

Primera prueba de evaluación continua.

Judith Pascual Pozurama

Tabla de contenidos

Introducción	1
Objetivos	1
Materiales y métodos.....	1
Resultados.....	2
Apéndice: código de R	14
Referencias	16

Introducción

La finalidad de esta primera PEC es iniciarse en el uso de herramientas clave para el análisis de datos ómicos, especialmente en lo que respecta al manejo de microarrays. Para ello, se llevará a cabo un análisis exploratorio utilizando un conjunto de datos obtenido de un repositorio de Github, específicamente el 2024-Cachexia, que se centra en el síndrome de deterioro progresivo que provoca la pérdida de músculo esquelético y grasa.

Objetivos

Como objetivos específicos podemos señalar los siguientes:

1. Identificar un conjunto de datos (“dataset”) de interés en la tabla proporcionada y descargarlo creando un contenedor del tipo SummarizedExperiment que contenga los datos y los metadatos.
2. Llevar a cabo un análisis exploratorio de los datos que proporcionar una visión general de las variables y de los individuos.

Materiales y métodos

Materiales

En el repositorio de Github se descargarán los datos de human-cachexia utilizando el paquete SummarizedExperiment. Este paquete permite descargar los datos indicados y crear con ellos una estructura de datos del tipo expressionSet que contiene una matriz de datos preprocesados (habitualmente normalizados), así como una tabla con información sobre covariables y otros aspectos del experimento, que, pueden variar de estudio a estudio. El código siguiente muestra como instalar los paquetes necesarios para el estudio.

Métodos

- El principal método bioinformático que se utiliza es la creación, automática, con el paquete SummarizedExperiment de Bioconductor, de clases contenedoras para datos de expresión, ExpressionSets,
- En cuanto a los métodos estadísticos podemos diferenciar:
 - Análisis uni y bivalente de los datos, mediante boxplots y/o histogramas para estudiar la forma general de los mismos y mediante correlogramas y matrices de distancias para estudiar la relación entre variables e individuos.
 - Análisis multivalente de los datos, mediante Análisis de Componentes Principales y Agrupamiento Jerárquico, para determinar si los grupos que aparezcan (en caso de hacerlo) parecen relacionarse con las fuentes de variabilidad del estudio o, si por el contrario, podrían haber otras fuentes de variabilidad como efectos batch.

Resultados

Selección de los datos

Los datos para este estudio se han escogido de un repositorio de GitHub. En este caso, se han obtenido los archivos de cachexia, tanto el dataset como el de metadatos.

Cargar los archivos

```
## standardGeneric for "spec" defined from package "genefilter"
##
## function (object)
## standardGeneric("spec")
## <bytecode: 0x00000234730ff6d8>
## <environment: 0x0000023473105c48>
## Methods may be defined for arguments: object
## Use showMethods(spec) for currently available ones.

## This is the famous cachexia dataset used in several MetaboAnalyst
tutorials
## Available from:
https://rest.xialab.ca/api/download/metaboanalyst/human_cachexia.csv
##
## Successfully passed sanity check!"
## [2] "Samples are not paired."
## [3] "2 groups were detected in samples."
## [6] "All data values are numeric."
## [7] "A total of 0 (0%) missing values were detected."
```

Crear un contenedor

Se crea un script para carga los datos y los metadatos, y posteriormente se crea un objeto SummarizedExperiment que contiene tanto la matriz de expresión como la información asociada a las columnas.

```
## class: SummarizedExperiment
## dim: 77 65
## metadata(0):
## assays(1): counts
## rownames: NULL
## rowData names(0):
## colnames(65): Patient ID Muscle loss ... pi-Methylhistidine
##   tau-Methylhistidine
## colData names(2): SampleID Metadata
```

Se observa la información que contiene sobre el estudio en forma de meta-datos que podemos utilizar como resumen de la información del estudio.

