente indetectables, además se observa valores atípicos.

```
boxplot(t(assay(se)), main = "Distribucion de metabolitos por muestra", las = 2)
```

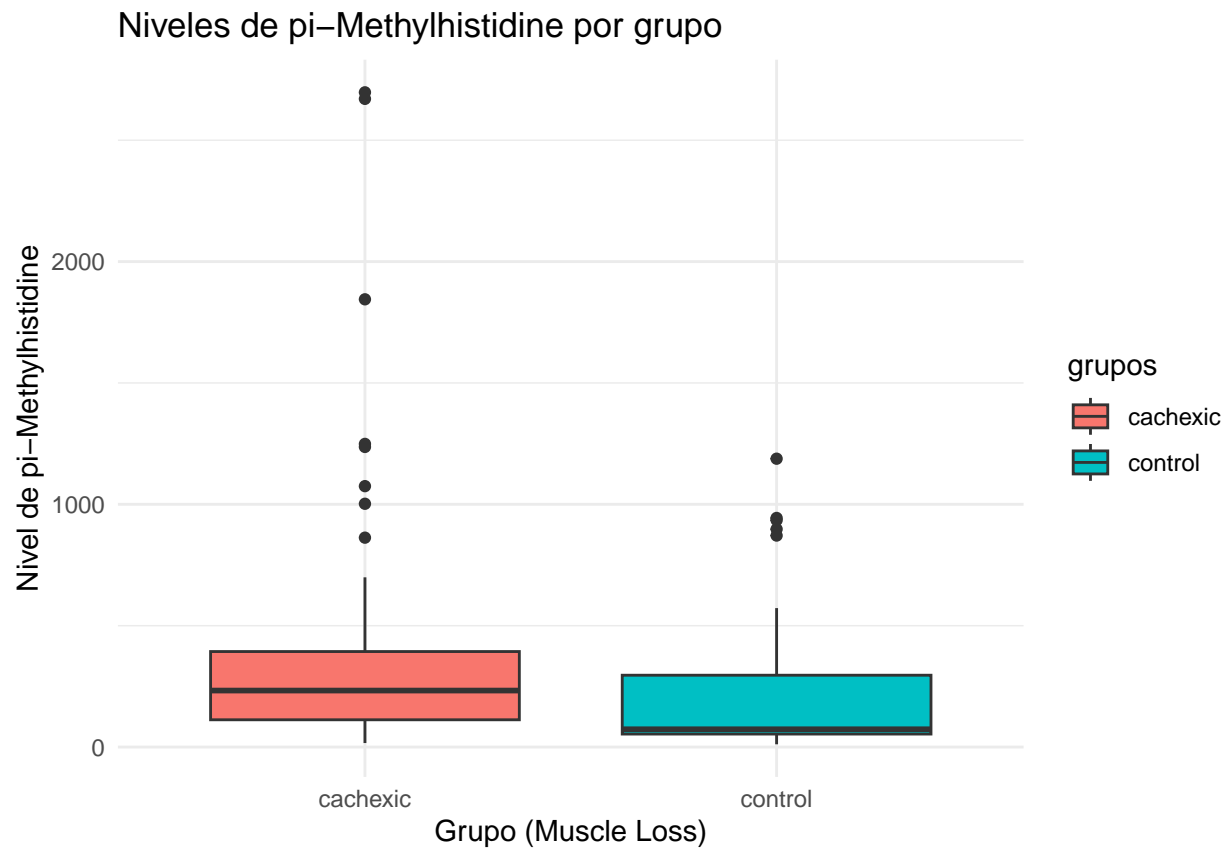
Distribucion de metabolitos por muestra



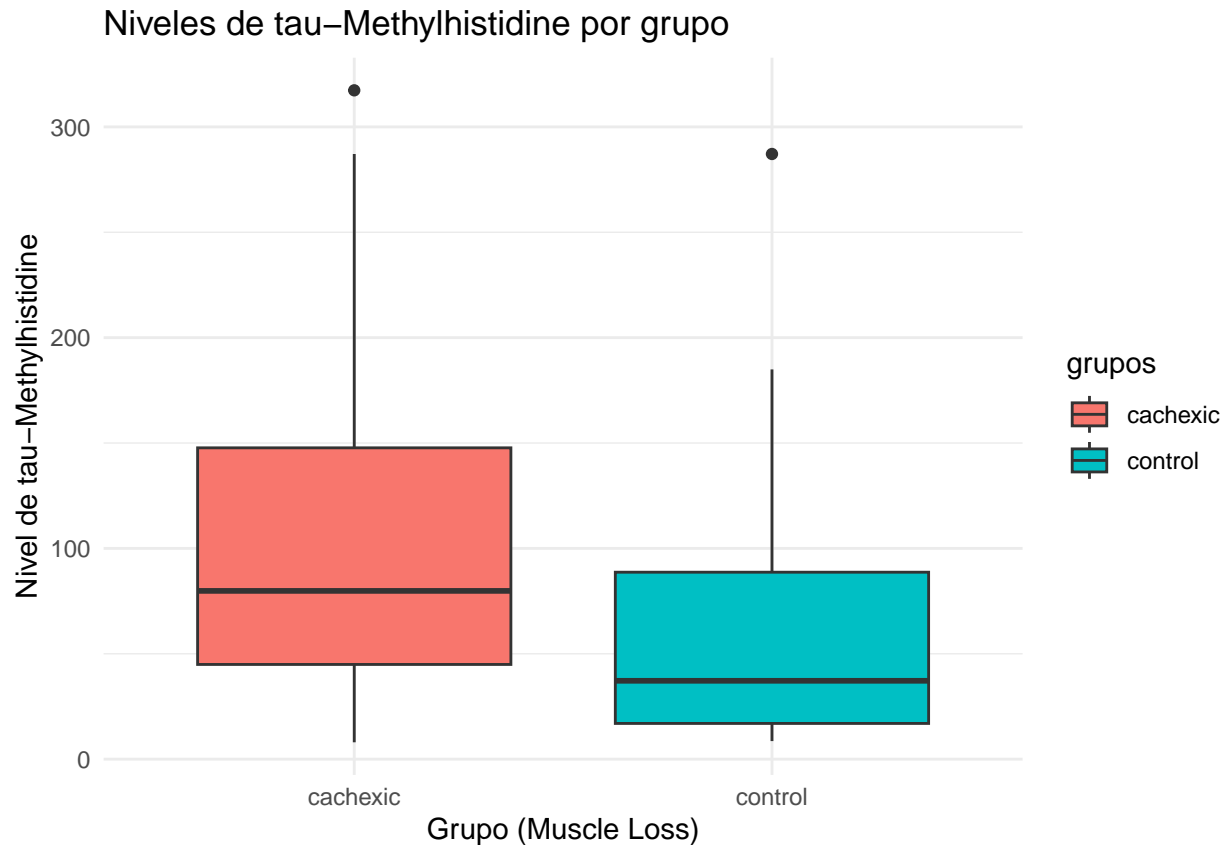
Al analizar pi.Methylhistidine y tau.Methylhistidine, se observa que los niveles de pi-metilhistidina son significativamente más altos en el grupo caquético en comparación con el grupo control, lo que coincide con la expectativa de que la degradación muscular sea más pronunciada en individuos con caquexia.

```
metabolitos_data <- assay(se)
grupos <- colData(se)$Muscle.loss

# Grafico los niveles de pi.Methylhistidine por grupo (boxplot)
ggplot() +
  geom_boxplot(aes(x = grupos, y = metabolitos_data["pi.Methylhistidine", ], fill = grupos)) +
  labs(title = "Niveles de pi-Methylhistidine por grupo",
       x = "Grupo (Muscle Loss)", y = "Nivel de pi-Methylhistidine") +
  theme_minimal()
```



