

PEC1

Alba Arranz

2024-11-04

Tabla de contenidos

1. Abstract.....	2
2. Objetivos del studio.....	2
3. Materiales y métodos.....	2
4. Resultados.....	2
5. Conclusiones y limitaciones.....	5
6. Anexo.....	6

Abstract

Este estudio presenta un proceso de análisis de datos hecho con R. Los datos han sido descargados de la base de datos MetabolomicsWorkbenchR en formato Summary Experiment. Los datos han sido cargados en el paquete POMA para hacer un análisis exploratorio que incluye un análisis univariante y multivariante. Los resultados sugieren un posible cambio en los metabolitos aunque no son concluyentes.

Objetivos del estudio

El objetivo de este estudio es la obtención de unos datos metabolómicos de un repositorio y llevar a cabo un análisis exploratorio de estos.

Materiales y métodos

Los datos corresponden a diferentes metabolitos detectados por espectrometría de masas (GCMS) en muestras intestinales pre y post trasplante. Los datos han sido obtenidos del paquete 'metabolomicsWorkbenchR' en un contenedor SummarizedExperiment del paquete con el mismo nombre. Para el análisis exploratorio se ha utilizado el paquete POMA. Primero se han buscado valores perdidos, y después los datos han sido normalizados. Los datos normalizados se han graficado en boxplots y también se han graficado los posibles outliers.

Para el análisis exploratorio primero se ha hecho un análisis univariante con el modelo t-student para hacer una comparación de la presencia de los diferentes metabolitos en los dos grupos antes y después del trasplante. Después se ha hecho un análisis multivariante mediante el análisis de componentes principales (PCA).