Exploración del dataset

```
##      Patient ID      Muscle loss 1,6-Anhydro-beta-D-glucose 1-
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"
##      2-Aminobutyrate 2-Hydroxyisobutyrate 2-Oxoglutarate 3-
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"
##      3-Hydroxybutyrate 3-Hydroxyisovalerate 3-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"
##      4-Hydroxyphenylacetate Acetate Acetone Adipate Alanine
Asparagine
## [1,] "120.30"      "126.47" " 9.49" " 38.09" " 314.19"
"159.17"
## [2,] "432.68"      "212.72" " 11.82" "327.01" " 871.31"
"157.59"
## [3,] "292.95"      "314.19" " 4.44" "131.63" " 464.05" "
89.12"
## [4,] "214.86"      " 37.34" "206.44" "144.03" " 589.93"
"273.14"
## [5,] " 97.51"      "407.48" " 44.26" " 15.03" "1118.79" "
42.52"
## [6,] "132.95"      " 81.45" " 14.44" " 25.28" " 237.46"
```

```

"157.59"
##      Betaine  Carnitine Citrate    Creatine  Creatinine Dimethylamine
## [1,] "109.95" "265.07" " 3714.50" " 196.37" "16481.60" " 632.70"
## [2,] "244.69" "120.30" " 2617.57" " 212.72" "15835.35" " 607.89"
## [3,] "116.75" " 25.03" "  862.64" " 221.41" "24587.66" " 735.10"
## [4,] "278.66" "200.34" "13629.61" "  85.63" "20952.22" "1064.22"
## [5,] "391.51" " 84.77" "  854.06" " 105.64" " 6768.26" " 242.26"
## [6,] " 66.69" " 40.04" " 1958.63" " 200.34" "15677.78" " 614.00"
##      Ethanolamine Formate   Fucose   Fumarate Glucose   Glutamine
Glycine
## [1,] " 645.48"      " 441.42" "336.97" "  7.69" " 395.44" " 871.31"
"2038.56"
## [2,] " 487.85"      " 252.14" "198.34" "18.92" "8690.62" " 601.85"
"1107.65"
## [3,] " 407.48"      " 249.64" "186.79" "  7.10" "1352.89" " 301.87" "
620.17"
## [4,] " 820.57"      " 468.72" "407.48" "96.54" " 862.64" "1685.81"
"5064.45"
## [5,] " 365.04"      " 114.43" " 26.05" "19.69" "6836.29" " 432.68" "
395.44"
## [6,] " 459.44"      " 314.19" "123.97" "  5.05" " 512.86" " 298.87" "
482.99"
##      Glycolate Guanidoacetate Hippurate  Histidine Hypoxanthine
Isoleucine
## [1,] "685.40" "154.47"      " 4582.50" " 925.19" " 97.51"      "
5.58"
## [2,] "651.97" "109.95"      " 1737.15" " 845.56" " 82.27"      "
8.17"
## [3,] "141.17" "183.09"      " 4315.64" " 284.29" "114.43"      "
9.30"
## [4,] " 70.81" "102.51"      "  757.48" "1043.15" "223.63"
"37.71"
## [5,] " 26.58" " 52.98"      " 1152.86" " 327.01" " 66.69"
"40.04"
## [6,] "428.38" " 57.97"      " 3568.85" " 459.44" " 62.80"      "
8.17"
##      Lactate   Leucine  Lysine   Methylamine Methylguanidine
## [1,] " 106.70" " 42.10" "146.94" "52.46"      "  9.97"
## [2,] " 368.71" " 77.48" "284.29" "23.57"      "  7.69"
## [3,] " 749.95" " 31.50" " 97.51" "18.73"      "  4.66"
## [4,] " 368.71" "103.54" "290.03" "48.91"      "141.17"
## [5,] "3640.95" "101.49" "122.73" "27.94"      "  5.31"
## [6,] " 113.30" " 28.79" "120.30" "36.97"      " 43.38"
##      N,N-Dimethylglycine O-Acetylcarnitine Pantothenate Pyroglutamate
Pyruvate
## [1,] " 23.34"      " 52.98"      " 25.79"      " 437.03"
" 21.12"
## [2,] " 87.36"      " 50.40"      "186.79"      " 437.03"
" 36.97"
## [3,] " 24.53"      "  5.58"      "145.47"      " 713.37"

```

" 29.37"						
## [4,]	" 40.04"	"254.68"	" 42.52"	" 566.80"		
" 64.07"						
## [5,]	" 46.06"	" 45.60"	" 74.44"	" 184.93"		
" 12.30"						
## [6,]	" 24.29"	" 13.46"	" 35.52"	" 432.68"		
" 32.79"						
##	Quinolate	Serine	Succinate	Sucrose	Tartrate	Taurine
Threonine						
## [1,]	"165.67"	" 284.29"	"154.47"	" 45.15"	" 97.51"	"1919.85"
"184.93"						
## [2,]	" 72.97"	" 391.51"	"244.69"	" 459.44"	" 32.79"	"1261.43"
"198.34"						
## [3,]	"192.48"	" 295.89"	"142.59"	" 160.77"	" 16.28"	"4272.69"
"109.95"						
## [4,]	" 86.49"	"1248.88"	"144.03"	" 111.05"	"837.15"	"1525.38"
"376.15"						
## [5,]	" 38.09"	" 206.44"	" 68.72"	" 75.19"	" 4.53"	" 468.72"
64.07"						
## [6,]	"112.17"	" 387.61"	" 33.45"	" 336.97"	" 24.05"	"2059.05"
"105.64"						
##	Trigonelline	Trimethylamine	N-oxide	Tryptophan	Tyrosine	Uracil
Valine						
## [1,]	" 943.88"	"2121.76"		"259.82"	"290.03"	"111.05"
" 86.49"						
## [2,]	" 208.51"	" 639.06"		" 83.10"	"167.34"	" 46.99"
"109.95"						
## [3,]	" 192.48"	"1152.86"		" 82.27"	" 60.34"	" 31.50"
" 59.15"						
## [4,]	" 992.27"	"1450.99"		"235.10"	"323.76"	" 30.57"
"102.51"						
## [5,]	" 86.49"	" 172.43"		"103.54"	"142.59"	" 44.26"
"160.77"						
## [6,]	" 862.64"	" 880.07"		"239.85"	"127.74"	" 29.67"
" 36.97"						
##	Xylose	cis-Aconitate	myo-Inositol	trans-Aconitate	pi-	
Methylhistidine						
## [1,]	" 72.24"	" 237.46"	"135.64"	" 51.94"	" 157.59"	
## [2,]	" 192.48"	" 333.62"	"376.15"	"217.02"	" 307.97"	
## [3,]	"2164.62"	" 330.30"	" 86.49"	" 58.56"	" 145.47"	
## [4,]	" 125.21"	"1863.11"	"247.15"	" 75.94"	" 249.64"	
## [5,]	" 186.79"	" 101.49"	"749.95"	" 98.49"	" 84.77"	
## [6,]	" 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"					

