

Informe PEC1 – Análisis de datos ómicos

Carlos Serrano Pinto

1. Introducción

En este trabajo se realizó un análisis de datos obtenidos desde la base de datos Metabolomics Workbench, donde en primera instancia se exploraron los distintos tópicos de estudio, donde se destacan algunos como: Cáncer, Diabetes, Obesidad, Enfermedades oculares, Bacterias, entre otros. En este caso, se optó por el tópico de Bacterias, precisamente el estudio **ST003521**. Se escogió este estudio debido a mi interés en bacterias y la resistencia a antibióticos. En este estudio, asociado a la publicación “Metabolic profiling unveils enhanced antibacterial synergy of polymyxin B and teixobactin against multi-drug resistant *Acinetobacter baumannii*” (<https://doi.org/10.1038/s41598-024-78769-6>), se aborda la sinergia antibacterial de dos compuestos: polymyxin B y Leu₁₀-teixobactin. Experimentalmente, se reveló una sinergia importante de los compuestos combinados contra *A. baumannii* resistente a polymyxin B, demostrando la inhibición del crecimiento de las bacterias luego de 24 horas.

Por otro lado, se estudió el metabolismo de las bacterias aplicando los compuestos por si solos y en combinación, donde se espera que las rutas metabólicas de lípidos, biogénesis de envoltura celular y respiración se vieran principalmente afectados.

Los resultados del artículo dan luces de un potencial tratamiento para prevenir la resistencia en *A. baumannii*.

2. Desarrollo

Para la realización de este trabajo, se escribió un flujo de trabajo detallado en RMarkdown, el cual se encuentra en el repositorio <https://github.com/cserranopinto/Serrano-Pinto-Carlos-PEC1> con el nombre de PEC1-Omica-CSP.Rmd y compilado a formato pdf con el mismo nombre.

Para la obtención de datos, se utilizó la librería **metabolomicsWorkbenchR**, la que permitió explorar algunos estudios directamente explorando los objetos creados en Rstudio. Esta librería permite fácilmente realizar consultas directamente a la base de datos Metabolomics Workbench con palabras claves en ciertos campos. Primero se realizó una consulta general de estudios

disponibles asociados a la palabra clave “Bacterial”, donde se obtuvieron alrededor de 50 opciones. Explorando la descripción de ellos, se escogió el estudio ST003521. Para obtener los datos, se realizó una nueva consulta con la palabra clave “ST003521” y con salida “summary” para obtener un poco más de información, donde, por ejemplo, se detalla el nombre del estudio, las especies involucradas, el instituto donde se desarrolló, el número de muestras y la tecnología utilizada. Además de lo mencionado anteriormente, otro factor para trabajar con este estudio fue que presentaba un artículo. Previamente se exploraron otros estudios, pero no siempre se contaba con un artículo publicado o simplemente los datos no se encontraban disponibles para obtenerlos mediante este método.

Posteriormente, se obtuvieron los resultados del estudio realizando una consulta con el código del estudio y especificando en el formato de salida “SummarizedExperiment”. De esta manera, se obtuvo una lista de dos objetos del tipo SummarizedExperiment. Este tipo de objeto es usado en distintas bases de datos de ómicas, la cual es una extensión de un primero objeto llamado ExpressionSet. Ambos son objetos que permiten contener información estructurada estandarizada para el análisis de datos, particularmente biológicos, sin embargo, el objeto ExpressionSet fue diseñado inicialmente para el análisis de microarreglos. Por otro lado, el objeto SummarizedExperiment ha sido utilizado generalmente para experimentos basados en secuenciación, pero también permite almacenar datos de otras tecnologías, por ejemplo, de HPLC. Otra característica para destacar entre ambos objetos es que las filas de ExpressionSet son características, mientras que en SummarizedExperiment las filas corresponden a GenomicRanges, rangos que representan expresión diferencial.