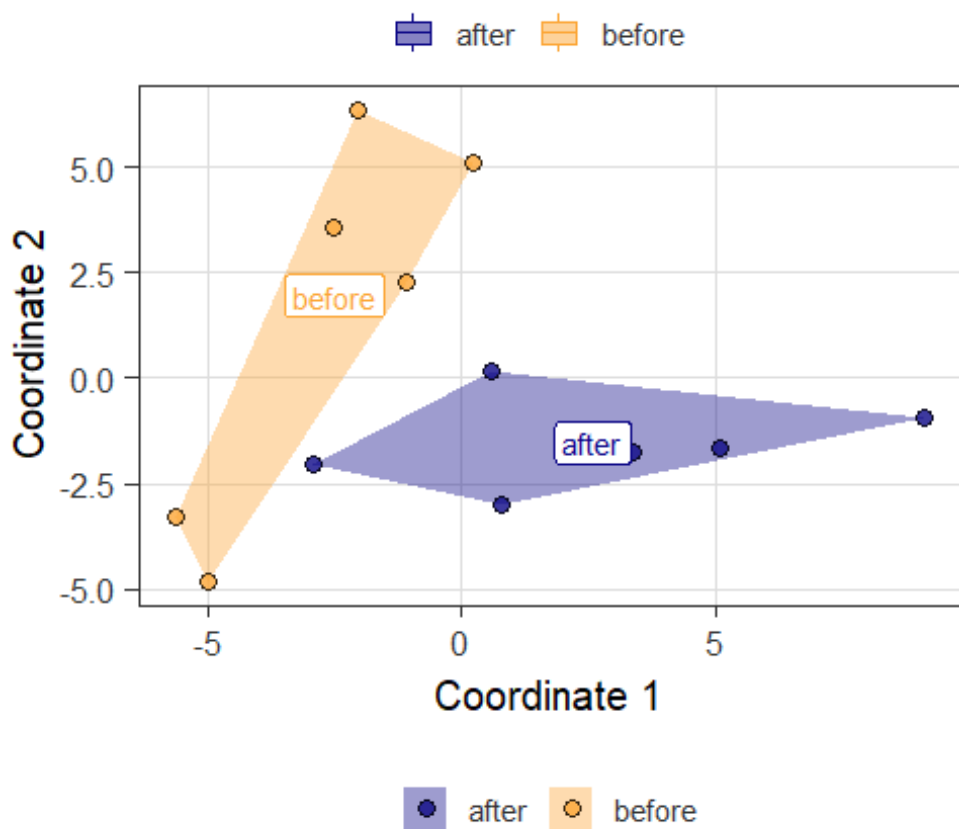
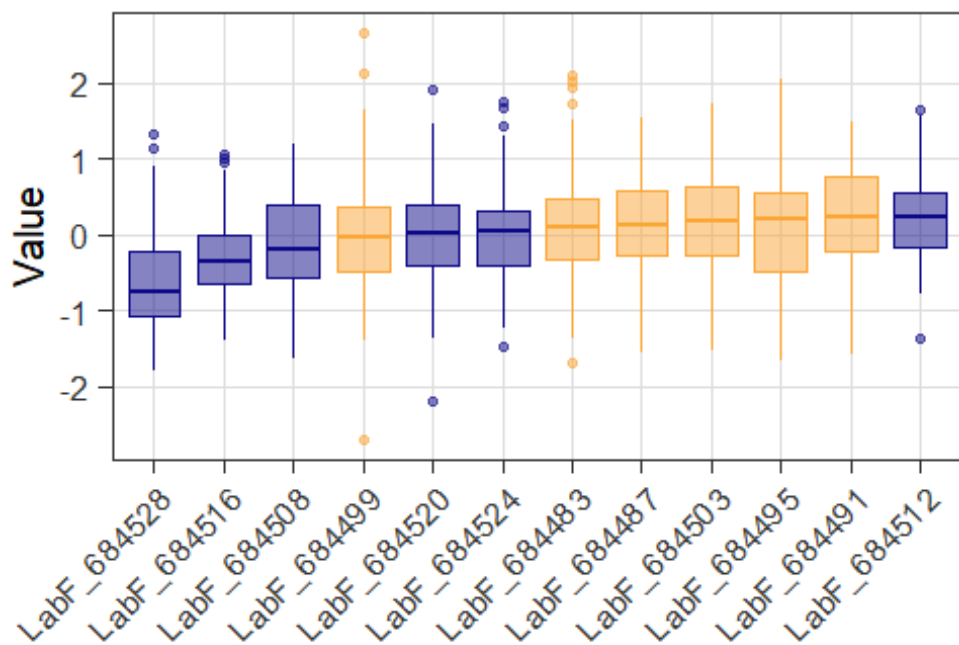
Resultados

Datos

Los datos del estudio contienen 12 muestras, de las cuáles 6 han sido obtenidas antes de un trasplante y las otras 6 restantes después. Para estas 12 muestras se ha analizado por espectrometría de masas la presencia de 142 metabolitos diferentes.

Con los datos normalizados, mediante un boxplot se observa que las muestras anteriores al trasplante presentan una ligera disminución en la presencia de metabolitos excepto para una muestra de post trasplante que es la que tiene la

media más elevada de metabolitos. A pesar de haber cierta dispersión dentro de muestras del mismo grupo no se detecta ningún outlier en ellas.