Proceder a la limpieza y convertir los datos a numéricos, para luego realizar el análisis exploratorio de los datos.

```
## Se encontraron valores no numéricos y se convirtieron a NA.
```

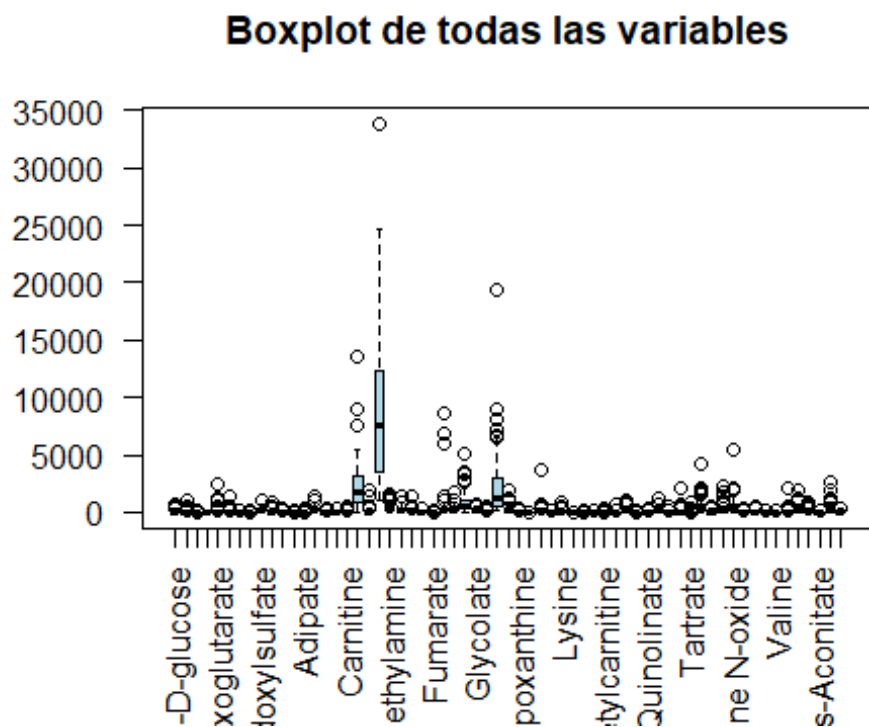
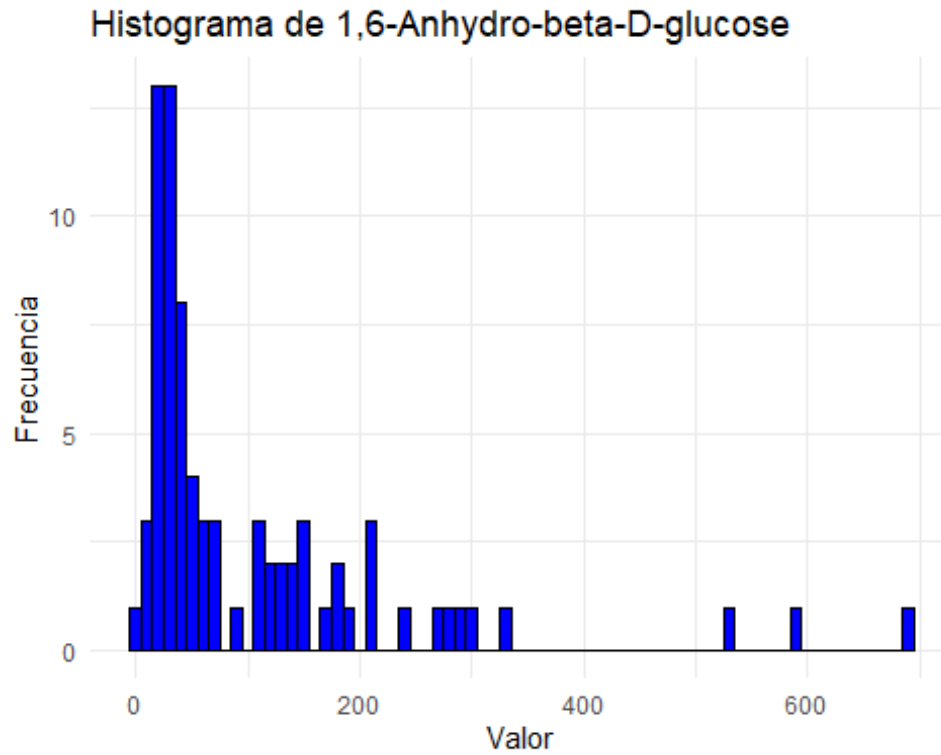
Comprobar que no hay valores numéricos:

```
## Columnas con valores no numéricos:
```

```
## character(0)
```