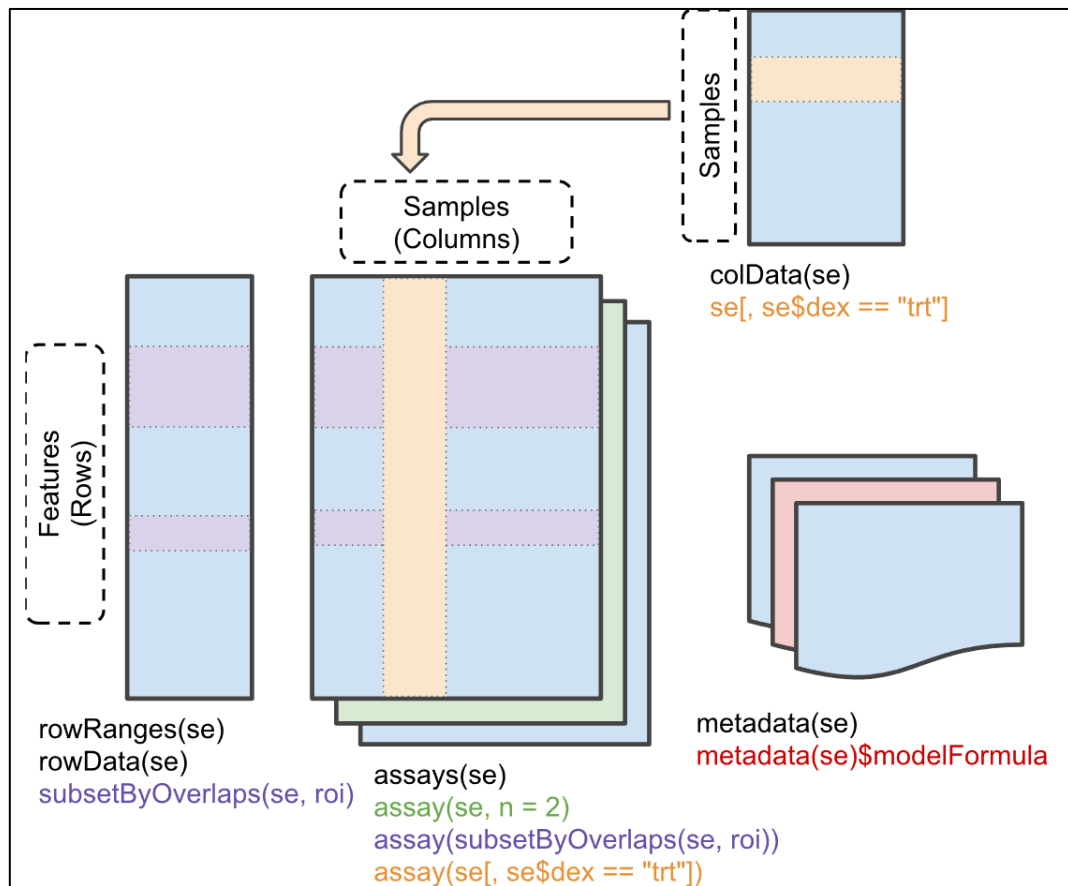


Figura 1: Esquema de objeto SummarizedExperiment. Extraído de: <https://bioconductor.org/packages/release/bioc/vignettes/SummarizedExperiment/inst/doc/SummarizedExperiment.html>

Una vez obtenidos los datos, se confirmó que los objetos de la lista correspondían a objetos del tipo SummarizedExperiment consultando su clase con la función `class()`. Se exploraron las distintas matrices que presenta el objeto: assays, Features, Samples y metadata.

Para acceder a la matriz assay se utilizó la función `assay()` para cada objeto de la lista obtenida. Para el primer objeto (AN005782), la matriz contaba con 513 fila y 58 columnas. Para el segundo (AN005783), la matriz contaba con 497 filas y 58 columnas. Las columnas se compartían en ambos objetos y correspondían a condiciones experimentales en las que fue sometida el organismo de estudio *A. baumannii*, siendo COM, PMB y TXB a distintos tiempos (1, 3 y 6 horas). Además, se contaban con réplicas. También se encontraban las condiciones de controles a los mismos tiempo y mediciones de blancos.

Se continuó explorando la matriz Features, donde se accedió usando la función `rowData()`. Aquí se obtuvieron la misma cantidad de filas que en assay, pero con 3 columnas. Las columnas corresponden al id de distintos metabolitos medidos, su nombre científico y nombre de referencia (No todos los metabolitos presentaban nombre de referencia).

La matriz Sample fue inspeccionada con la función `colData()`. Aquí se presentaban la misma cantidad de filas que en assay pero con 7 columnas. Las columnas presentaban mayor detalle descriptivo de las condiciones a las que fue medido el organismo.

Por último, se accedió a la matriz metadata, mediante a función `metadata()`. Aquí se obtiene una lista con información como: Base de datos, id del estudio, código del análisis, unidad de medida, resumen de técnica y descripción del estudio.