Análisis univariante

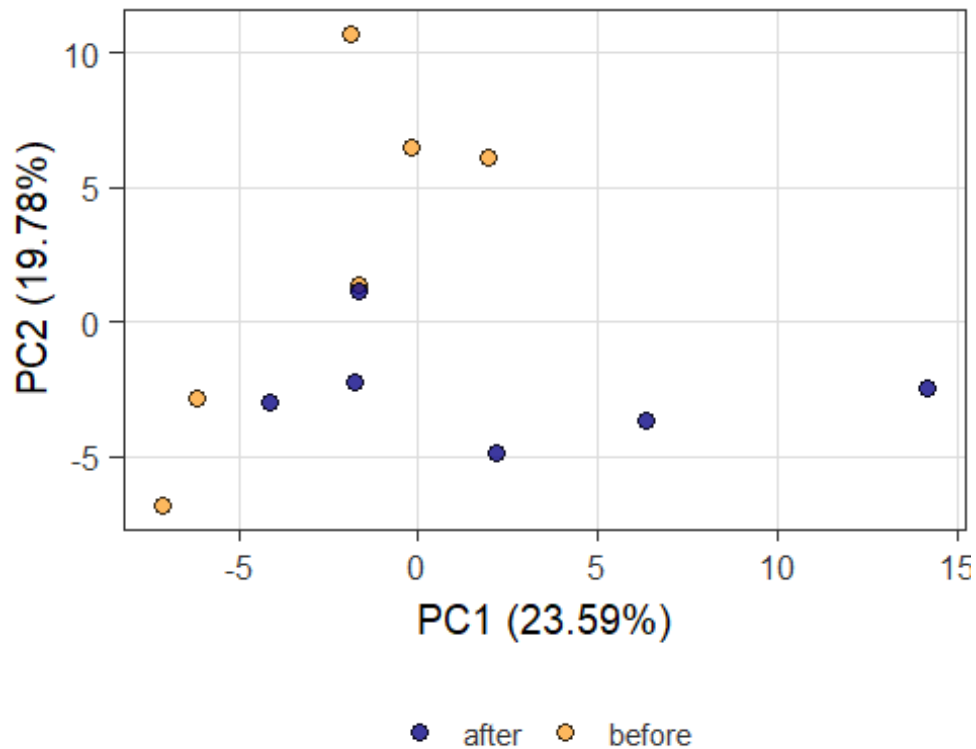
Del análisis univariante podemos observar 5 metabolitos con un p valor inferior a 0.1 y por tanto con un diferencia significativa. Estos 5 metabolitos son: ácido fosfórico, ácido treónico, piruvato, 2-monooctadecanoato, levanbiosa. Únicamente el ácido fosfórico aumenta tras el transplante, mientras que el resto disminuyen.

Variación significativa de metabolitos

metabolitos	result.fo ld_chan ge	result.d iff_me ans	resul t.pval ue	result.a dj_pval ue	result. mean_ after	result.m ean_be fore	result. sd_aft er	result. sd_bef ore
Ácido fosfórico	-1	-1.158	0.00 2151 8	0.0890 441	0.5789 953	- 0.5789 953	0.551 5047	0.2912 252
Ácido treónico	-1	1.001	0.00 2246 3	0.0890 441	- 0.5005 952	0.5005 952	0.398 3429	0.4483 731
Piruvato	-1	1.214	0.00 2516 9	0.0890 441	- 0.6070 583	0.6070 583	0.594 2935	0.3135 637
2- monooct adecano ato	-1	1.218	0.00 2773 9	0.0890 441	- 0.6089 716	0.6089 716	0.297 1105	0.6047 527
Levanbi osa	-1	1.388	0.00 3135 4	0.0890 441	- 0.6940 017	0.6940 017	0.416 0816	0.7062 611

Análisis multivariante

El análisis de las componentes principales muestra un ligero agrupamiento de las muestras de antes del transplante más abajo mientras que las de después se agrupan más a la derecha. Este agrupamiento no es muy obvio y además las dos primeras componentes explican un 43.29% de la variabilidad.



Conclusiones y limitaciones.

Los resultados obtenidos del análisis realizado sugieren que podría haber cierto cambio en la presencia de metabolitos en muestras intestinales antes y después del trasplante tendiendo a disminuir. Pero es cierto que en el análisis de componentes principales podría mostrar cierto efecto batch en los resultados. También el bajo número de muestras analizadas impide sacar conclusiones más conclusivas. Se sugiere en un futuro repetir el experimento aumentando el número de muestras a analizar.