[Realizar un análisis exploratorio](#)

[Visualización de distribución de variables](#)

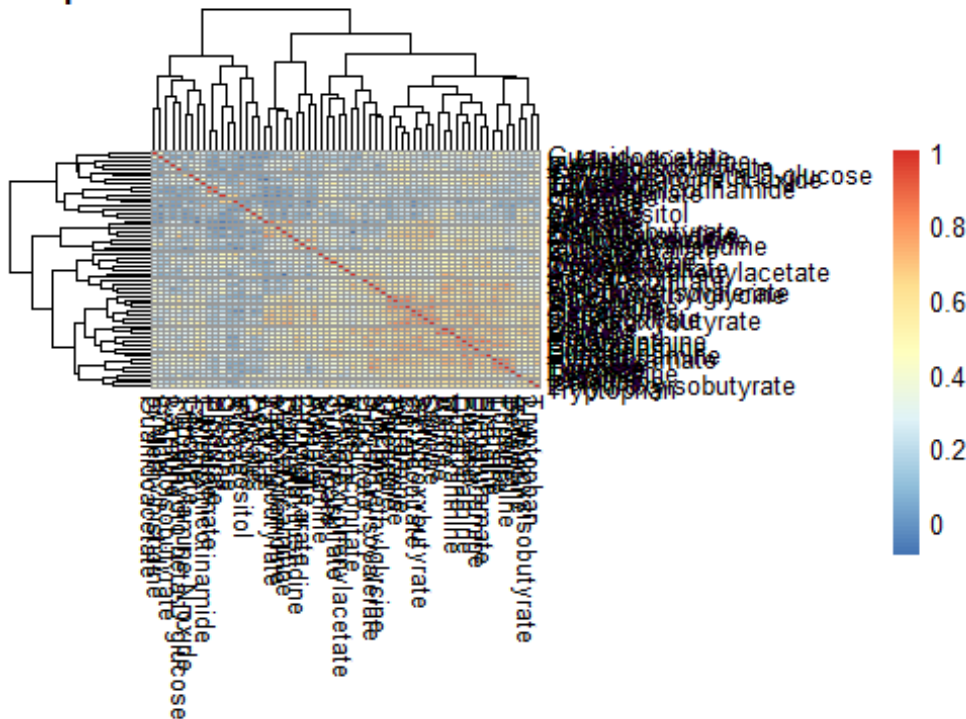


- Histograma: La distribución de la variable sugiere una heterogeneidad metabólica entre los pacientes. Algunos pueden tener niveles de “1,6-Anhydro-beta-D-glucose” que son notablemente más altos. Esta variabilidad puede ser el resultado de diferencias en la etapa de la enfermedad, en la respuesta

inflamatoria, o en factores genéticos y ambientales. Si los niveles son bajos, los pacientes han experimentado un nivel considerable de cachexia.

- Boxplot: Los metabolitos como la creatina y la carnitina, que tienen cajas y bigotes altos, podrían estar relacionados con procesos de degradación muscular, ya que son relevantes en el metabolismo energético y muscular. En cachexia, el desgaste de masa muscular es un síntoma central, y variaciones en estos metabolitos pueden reflejar diferentes grados de catabolismo muscular en los pacientes.