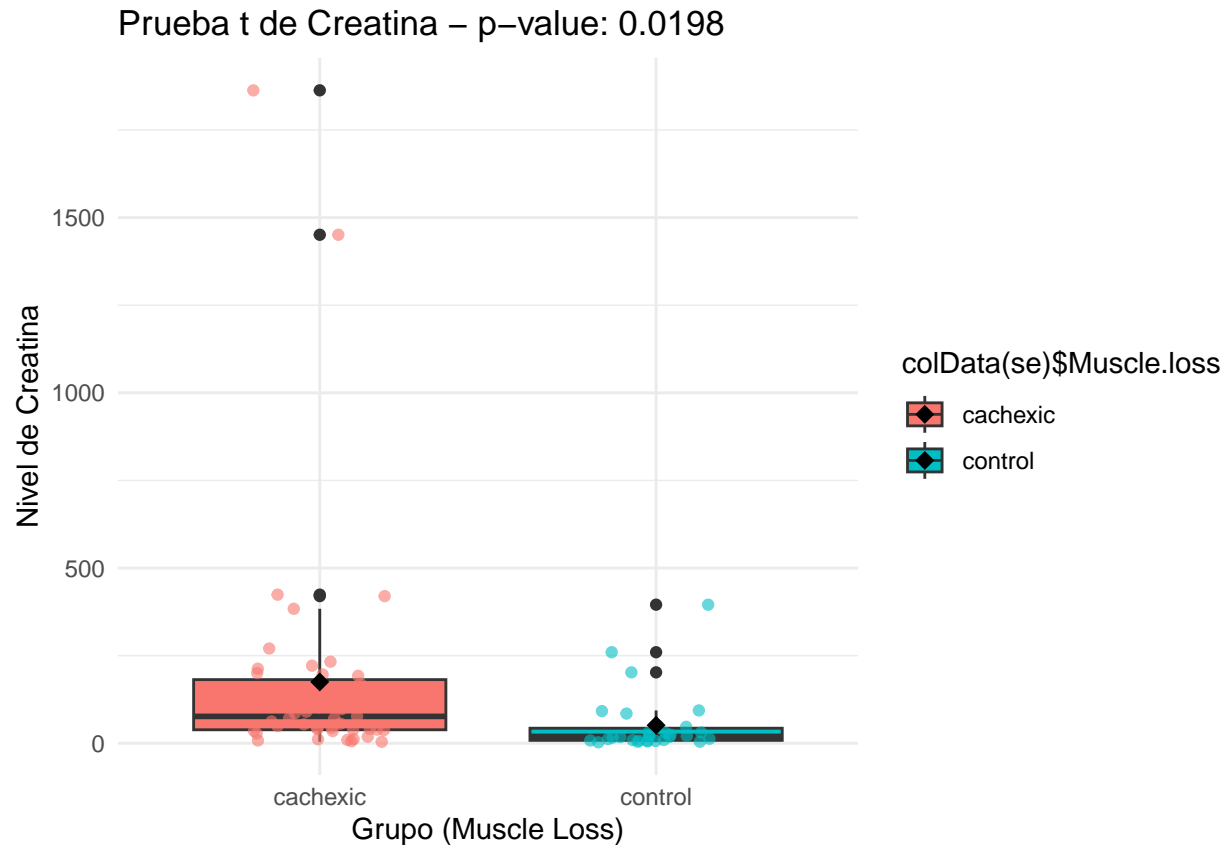
```
# Grafico los niveles de tau.Methylhistidine por grupo (boxplot)
ggplot() +
  geom_boxplot(aes(x = grupos, y = metabolitos_data["tau.Methylhistidine", ], fill = grupos)) +
  labs(title = "Niveles de tau-Methylhistidine por grupo",
        x = "Grupo (Muscle Loss)", y = "Nivel de tau-Methylhistidine") +
  theme_minimal()
```



Se compararon los niveles de creatina entre los grupos de pérdida muscular (cachexia) y control utilizando un test t . Dado que el valor p fue menor a 0.05, se rechazó la hipótesis nula, que planteaba que no había diferencia en las medias de creatina entre los grupos. Esto indica que los niveles de creatina difieren significativamente entre el grupo cachexia y el grupo control. Además, se realizó un boxplot donde se muestra la diferencia en los niveles de creatina entre ambos grupos.

```
t_test_result <- t.test(assay(se)["Creatine", ] ~ colData(se)$Muscle.loss)

# Gráfico boxplot para comparar los niveles de Creatina entre los grupos
ggplot(data = NULL, aes(x = colData(se)$Muscle.loss, y = assay(se)["Creatine", ], fill = colData(se)$Muscle.loss)) +
  geom_boxplot() +
  geom_jitter(width = 0.2, aes(color = colData(se)$Muscle.loss), alpha = 0.6) +
  labs(title = paste("Prueba t de Creatina - p-value:", round(t_test_result$p.value, 4)),
       x = "Grupo (Muscle Loss)",
       y = "Nivel de Creatina") +
  theme_minimal() +
  stat_summary(fun = mean, geom = "point", shape = 18, size = 3, color = "black")
```