Revisando cada una de las matrices, se pudo obtener una idea de como se organizan los datos, qué posibles caminos se deben realizar para analizar los datos e información detallada de los nombres de metabolitos que se pueden estudiar.

3. Análisis

Como se observó anteriormente, la matriz assay cuenta con columnas que corresponden a replicados. Se decidió promediar los valores de las condiciones si correspondían a duplicados, de tal manera de reducir las variables. Para ello, las matrices assay se convirtieron a dataframe para poder manejar fácilmente sus columnas y filas. Se tomaron los nombres más detallados de las variables y también los nombres químicos de los metabolitos, que originalmente eran códigos, y se renombraron filas y columnas. Se determinaron nombres únicos de las condiciones y se utilizaron como variable para poder agrupar y promediar los valores correspondientes. Esto se realizó para ambos objetos. Una vez que se tienen ambos dataframes renombrados, se unieron, obteniendo un dataframe de 1010 filas y 12 columnas.

Como se quiere estudiar el cambio producido por la acción de los compuestos, se generó un nuevo dataframe que contiene el nivel de cambio de los metabolitos respecto a su muestra control. Para ello se calculó el logaritmo de

FoldChange ($\log_2 FC$), el cual representa cuánto ha variado la medición de un metabolito respecto a un control. Este se calculó con:

$$\log_2(FC) = \log_2\left(\frac{\textit{Condicion Experimental}}{\textit{Condicion Control}}\right)$$

Para visualizarlo, se utilizó la librería ComplexHeatmap, el cual permite graficar Heatmaps. Dada la gran cantidad de datos y para fines prácticos se decidió graficar solo un set de datos. Se procuró ordenar los datos por tiempo y tipo de control, por lo que el heatmap se generó separándolo en 3, cada sección correspondiente a un tiempo. En filas se indica el nombre del metabolito y en columnas y parte inferior el control correspondiente. A la derecha del heatmap se muestra la escala de colores. En 0, que representa que no hubo cambio, se presenta con color blanco. Una mayor expresión respecto al control se presenta con tonos rojizos, mientras que una disminución en la expresión se presenta con colores azulados. A continuación, se muestra la figura obtenida (Revisar el pdf correspondiente a la compilación de RMarkdown para mayor detalle).

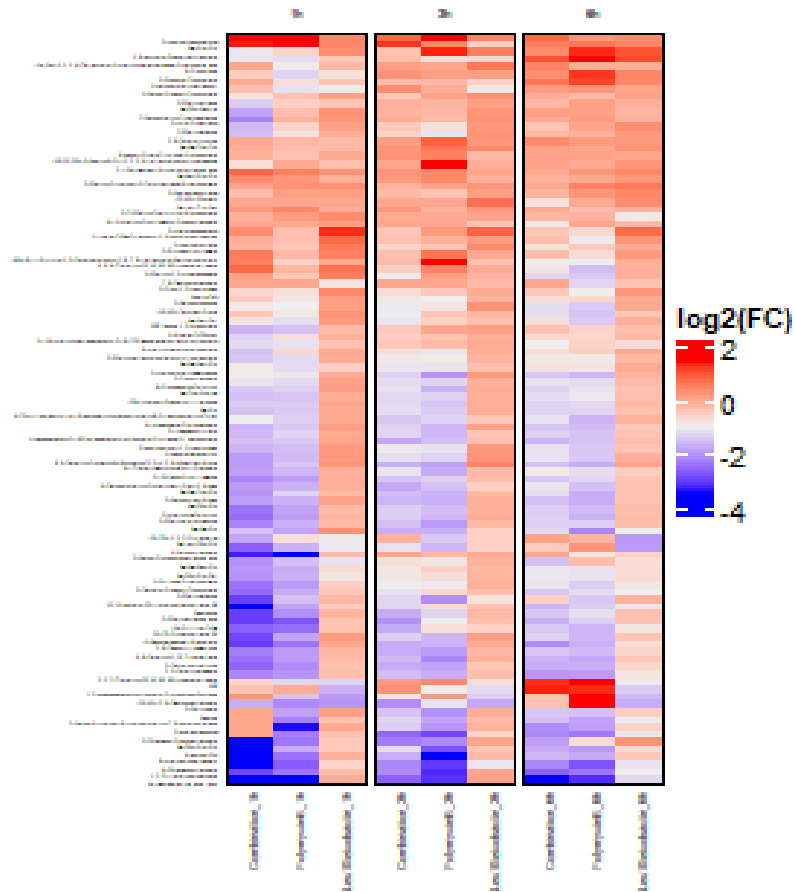


Figura 2: Heatmap log2(FC) de subset de datos, agrupados por condiciones y tiempo