Mapa de calor de la matriz de correlación

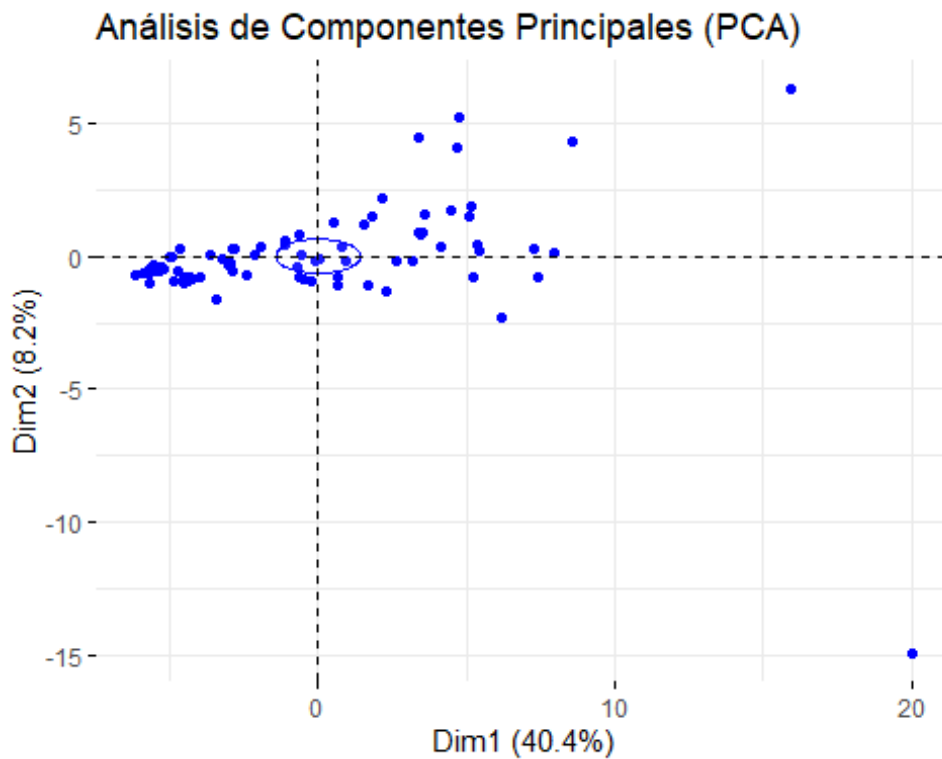


La áreas en rojo indican correlaciones altas, pueden ser resultado de procesos metabólicos conjuntos, donde los niveles de estos metabolitos cambian en respuesta a un mismo factor (como el avance de la cachexia). Las áreas en azul representan pares de metabolitos que no están correlacionados o tienen correlaciones negativas, lo cual indica que no están relacionados entre sí en términos de sus niveles en el conjunto de datos.

La correlación entre ciertos metabolitos puede indicar rutas metabólicas específicas que están alteradas en los pacientes con cachexia. Por ejemplo, si los metabolitos de una vía energética están altamente correlacionados, esto puede reflejar una disfunción metabólica en el manejo de la energía en estos pacientes.

Algunos metabolitos específicos, como aquellos relacionados con el metabolismo de aminoácidos (como alanina y glutamina) o el metabolismo de energía (creatina y carnitina), podrían estar altamente correlacionados entre sí debido a la forma en que la cachexia afecta estos sistemas metabólicos.

Análisis de componentes principales (PCA)

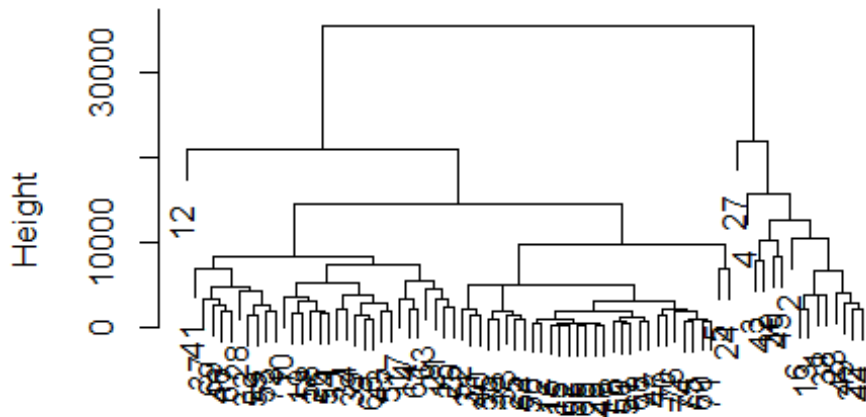


Los puntos alejados del grupo central podrían indicar subgrupos de pacientes. Por ejemplo, los puntos alejados en la dirección positiva de Dim1 podrían corresponder a individuos con un perfil metabólico más distintivo, quizás relacionado con un estado avanzado de cachexia.

Esto sugiere la posibilidad de que algunos pacientes tengan un perfil de metabolitos que los separe de los demás, lo cual podría estar asociado con diferentes etapas o manifestaciones de la enfermedad.

Análisis de agrupamiento

Dendrograma de Agrupamiento Jerárquico



Los grupos identificados pueden corresponder a diferentes etapas o severidades de la cachexia, es decir, algunos grupos pueden incluir pacientes con un perfil metabólico más extremo, asociado con una mayor pérdida de masa muscular, mientras que otros pueden tener un perfil más moderado.

Al analizar los metabolitos que caracterizan cada grupo, se puede identificar perfiles metabólicos específicos en función de la progresión de la enfermedad.

Estos grupos podrían usarse para desarrollar enfoques de tratamiento personalizados. Si ciertos perfiles metabólicos están asociados con respuestas específicas al tratamiento o con pronósticos diferentes, el agrupamiento jerárquico podría ser una herramienta útil para guiar la terapia en función del perfil de cada paciente.

