

Análisis de datos ómicos - PEC1

Juan Luis Hernández García

<https://github.com/juanluishg/Hernandez-Garcia-Juan-Luis-PEC1>

Descarga de datos y la reposición de los datos en github

Para la descarga de los datos a utilizar para este proceso, primero debemos clonarnos el repositorio que contiene los datasets

```
git clone https://github.com/nutrimetabolomics/metaboData.git
```

Después accedemos a la carpeta "Datasets" y seleccionamos el deseado para hacer la exploración de los datos, en mi caso he escogido "2023-UGrX-4MetaboAnalystTutorial".

Creamos un repositorio en github, a través de su entono web y posteriormente lo inicializamos el repositorio en una carpeta local que hayamos creado donde previamente hemos copiado el dataset a utilizar:

```
git init
git remote add origin
https://github.com/juanluishg/Hernandez-Garcia-Juan-Luis-PEC1.git
```

Cada vez que queramos subir las modificaciones a github, añadimos los ficheros que deseemos, después hacemos commit con un mensaje y por último hacer push para subir los cambios a github.

```
git add .
git commit -m "Dataset"
git push -u origin main
```

Creación del contenedor

Para la creación del contenedor de datos SummarizedExperiment, primero debemos instalar este paquete de Bioconductor, con la ayuda de BioCManager.

```
BiocManager::install("SummarizedExperiment")
```

Posteriormente procedemos a utilizarlo. Primero debemos de leer el fichero “ST000002_AN000002.txt” que contiene toda la información del dataset, después separamos cada una de las secciones que vamos a utilizar para crear el objeto SummarizedExperiment.

- MS_METABOLITE_DATA_START: Contiene los datos del experimento, es decir, los metabolitos, las muestras y los valores de estos
- SUBJECT_SAMPLE_FACTORS: Contiene la lista de muestras asi como metadatos de si estas muestras se obtuvieron antes o después del trasplante.

Con estos valores extraídos, creamos el SummarizedExperiment donde el assays sera los datos del experimento, el colData los metadatos y el rowData los nombres de los metabolitos.

Exploración de datos

Primero hacemos un análisis exploratorio de los datos, en el que mostramos las primeras filas, en las que podemos ver algunos valores para cada una de las muestras.