Siguiendo la misma lógica, se reordenaron los datos para visualizar el cambio de los metabolitos, esta vez mediante gráfico de barras. En el siguiente caso, solo se compararon algunos metabolitos en tiempo 1. Sin embargo, se alcanza a vislumbrar

el cambio de la detección para la mayoría de ellos.

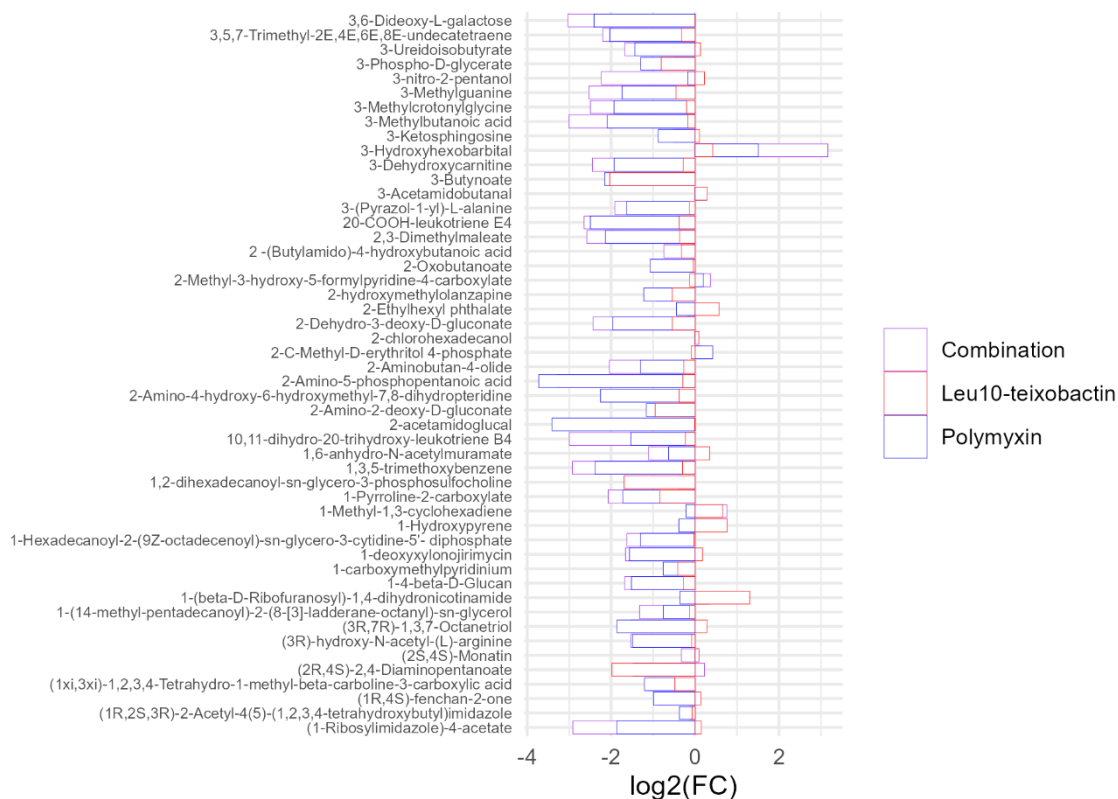


Figura 3: Comparativa de cambios $\log_2(FC)$ para distintos metabolitos. En azul el compuesto Polymixin B, en rojo Leu10-teixobactin y en morado la combinación de compuestos.

4. Conclusión

Se logró obtener los datos metabolómicos correspondientes al estudio seleccionado, donde se pudo almacenar e inspeccionar mediante el objeto SummarizedExperiment. Se exploraron sus principales características y con la observación de datos se logró inferir como trabajar con ellos. Se logró observar cambios en la medición de metabolitos en distintas condiciones, sin embargo, falta un trabajo más detallado para: Identificar cuales son los compuestos más afectados por la combinación de compuestos y relacionarlos con las rutas metabólicas, de tal manera de aportar información respecto a la hipótesis del trabajo.