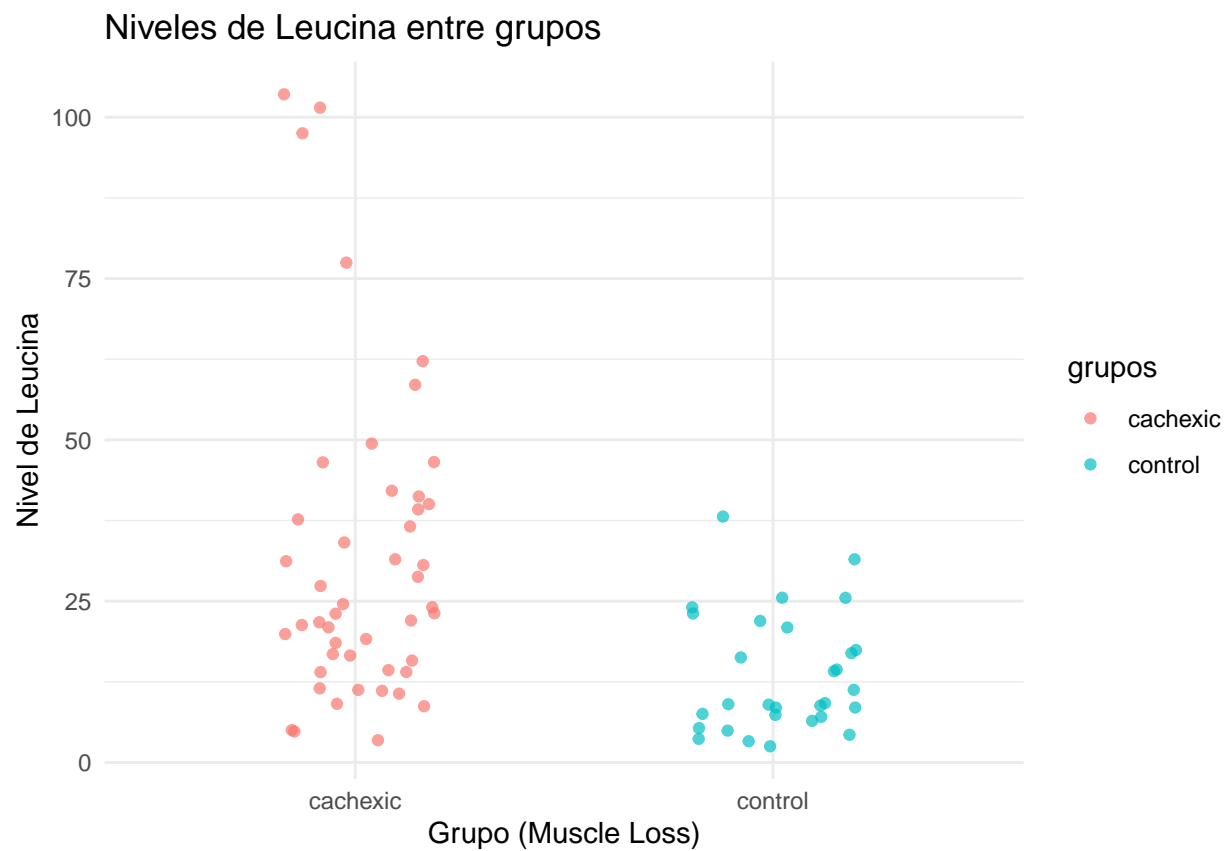
```
t_test_result
```

```
##
## Welch Two Sample t-test
##
## data:  assay(se)["Creatine", ] by colData(se)$Muscle.loss
## t = 2.3988, df = 55.284, p-value = 0.01985
## alternative hypothesis: true difference in means between group cachexic and group control is not equal to 0
## 95 percent confidence interval:
##  20.3217 226.4964
## sample estimates:
## mean in group cachexic mean in group control
##          174.91340          51.50433
```