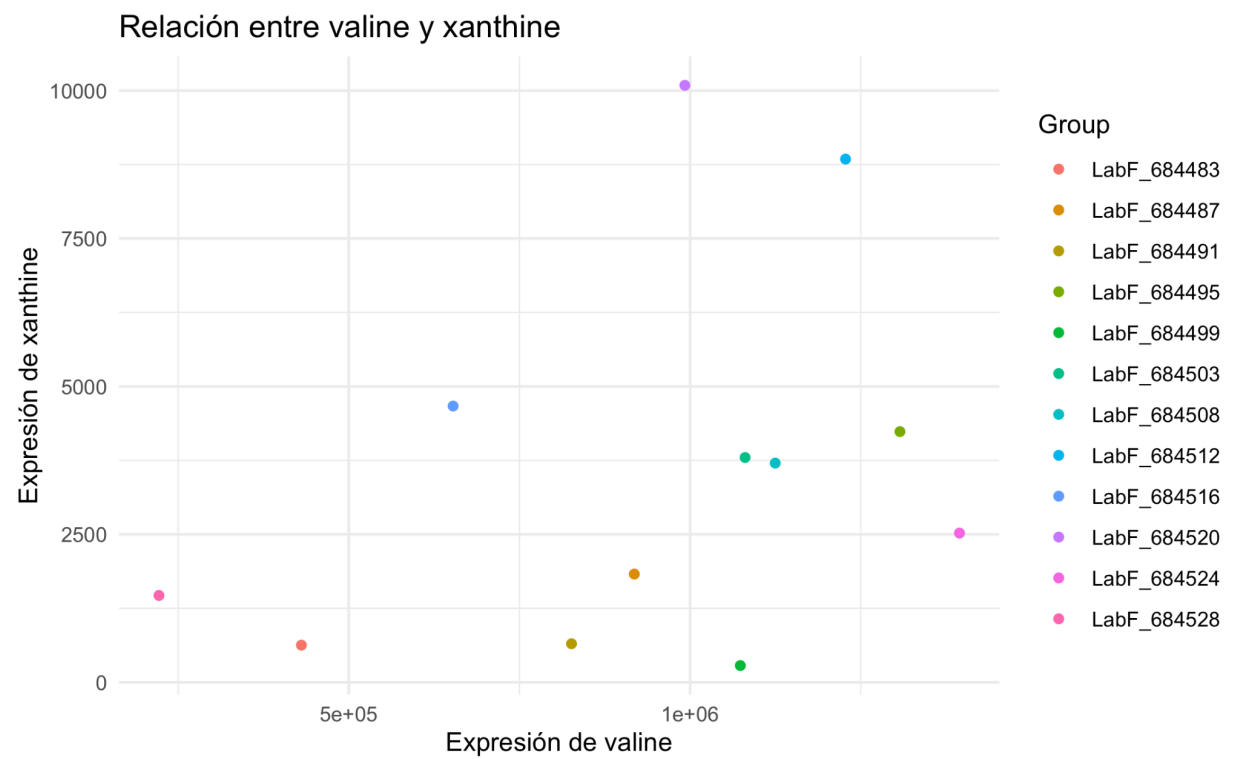
Después imprimimos los metadatos, que muestran cada una de las muestras y si se han tomado antes o después del trasplante.

También escribimos los distintos metabolitos que tenemos en este experimento.