Repositorio github: <https://github.com/albaag96/ARRANZ-Garcia-Alba-PEC1>

Anexo

```

if (!require("BiocManager", quietly = TRUE))
  install.packages("BiocManager")

## Bioconductor version '3.19' is out-of-date; the current release
## version '3.20' is available with R version '4.4'; see
## https://bioconductor.org/install

if (!require("SummarizedExperiment", quietly = TRUE))
  install.packages("SummarizedExperiment")
if (!require("metabolomicsWorkbenchR", quietly = TRUE))
  install.packages("metabolomicsWorkbenchR")

library(SummarizedExperiment)
library(metabolomicsWorkbenchR)

# Descargar estudio de metabolomicsWorkbenchR
experiment = do_query(
  context = 'study',
  input_item = 'study_id',
  input_value = 'ST000002',
  output_item = 'SummarizedExperiment'
)

# Descripción de los datos
metadata(experiment)

## $data_source
## [1] "Metabolomics Workbench"
## $study_id
## [1] "ST000002"
## $analysis_id
## [1] "AN000002"
## $analysis_summary
## [1] "GCMS positive ion mode"
## $units
## [1] "Peak height"
## $name
## [1] "ST000002:AN000002"
## $description
## [1] "Intestinal Samples II pre/post transplantation"

```

```
##                                                                 $subject_type
## [1] NA

colData(experiment)

##      DataFrame      with      12      rows      and      6      columns
##      local_sample_id      study_id      sample_source      mb_sample_id
raw_data
##      <character> <character>      <character>      <character>
<character>
## LabF_684483      LabF_684483      ST000002      Intestine      SA000025
## LabF_684487      LabF_684487      ST000002      Intestine      SA000026
## LabF_684491      LabF_684491      ST000002      Intestine      SA000027
## LabF_684495      LabF_684495      ST000002      Intestine      SA000028
## LabF_684499      LabF_684499      ST000002      Intestine      SA000029
## ...              ...              ...              ...              ...
...
## LabF_684512      LabF_684512      ST000002      Intestine      SA000032
## LabF_684516      LabF_684516      ST000002      Intestine      SA000033
## LabF_684520      LabF_684520      ST000002      Intestine      SA000034
## LabF_684524      LabF_684524      ST000002      Intestine      SA000035
## LabF_684528      LabF_684528      ST000002      Intestine      SA000036
##                                                                 Transplantation
##                                                                 <factor>
##              LabF_684483      Before      transplantation
##              LabF_684487      Before      transplantation
##              LabF_684491      Before      transplantation
##              LabF_684495      Before      transplantation
##              LabF_684499      Before      transplantation
## ...              ...
##              LabF_684512      After      transplantation
##              LabF_684516      After      transplantation
##              LabF_684520      After      transplantation
##              LabF_684524      After      transplantation
## LabF_684528 After transplantation

colData(experiment)$Transplantation

##      [1] Before transplantation Before transplantation Before
transplantation
##      [4] Before transplantation Before transplantation Before
transplantation
##      [7] After transplantation After transplantation After
transplantation
##     [10] After transplantation After transplantation After
transplantation
## Levels: After transplantation Before transplantation

rowData(experiment)
```

```
##      DataFrame      with      142      rows      and      3      columns
##      metabolite_name metabolite_id      refmet_name
##      <character>      <character>      <character>
## ME641269      1-monoolein      ME641269      MG 18:1(9Z)/0:0/0:0
## ME641270      1-monostearin      ME641270      MG 18:0/0:0/0:0
## ME641271 2-hydroxybutanoic acid      ME641271 2-Hydroxybutyric acid
## ME641272 2-hydroxyglutaric acid      ME641272 2-Hydroxyglutaric acid
## ME641273 2-ketoisocaproic acid      ME641273      Ketoleucine
## ...      ...      ...
## ME641406      uric acid      ME641406      Uric acid
## ME641407      uridine      ME641407      Uridine
## ME641408      valine      ME641408      Valine
## ME641409      xanthine      ME641409      Xanthine
## ME641410      xylose      ME641410      Xylose
```

```
assay(experiment)
```

```
##      LabF_684483 LabF_684487 LabF_684491 LabF_684495 LabF_684499
LabF_684503
##      <num>      <num>      <num>      <num>      <num>
<num>
## 1:      244142      6968      1928      19228      3029
23277
## 2:      16848      10206      9398      1013      4190
11114
## 3:      11587      65635      32433      1823      4429
30427
## 4:      8276      12402      20964      25913      2709
70972
## 5:      2096      3472      10669      432      1055
1005
##      ---
## 138:      269872      1896      150207      66759      327982
136704
## 139:      528      253      464      2915      190
246
## 140:      430644      918195      826226      1307288      1073522
1080473
## 141:      627      1828      651      4236      282
3798
## 142:      9049      12730      9733      6323      9324
1007
##      LabF_684508 LabF_684512 LabF_684516 LabF_684520 LabF_684524
LabF_684528
##      <num>      <num>      <num>      <num>      <num>
<num>
## 1:      6047      2902      1452      3428      2985
16334
## 2:      9771      6521      1302      2781      5789
4338
```



```

##      3:      13238      29774      4134      4419      13334
2115
##      4:      7160      11501      3202      17238      20376
1109
##      5:       812      2011      738      2550      871
628
##
## 138:      30387      714      3899      1038      523718
580
## 139:       351      3894      283      4108      637
242
## 140:     1124633     1227642     652835     992247     1394436
222010
## 141:       3704      8842      4668      10089      2521
1467
## 142:       3533      3267      1848      12902      2070
758

assay_data      <-      assay(experiment)
assay_data      <-      assay_data[,      -1]

#      Guardar      el      Summarized      Experiment      en      formato      binario
save(experiment,      file      =      'experiment.Rda')

if      (!require("POMA",      quietly      =      TRUE))
  BiocManager::install("POMA")

library(POMA)
library(ggtext)
library(magrittr)
library(readxl)
library(dplyr)
library(ggplot2)
library(readr)

#      Modificar      datos      y      cargar      en      POMA
metadata      <-      as.data.frame(colData(experiment))[,      c(1,      6)]
metadata      <-      metadata
  mutate(
    Transplantation      =      recode(Transplantation,
      "Before      transplantation"      =      "before",
      "After      transplantation"      =      "after")
  )
View(metadata)
View(assay(experiment))
data      <-      PomaCreateObject(metadata      =      metadata,      features      =
t(assay(experiment)))