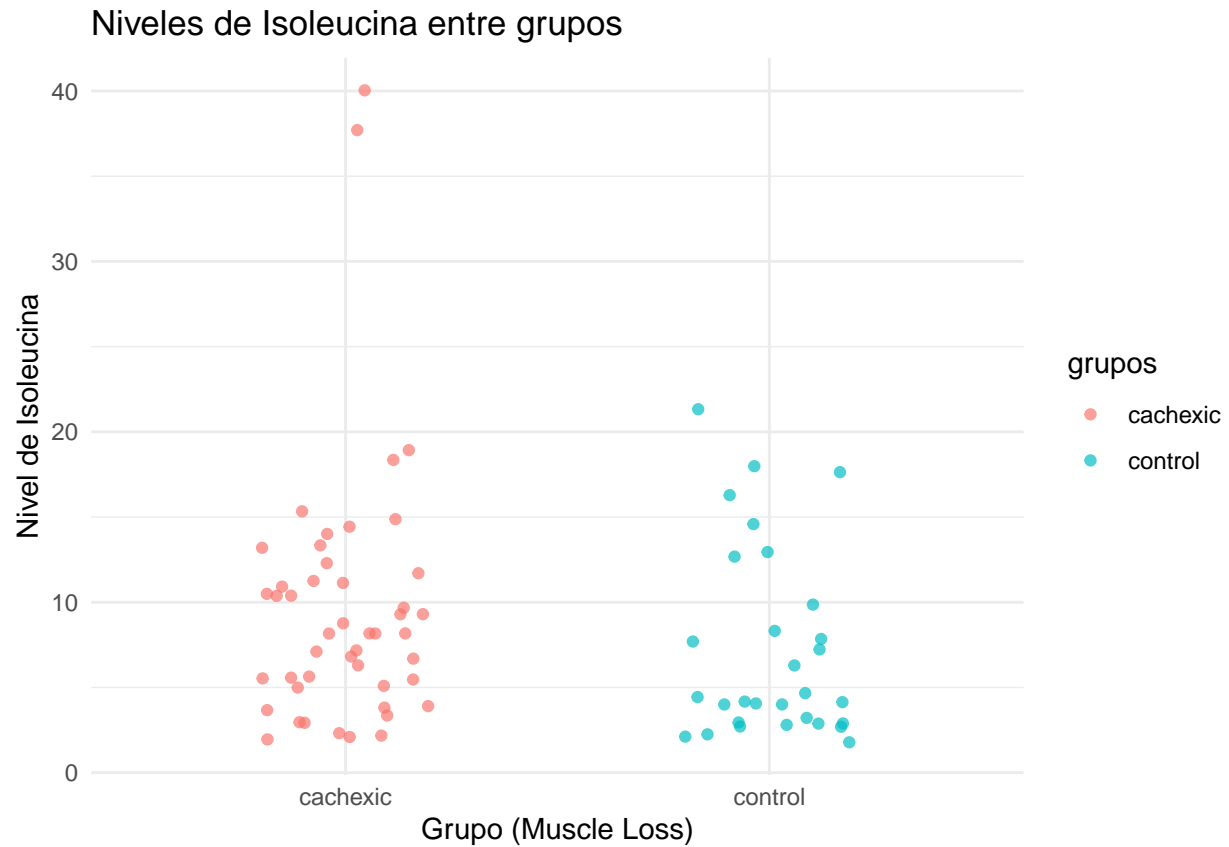
En los graficos de dispersion se observa los niveles de leucina y isoleucina fueron notablemente más altos en el grupo caquéxico, y se observó una mayor variabilidad en estos metabolitos en dicho grupo. Por otro lado, la diferencia en los niveles de Valina entre los grupos fue menor y la variabilidad fue más uniforme en ambos. Esto sugiere que, aunque leucina y Isoleucina presentan una mayor dispersión en el grupo caquéxico, la diferencia en los niveles de valina es menos marcada.

```
#Leucina
ggplot() +
  geom_jitter(aes(x = grupos, y = metabolitos_data["Leucine", ], color = grupos),
    width = 0.2, alpha = 0.7) + # Puntos dispersos
  labs(title = "Niveles de Leucina entre grupos",
```