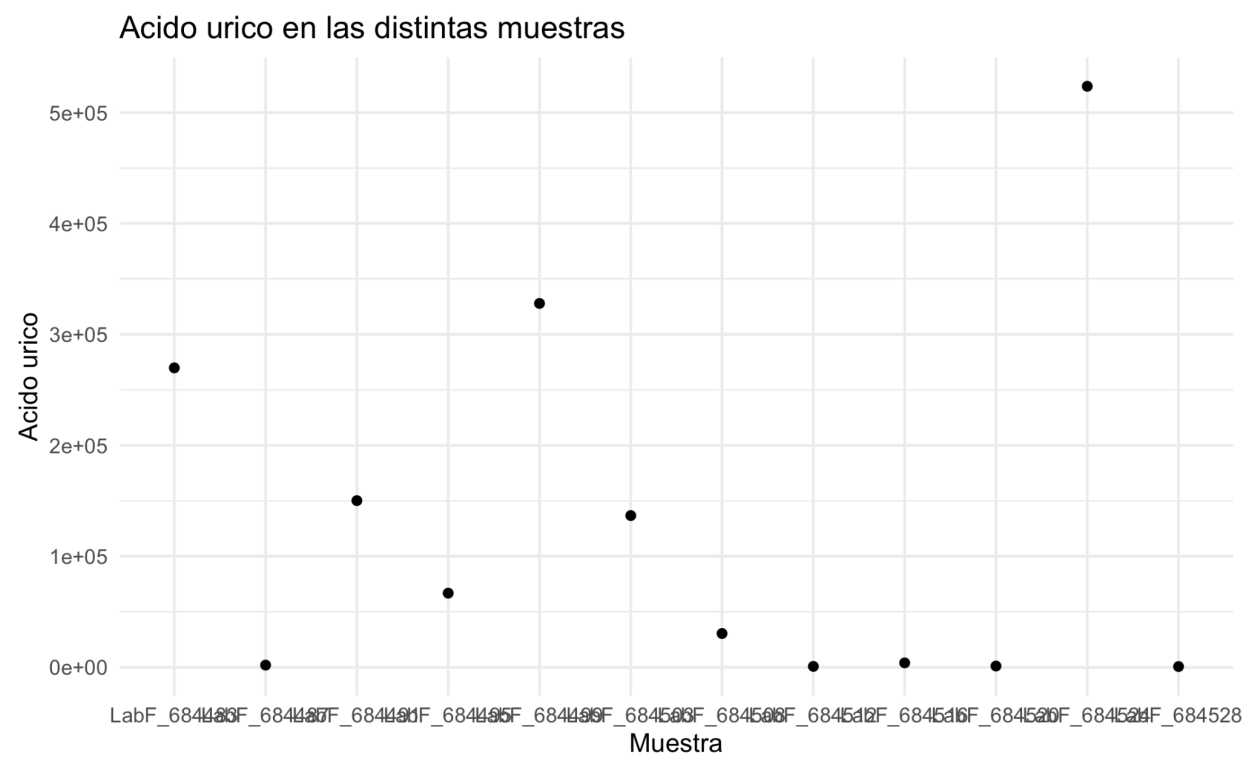
```
# A SummarizedExperiment-tibble abstraction: 1,704 x 7
# Features=142 | Samples=12 | Assays=counts
  .feature      .sample    counts Subject      Sample      Factors      metabolite
  <chr>         <chr>      <dbl> <chr>      <chr>      <chr>      <chr>
1 1-monoolein   LabF_684508    6047 SUBJECT_SAMPLE_FACTORS LabF_684508 Transplantation:After transplantation 1-monoolein
2 1-monostearin LabF_684508    9771 SUBJECT_SAMPLE_FACTORS LabF_684508 Transplantation:After transplantation 1-monostearin
3 2-hydroxybutanoic acid LabF_684508   13238 SUBJECT_SAMPLE_FACTORS LabF_684508 Transplantation:After transplantation 2-hydroxybutanoic acid
4 2-hydroxyglutaric acid LabF_684508    7160 SUBJECT_SAMPLE_FACTORS LabF_684508 Transplantation:After transplantation 2-hydroxyglutaric acid
5 2-ketoisocaproic acid LabF_684508     812 SUBJECT_SAMPLE_FACTORS LabF_684508 Transplantation:After transplantation 2-ketoisocaproic acid
6 2-monopalmitin LabF_684508    1511 SUBJECT_SAMPLE_FACTORS LabF_684508 Transplantation:After transplantation 2-monopalmitin
7 2-monostearin NIST LabF_684508    2060 SUBJECT_SAMPLE_FACTORS LabF_684508 Transplantation:After transplantation 2-monostearin NIST
8 3-aminoisobutyric acid LabF_684508    9147 SUBJECT_SAMPLE_FACTORS LabF_684508 Transplantation:After transplantation 3-aminoisobutyric acid
9 3-hydroxybutanoic acid LabF_684508   19381 SUBJECT_SAMPLE_FACTORS LabF_684508 Transplantation:After transplantation 3-hydroxybutanoic acid
10 3-hydroxypropionic acid LabF_684508    2166 SUBJECT_SAMPLE_FACTORS LabF_684508 Transplantation:After transplantation 3-hydroxypropionic acid
# i 40 more rows
```

Por último pasamos a hacer algunos gráficos para representar los datos de una forma más visual.

El primer gráfico representa la relación entre dos metabolitos (valine y xanthine) entre las distintas muestras.



El segundo gráfico representa los valores de un metabolito en concreto (uric acid) para las distintas muestras.



Exportar datos

Exportamos los datos en formato Rda con la función `save`, primero del objeto `SummarizedExperiment` completo y luego del data frame que contiene los metadatos.

Después exportamos los datos, extrayéndolos con la función `assay` y lo escribimos como texto en un fichero.

Por último vamos a crear un fichero markdown con los metadatos, en él vamos a incluir el número de metabolitos, de muestras, así como algunas filas de los datos disponibles.