```

```

# Detectar y eliminar valores perdidos
imputed <- PomaImpute(method = "knn", zeros_as_na = TRUE, remove_na = TRUE, cutoff = 20)

## No missing values detected

# Normalización
normalized <- PomaNorm(method = "log_pareto", imputed)

# Visualización de datos no-normalizados y normalizados
data$metada

## NULL

PomaBoxplots(imputed, x = "samples")

```

```

PomaBoxplots(normalized, x = "samples")

```

```

PomaOutliers(normalized)$polygon_plot

```

```

# Análisis univariante
poma_uni <- PomaUnivariate(normalized)

## method argument is empty. T-test will be used

poma_uni$result

## # A tibble: 142 x 9
##   feature fold_change diff_means pvalue adj_pvalue mean_after
##   <chr>      <dbl>      <dbl>   <dbl>    <dbl>      <dbl>
## 1 V114      -1      -1.16  0.00215  0.0890    0.579
## 2 V129      -1       1.00  0.00225  0.0890   -0.501
## 3 V117      -1       1.21  0.00252  0.0890   -0.607
## 4 V7        -1       1.22  0.00277  0.0890   -0.609
## 5 V86       -1       1.39  0.00314  0.0890   -0.694

```

```
## 6 V94 -1 -1.59 0.00445 0.105 0.795 -
0.795
## 7 V82 -1 1.12 0.00952 0.193 -0.559
0.559
## 8 V43 -1 1.12 0.0114 0.202 -0.559
0.559
## 9 V122 -1 1.28 0.0157 0.232 -0.639
0.639
## 10 V19 -1 0.731 0.0175 0.232 -0.366
0.366
## # 132 more rows
## # 2 more variables: sd_after <dbl>, sd_before <dbl>

# Análisis de componentes principales
poma_pca <- PomaPCA(normalized)
poma_pca$factores_plot
```

```
View(poma_pca$loadings)
```