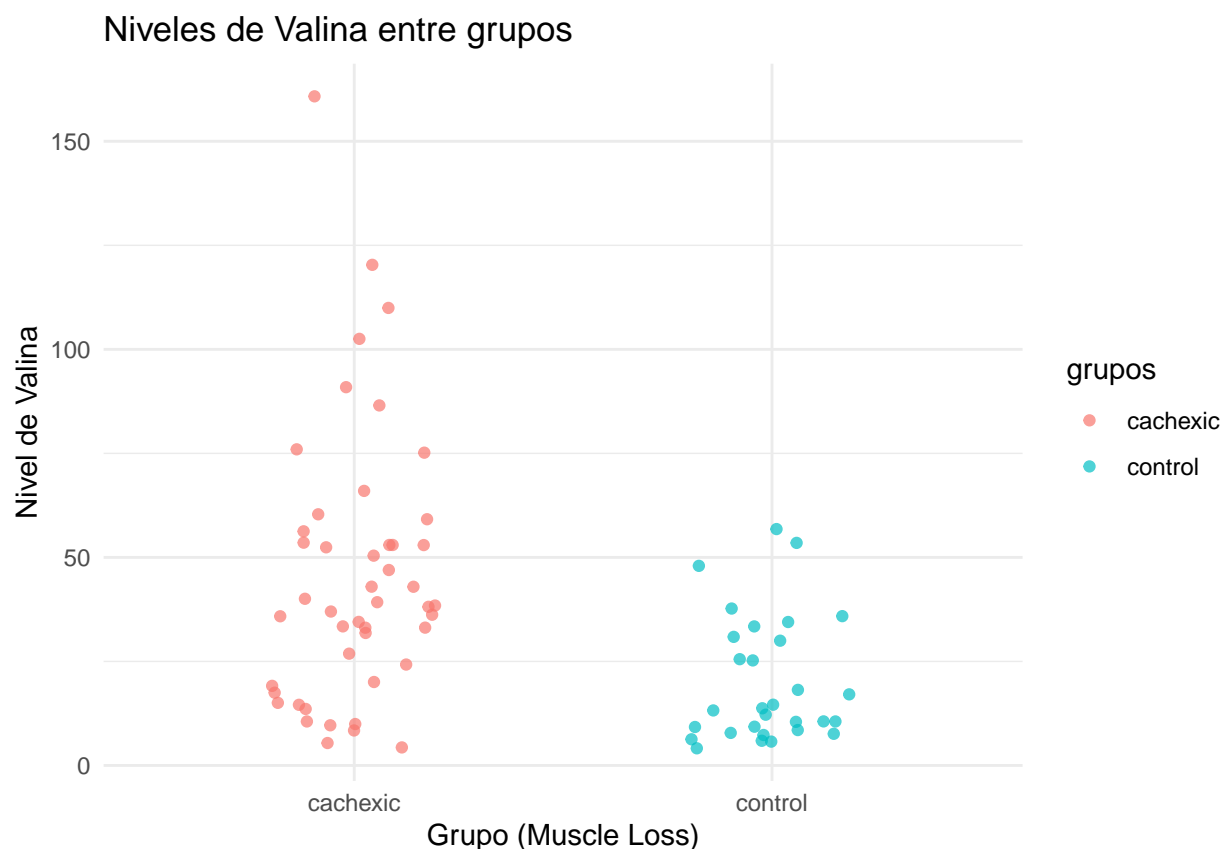
```
x = "Grupo (Muscle Loss)", y = "Nivel de Leucina") +  
theme_minimal()
```



```
# Isoleucina  
ggplot() +  
  geom_jitter(aes(x = grupos, y = metabolitos_data["Isoleucine", ], color = grupos),  
              width = 0.2, alpha = 0.7) + # Puntos dispersos  
  labs(title = "Niveles de Isoleucina entre grupos",  
        x = "Grupo (Muscle Loss)", y = "Nivel de Isoleucina") +  
  theme_minimal()
```

```
# Valina
ggplot() +
  geom_jitter(aes(x = grupos, y = metabolitos_data["Valine", ], color = grupos),
    width = 0.2, alpha = 0.7) + # Puntos dispersos
  labs(title = "Niveles de Valina entre grupos",
    x = "Grupo (Muscle Loss)", y = "Nivel de Valina") +
  theme_minimal()
```