Estadísticas descriptivas

##	1,6-Anhydro-beta-D-glucose	1-Methylnicotinamide	2-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	
##	2-Hydroxyisobutyrate	2-Oxoglutarate	3-Aminoisobutyrate	3-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			
## 3-Hydroxyisovalerate	3-Indoxylsulfate	4-Hydroxyphenylacetate	
Acetate			
## Min. : 0.92	Min. : 27.66	Min. : 15.49	Min.
: 3.49			
## 1st Qu.: 5.26	1st Qu.: 82.27	1st Qu.: 41.68	1st
Qu.: 16.28			
## Median : 12.55	Median : 144.03	Median : 70.11	Median
: 39.65			
## Mean : 21.65	Mean : 218.88	Mean :112.02	Mean
: 66.14			
## 3rd Qu.: 30.27	3rd Qu.: 333.62	3rd Qu.:145.47	3rd
Qu.: 86.49			
## Max. :164.02	Max. :1043.15	Max. :796.32	Max.
:411.58			
## Acetone	Adipate	Alanine	Asparagine
## Min. : 2.29	Min. : 1.55	Min. : 16.78	Min. : 6.69
## 1st Qu.: 4.95	1st Qu.: 6.11	1st Qu.: 78.26	1st Qu.: 20.49
## Median : 7.10	Median : 10.18	Median : 194.42	Median : 42.10
## Mean : 11.43	Mean : 24.76	Mean : 273.56	Mean : 62.28
## 3rd Qu.: 10.49	3rd Qu.: 19.11	3rd Qu.: 399.41	3rd Qu.: 89.12
## Max. :206.44	Max. :327.01	Max. :1312.91	Max. :273.14
## Betaine	Carnitine	Citrate	Creatine
## Min. : 2.29	Min. : 2.18	Min. : 59.74	Min. : 2.75
## 1st Qu.: 28.79	1st Qu.: 14.44	1st Qu.: 788.40	1st Qu.: 17.64
## Median : 64.72	Median : 23.81	Median : 1790.05	Median : 44.26
## Mean : 90.32	Mean : 52.09	Mean : 2235.35	Mean : 126.83
## 3rd Qu.:127.74	3rd Qu.: 60.95	3rd Qu.: 3071.74	3rd Qu.: 117.92
## Max. :391.51	Max. :487.85	Max. :13629.61	Max. :1863.11
## Creatinine	Dimethylamine	Ethanolamine	Formate
## Min. : 1002	Min. : 41.26	Min. : 16.12	Min. : 6.42
## 1st Qu.: 3498	1st Qu.: 142.59	1st Qu.: 86.49	1st Qu.: 53.52
## Median : 7631	Median : 304.90	Median : 204.38	Median : 95.58
## Mean : 8734	Mean : 358.17	Mean : 276.26	Mean : 147.40
## 3rd Qu.:12333	3rd Qu.: 454.86	3rd Qu.: 407.48	3rd Qu.: 167.34
## Max. :33860	Max. :1556.20	Max. :1436.55	Max. :1480.30
## Fucose	Fumarate	Glucose	Glutamine
## Min. : 5.70	Min. : 0.79	Min. : 26.84	Min. : 23.34
## 1st Qu.: 29.37	1st Qu.: 2.23	1st Qu.: 80.64	1st Qu.: 113.30
## Median : 61.56	Median : 4.10	Median : 210.61	Median : 225.88
## Mean : 88.67	Mean : 8.44	Mean : 559.85	Mean : 306.87

