Informe

Jairo Aguirre

Noviembre - 2024

Table of Contents

## 1. Resumen

En esta prueba de evaluación continua hemos creado un contenedor del tipo SummarizedExperiment que contiene datos y metadatos de un dataset que ha sido descargado del repositorio del [repositorio de GitHub](https://github.com/nutrimetabolomics/metaboData/tree/main/Datasets/2023-CIMCBTutorial). Se ha llevado a cabo una exploración del dataset, proporcionando una visión general sobre cómo trabajar con este tipo de clases en R.

El proceso para explorar los datos fue el siguiente:

1. Descargar los datos del repositorio GitHub
2. Crear nuestro SummarizedExperiment
3. Explorar el objeto (se) para entender su estructura y contenido
4. Visualizar y explicar algunos de los resultados observados

## 2. Objetivos

El objetivo principal de esta actividad es consolidar los conocimientos y competencias adquiridos sobre los diferentes tipos de herramientas y su manejo bioinformático con herramientas como Galaxy o Bioconductor, a través del estudio de un SummarizedExperiment creado a partir de un dataset.

Adicionalmente, se evaluará la calidad de los datos utilizando las métricas Perc\_missing y QC\_RSD, las cuales se explican más adelante.

## 3. Preparación del sistema

### 3.1 Instalación y carga de paquetes

Se deben cargar las librerías necesarias para la realización de esta PEC

### 3.2 Preparar los componentes de nuestra clase

Lo primero que hacemos es leer los datos con los que trabajaremos. Nuesto excel contiene dos hojas, por lo tanto asignaremos cada hoja a un objeto.

Para crear nuestra clase debemos separar nuestros datos de expresión (assays) y los metadatos (información fenotípica o covariables de las muestras).

Los datos de expresión los encontramos en la hoja 1 y se corresponden con las columnas que contienen los valores númericos (M1 - M149) Los metadatos de las muestras los encontramos en la hoja 2 que hemos denominado peak.

### 3.3 Creación de nuestra clase

Ahora procedemos a crear nuestra clase SummarizedExperiment

Con rowData vamos a añadir caracteristicas de cada muestra

## 4 Exploración del SummarizedExperiment

Cuando ejecutamos class(colData(se)) obtenemos la clase del elemento colData en el objeto SummarizedExperiment (se)

Podemos observar las primeras filas de nuestro colData con head(colData(se))

Con el comando head(assay()) observamos las primeras filas de nuestro assay, es decir, nuestra matriz de datos con los valores observados en cada muestra (columnas) por cada metabolito (filas).

Al ejecutar head(rowData()) observaremos las primeras filas de una matriz que nos ofrece información descriptiva específica sobre las muestras.

## 5.Resultados

Una vez tenemos clara esta información, podríamos inferir:

Por ejemplo, el valor de 11.428571 en la primera fila significa que aproximadamente el 11.43% de las mediciones para “1\_3-Dimethylurate” no están presentes o no se detectaron.

El metabolito “2-Aminobutyrate” presenta un valor de QC\_RSD de 46.9771 lo que podría indicar mayor variabilidad y posibles problemas en la detección o cuantificación. Un valor alto de QC\_RSD puede sugerir inconsistencias en la cuantificación de los metabolitos.

## 6. Discusión

En esta actividad, se ha creado un contenedor SummarizedExperiment que permite integrar datos y metadatos obtenidos de un análisis de metabolómica. La exploración realizada ha proporcionado una visión clara sobre la estructura de los datos y ha facilitado identificación relevante de patrones en los metabolitos analizados.

La interpretación biológica de estos resultados puede ser compleja. Los metabolitos pueden estar influenciados por múltiples factores, como la dieta, el estilo de vida y condiciones ambientales, lo que complica su análisis. Hay que tener en cuenta n tamaño de muestra insuficiente puede limitar la capacidad de generalizar los resultados y afectar la validez estadística de las conclusiones (Sigma-Aldrich, n.d).

Esta actividad demuestra la utilidad de los objetos SummarizedExperiment para manejar datos de metabolómica de manera eficiente. La integración de datos y metadatos en un único objeto facilita la interpretación y el análisis posterior.

## 7. Referencias

Roland (2013). What is the practical difference between data.frame and data.table in R? Stack Overflow. Retrieved from <https://stackoverflow.com/questions/18001120/what-is-the-practical-difference-between-data-frame-and-data-table-in-r>

Sigma-Aldrich. (n.d.). Metabolomics research. Sigma-Aldrich. <https://www.sigmaaldrich.com/ES/es/applications/research-disease-areas/metabolomics-research?srsltid=AfmBOopAGXHGMoMymGHUDLDu2Qw7GCmsGVKyqa_ejkC-PlTFBLXMiKSG>

## 8. Link GitHub

[Link](https://github.com/jairoaguirre2/Aguirre-Bedoya-Jairo-PEC1)