Discusión y Conclusiones

En este estudio, se realizó un análisis de datos metabolómicos de pacientes con caquexia y un grupo control utilizando el contenedor SummarizedExperiment. El análisis incluyó una prueba estadística t-test para evaluar si los niveles de creatina eran significativamente diferentes entre ambos grupos. Los resultados mostraron un valor p menor a 0.05, lo que indica diferencias significativas en los niveles de creatina entre los pacientes con caquexia y el grupo control. Para la visualización de los datos, se utilizó ggplot2, generando gráficos de cajas (boxplots) y de dispersión que ilustraron claramente las diferencias en variables como pimetilhistidina, glucosa y algunos aminoácidos (leucina, isoleucina y valina) entre los grupos, lo que facilitó la comprensión de la distribución de los datos. Además, algunos metabolitos, como la glucosa y los aminoácidos de cadena ramificada, no mostraron resultados consistentes con lo reportado en la bibliografía.

Sin embargo, el análisis tiene limitaciones. El tamaño de la muestra, con 77 pacientes, es relativamente pequeño, lo que dificulta aplicar los resultados a una población más amplia. Además, la falta de unidades de medida en los datos de metabolitos dificultó la interpretación precisa de sus niveles y complicó la comparación con otros estudios. También, la presencia de valores atípicos en algunos metabolitos afectó la interpretación de los resultados. Finalmente, la falta de datos clínicos adicionales limita la comprensión total de los hallazgos.

Bibliografía:

Instituto Nacional del Cáncer. (s. f.). Caquexia: Un síndrome complejo en la investigación del cáncer. Instituto Nacional del Cáncer. Recuperado el 3 de noviembre de 2024 de <https://www.cancer.gov/espanol/>

cancer/tratamiento/investigacion/caquexia

Nishikawa, H., Goto, M., Fukunishi, S., Asai, A., Nishiguchi, S., & Higuchi, K. (2021). Cancer cachexia: Its mechanism and clinical significance. *International Journal of Molecular Sciences*, 22(16), 8491. <https://doi.org/10.3390/ijms22168491>

Yang, Q.-J., Zhao, J.-R., Hao, J., Li, B., Huo, Y., Han, Y.-L., Wan, L.-L., Li, J., Huang, J., Lu, J., Yang, G.-J., & Guo, C. (2017). Serum and urine metabolomics study reveals a distinct diagnostic model for cancer cachexia. *Journal of Cachexia, Sarcopenia and Muscle*, 8(5), 749-760. <https://doi.org/10.1002/jcsm.12246>

Chapela, S., & Martinuzzi, A. (2018). Pérdida de masa muscular en el paciente críticamente enfermo: ¿caquexia, sarcopenia y/o atrofia? impacto en la respuesta terapéutica y la supervivencia. *Revista Cubana de Alimentación y Nutrición*, 28(2), 393-416.

Anexos

Reposición de los Datos en GitHub

Para subir los archivos al repositorio de GitHub, se hizo lo siguiente: Primero se debe tener una cuenta en GitHub, después se creó un repositorio con el nombre de Galvan-Villavicencio-Ximena-PEC1, se cargaron los archivos: el informe en pdf, el objeto contenedor .Rda (cachexia_datos.Rda) y el archivo Rmarkdown.

El repositorio está disponible en: <https://github.com/xgalvan/Galvan-Villavicencio-Ximena-PEC1.git>

Guardamos el objeto SummarizedExperiment en un archivo .Rda

```
save(se, file = "cachexia_datos.Rda")
```