## 3rd Qu.:123.97	3rd Qu.: 7.85	3rd Qu.: 407.48	3rd Qu.: 445.86
## Max. :407.48	Max. :96.54	Max. :8690.62	Max. :1685.81
## Glycine	Glycolate	Guanidoacetate	Hippurate
## Min. : 38.09	Min. : 5.42	Min. : 7.03	Min. : 92.76
## 1st Qu.: 262.43	1st Qu.: 50.91	1st Qu.: 33.78	1st Qu.: 492.75
## Median : 528.48	Median :130.32	Median : 64.72	Median : 1224.15
## Mean : 880.72	Mean :187.99	Mean : 86.37	Mean : 2286.84
## 3rd Qu.:1096.63	3rd Qu.:267.74	3rd Qu.:108.85	3rd Qu.: 2921.93
## Max. :5064.45	Max. :720.54	Max. :561.16	Max. :19341.34
## Histidine	Hypoxanthine	Isoleucine	Lactate
## Min. : 14.15	Min. : 3.78	Min. : 1.790	Min. : 7.32
## 1st Qu.: 66.69	1st Qu.: 20.70	1st Qu.: 3.900	1st Qu.: 35.52
## Median : 174.16	Median : 40.04	Median : 7.170	Median : 81.45
## Mean : 292.64	Mean : 61.10	Mean : 8.709	Mean : 158.46
## 3rd Qu.: 419.89	3rd Qu.: 83.93	3rd Qu.:11.250	3rd Qu.: 139.77
## Max. :1863.11	Max. :265.07	Max. :40.040	Max. :3640.95
## Leucine	Lysine	Methylamine	Methylguanidine
## Min. : 2.51	Min. : 10.49	Min. : 1.51	Min. : 1.70
## 1st Qu.: 9.12	1st Qu.: 30.27	1st Qu.: 5.26	1st Qu.: 4.26
## Median : 19.11	Median : 69.41	Median :14.73	Median : 7.85
## Mean : 24.36	Mean :108.79	Mean :17.38	Mean : 15.32
## 3rd Qu.: 31.19	3rd Qu.:121.51	3rd Qu.:24.05	3rd Qu.: 19.30
## Max. :103.54	Max. :788.40	Max. :52.46	Max. :141.17
## N,N-Dimethylglycine	O-Acetylcarnitine	Pantothenate	Pyroglutamate
## Min. : 0.79	Min. : 1.23	Min. : 2.59	Min. :
21.33			
## 1st Qu.: 7.03	1st Qu.: 3.94	1st Qu.: 11.13	1st Qu.:
68.72			
## Median : 21.98	Median : 11.47	Median : 22.65	Median :
157.59			
## Mean : 26.35	Mean : 19.73	Mean : 44.88	Mean :
211.45			
## 3rd Qu.: 40.04	3rd Qu.: 20.91	3rd Qu.: 41.26	3rd Qu.:
301.87			
## Max. :120.30	Max. :254.68	Max. :692.29	Max.:
:1064.22			
## Pyruvate	Quinolate	Serine	Succinate
## Min. : 0.90	Min. : 5.21	Min. : 16.12	Min. : 1.72
## 1st Qu.: 4.85	1st Qu.: 26.58	1st Qu.: 83.10	1st Qu.: 8.58
## Median : 13.46	Median : 51.42	Median : 142.59	Median : 30.88
## Mean : 21.29	Mean : 66.44	Mean : 197.69	Mean : 60.23
## 3rd Qu.: 29.08	3rd Qu.: 87.36	3rd Qu.: 270.43	3rd Qu.: 74.44
## Max. :184.93	Max. :259.82	Max. :1248.88	Max. :589.93
## Sucrose	Tartrate	Taurine	Threonine
## Min. : 6.49	Min. : 2.20	Min. : 17.81	Min. : 8.25
## 1st Qu.: 19.30	1st Qu.: 6.89	1st Qu.: 99.48	1st Qu.: 31.82
## Median : 40.85	Median : 12.94	Median : 249.64	Median : 64.07
## Mean : 113.23	Mean : 40.00	Mean : 525.12	Mean : 95.36
## 3rd Qu.: 94.63	3rd Qu.: 25.79	3rd Qu.: 665.14	3rd Qu.:137.00
## Max. :2079.74	Max. :837.15	Max. :4272.69	Max. :450.34

##	Trigonelline	Trimethylamine N-oxide	Tryptophan	Tyrosine
##	Min. : 10.07	Min. : 55.7	Min. : 8.67	Min. :
##	1st Qu.: 53.52	1st Qu.: 175.9	1st Qu.: 21.33	1st Qu.:
##	Median : 114.43	Median : 383.8	Median : 46.99	Median :
##	Mean : 270.44	Mean : 652.2	Mean : 66.24	Mean :
##	3rd Qu.: 340.36	3rd Qu.: 735.1	3rd Qu.: 96.54	3rd
##	Max. : 2252.96	Max. : 5486.2	Max. : 259.82	Max.
##	Uracil	Valine	Xylose	cis-Aconitate
##	Min. : 3.10	Min. : 4.10	Min. : 10.07	Min. : 12.94
##	1st Qu.: 11.94	1st Qu.: 12.18	1st Qu.: 29.96	1st Qu.: 36.23
##	Median : 27.39	Median : 33.12	Median : 50.40	Median : 129.02
##	Mean : 35.56	Mean : 35.67	Mean : 100.93	Mean : 204.22
##	3rd Qu.: 44.26	3rd Qu.: 50.40	3rd Qu.: 89.12	3rd Qu.: 254.68
##	Max. : 179.47	Max. : 160.77	Max. : 2164.62	Max. : 1863.11
##	myo-Inositol	trans-Aconitate	pi-Methylhistidine	tau-
##	Methylhistidine			
##	Min. : 11.59	Min. : 4.90	Min. : 11.36	Min. : 8.00
##	1st Qu.: 30.27	1st Qu.: 12.43	1st Qu.: 67.36	1st Qu.: 27.39
##	Median : 78.26	Median : 26.84	Median : 162.39	Median : 68.72
##	Mean : 135.40	Mean : 40.63	Mean : 370.29	Mean : 89.69
##	3rd Qu.: 167.34	3rd Qu.: 57.40	3rd Qu.: 387.61	3rd Qu.: 130.32
##	Max. : 854.06	Max. : 217.02	Max. : 2697.28	Max. : 317.35

Estos datos reflejan una alta variabilidad en los niveles de metabolitos en los pacientes estudiados, lo cual es consistente con lo que se esperaría en el contexto de una enfermedad compleja y diversa como la cachexia. Algunos patrones observados en los datos:

- Sesgo en la distribución: muchos metabolitos tienen una distribución sesgada a la derecha, lo que indica la presencia de valores muy altos en un subgrupo de pacientes.

- Presencia de valores atípicos: metabolitos como creatinina, 3-Aminoisobutyrate, y taurina muestran valores máximos muy altos, sugiriendo la existencia de outliers. Estos outliers pueden representar subgrupos específicos de pacientes con una alteración metabólica más pronunciada.

- Indicadores potenciales de severidad: metabolitos como creatinina, glucosa, y 3-Indoxylsulfate, que tienen altos valores en algunos pacientes, podrían servir como biomarcadores de la severidad de la cachexia.

Apéndice: código de R

Paquetes:

```
if (!require(BiocManager)) install.packages("BiocManager")
installifnot <- function (pkg){
  if (!require(pkg, character.only=T)){
    BiocManager::install(pkg)
  }
}
installifnot("SummarizedExperiment")
installifnot("PCAtools")
installifnot("pd.mogene.1.0.st.v1")
installifnot("mogene10sttranscriptcluster.db")
installifnot("oligo")
installifnot("limma")
installifnot("Biobase")
installifnot("arrayQualityMetrics")
installifnot("genefilter")
installifnot("annotate")
installifnot("xtable")
installifnot("gplots")
installifnot("GOstats")
```

Cargar de archivos:

```
library(readr)
genefilter::spec

# Cargar el archivo description.md
metadata <- read_file("C:/Users/judii/Desktop/MÁSTER BIOINFORMÁTICA Y
BIOESTADÍSTICA/ANÁLISIS DATOS ÓMICOS/2/PEC1-ADO/description.md")
metadata_info <- paste(metadata, collapse = "\n")

# Imprimir el contenido del archivo de descripción
cat(metadata_info)
```

Crear contenedor y limpiar datos:

```
# Crear la matriz de expresión
expression_matrix <- as.matrix(human_cachexia)

# Crear un objeto DataFrame para las columnas (muestras)
colData <- DataFrame(SampleID = colnames(expression_matrix))

# Agregar metadatos al objeto colData
colData$Metadata <- rep(metadata_info, ncol(expression_matrix))

# Crear un objeto SummarizedExperiment
se <- SummarizedExperiment(
  assays = list(counts = expression_matrix),
  colData = colData
)
```

```

# Mostrar información general del objeto
se

# Mostrar Las primeras filas de La matriz de expresión
head(assay(se))

# Convertir Los datos a numéricos, forzando NA si hay valores no
numéricos
expression_data <- as.data.frame(assay(se))
expression_data[] <- lapply(expression_data, function(x)
as.numeric(as.character(x)))

# Verificar si La conversión generó NAs
if (any(is.na(expression_data))) {
  cat("Se encontraron valores no numéricos y se convirtieron a NA.\n")
}

# Eliminar columnas con todos Los valores NA (si Las hay)
expression_data <- expression_data[, colSums(is.na(expression_data)) <
nrow(expression_data)]

# Recalcular La matriz de correlación
corr_matrix <- cor(expression_data, use = "pairwise.complete.obs")

non_numeric_cols <- sapply(expression_data, function(x)
sum(is.na(as.numeric(as.character(x)))))
problematic_cols <- names(non_numeric_cols[non_numeric_cols > 0])
cat("Columnas con valores no numéricos:\n")
print(problematic_cols)
expression_data[] <- lapply(expression_data, function(x) {
  x <- as.character(x)
  x[grepl("[^0-9.]", x)] <- NA # Reemplaza cualquier cosa que no sea un
número o un punto decimal con NA
  as.numeric(x)
})

# Recalcular La matriz de correlación después de La limpieza
corr_matrix <- cor(expression_data, use = "pairwise.complete.obs")

```

Análisis exploratorio:

Visualización de distribución de variables

```

library(ggplot2)

# Histograma de una variable seleccionada
ggplot(expression_data, aes(x = `1,6-Anhydro-beta-D-glucose`)) +
  geom_histogram(binwidth = 10, fill = "blue", color = "black") +
  theme_minimal() +
  labs(title = "Histograma de 1,6-Anhydro-beta-D-glucose", x = "Valor", y
= "Frecuencia")

```

```
# Boxplot para todas las variables
boxplot(expression_data, main = "Boxplot de todas las variables", las =
2, col = "lightblue")

library(pheatmap)

# Mapa de calor de la matriz de correlación
pheatmap(corr_matrix, main = "Mapa de calor de la matriz de correlación")
```

Análisis de componentes principales (PCA)

```
library(FactoMineR)
library(factoextra)

# Ejecutar PCA
pca_result <- PCA(expression_data, graph = FALSE)

# Visualizar las primeras dos componentes principales
fviz_pca_ind(pca_result, geom.ind = "point", col.ind = "blue",
             addEllipses = TRUE, ellipse.type = "confidence",
             title = "Análisis de Componentes Principales (PCA)")
```

Análisis de agrupamiento

```
# Calcular la distancia entre las muestras
dist_matrix <- dist(expression_data)
hclust_result <- hclust(dist_matrix)

# Visualizar el dendrograma
plot(hclust_result, main = "Dendrograma de Agrupamiento Jerárquico", xlab
= "", sub = "")
```

Estadísticas descriptivas

```
summary_stats <- summary(expression_data)
print(summary_stats)
```

Referencias