

Informe del Estudio Metabolómico sobre el Síndrome de la Fatiga Crónica

Alvar Manjón

Tabla de contenidos

	Página:
Resumen	1
Objetivos	1
Metodología	2
Resultados	3
Discusión	6
Conclusiones	9
Referencias	9

Resumen

La encefalomiелitis miálgica/síndrome de fatiga crónica (ME/CFS) es una enfermedad aún poco conocida, caracterizada por fatiga extrema, entre otros síntomas. Aún no existen diagnósticos de laboratorio ni se conoce la causa, por lo que los estudios metabolómicos de esta enfermedad pueden arrojar luz. En este informe se ha realizado un estudio multivariante de una base de datos metabolómica compuesta por muestras de líquido cerebro espinal de enfermos de ME/CFS y controles. Se han realizado análisis de componentes principales y agrupamientos jerárquicos con mapas de calor. Los resultados muestran que no existe una gran diferenciación entre enfermos y controles en cuanto al metaboloma por lo que no podría usarse con estos análisis para diagnosticar la enfermedad. Sin embargo, en las muestras de los enfermos sí se detecta una menor concentración de metabolitos involucrados en rutas metabólicas de obtención de energía, lo cual podría estar relacionado con los síntomas de la enfermedad.

Objetivos

Los objetivos de este informe son los siguientes:

- Aprender a manejar datos ómicos mediante técnicas bioinformáticas siguiendo el software R.
- Manejar un objeto de la clase SummarizedExperiment.
- Preprocesar una base de datos preparándola para su posterior análisis estadístico.

- Aplicar un análisis multivariante a los datos del estudio metabolómico.
- Extraer conclusiones biológicas a partir de los resultados estadísticos.
- Concretamente para el estudio elegido, comprobar si el cuadro metabolómico para los enfermos de encefalomiелitis miálgica/síndrome de fatiga crónica puede diferenciarse estadísticamente de otras enfermedades neurológicas o de individuos sanos, usando análisis de componentes principales y agrupamientos jerárquicos.

Métodos

La obtención de una base de datos sobre el metaboloma se ha llevado a cabo a través del repositorio online Metabolomics Workbench (Sud et al., 2016). Allí, se han realizado búsquedas de estudios científicos que reuniesen un número considerable de muestras sobre las que se han medido las cantidades de una variedad de metabolitos. Finalmente, el trabajo de Nagy-Szakal et al. (2018) ha sido el seleccionado para este informe. Este estudio trata sobre la encefalomiелitis miálgica/síndrome de fatiga crónica (ME/CFS), la cual es una enfermedad neurológica que afecta a varios sistemas y se alarga más de 6 meses en el tiempo. El cuadro médico se caracteriza por fatiga inexplicable, malestar tras realizar actividades físicas, sueño insuficiente y disfunción cognitiva, entre otros síntomas. Por el momento no existe cura o tratamiento claro para esta enfermedad, la cual sufren entre 800.000 y 2,5 millones de personas en Estados Unidos. Además, no hay técnicas de diagnóstico en laboratorio aprobadas, tan sólo las basadas en historial y examinación física. De hecho, no se conoce el origen de la enfermedad, aunque existen hipótesis sobre su causa, como un origen genético, infección, desbalance inmunitario, alteración de la microbiota o estrés emocional. Por ello, estos estudios metabolómicos son de gran importancia ya que pueden ayudar a caracterizar la enfermedad, identificar posibles desbalances metabólicos causantes de la enfermedad y diagnosticarla.

Este estudio se ha seleccionado para este informe debido a la importancia que suponen los trabajos sobre esta enfermedad aún desconocida en la actualidad y que padece tanta gente. Además, la base de datos seleccionada cuenta con un buen número de muestras y la cantidad de metabolitos medidos es considerable. Todo ello permite un análisis multivariante más robusto. Dicha base de datos consiste en muestras de líquido cerebro espinal de hasta 91 sujetos humanos, 32 de los cuales diagnosticados con ME/CFS. El resto son muestras control que consisten en 40 procedentes de pacientes con esclerosis múltiple (MS) y 19 individuos sin enfermedad (ND). Las muestras se obtuvieron mediante punciones lumbares y se guardaban en los bio repositorios de Sierra Internal Medicine (SIM). Las muestras fueron preparadas y posteriormente sometidas a análisis metabolómicos no dirigidos para metabolitos primarios, aminas biogénicas y lípidos complejos. Estos análisis se realizaron mediante cromatografía de gases con tiempo de vuelo y mediante espectrometría de masas en tándem con cromatografía líquida. Esto dio lugar a una tabla de datos de las cantidades de 94 metabolitos.

Para este informe, los datos fueron tratados estadísticamente usando el software estadístico Rstudio (R Core Team, 2025). La mayoría de acciones dentro del software se hicieron usando la librería centrada en bioinformática Bioconductor (Huber et al., 2015). En primer lugar, se descargó la base de datos del estudio seleccionado desde el repositorio Metabolomics Workbench usando el paquete metabolomicsWorkbenchR (Lloyd & Weber,

2024) e introduciendo el ID del estudio (ST000910). De esta manera, se obtienen directamente los datos en formato de objeto de clase SummarizedExperiment. Otra opción para hacerlo sería descargar el archivo .csv del repositorio para construir el objeto de forma manual. Sin embargo, fue imposible obtener dicho archivo desde la página web del repositorio. Posteriormente, los datos fueron preprocesados siguiendo el *Workflow* del paquete de bioconductor POMA (Castellano-Escuder et al., 2021). Mediante dicho paquete se eliminaron los posibles valores nulos, se normalizaron los resultados y se buscaron los posibles valores extremos (*outliers*) para después proceder a su eliminación.

Con los datos ya preparados para su estudio estadístico se realizaron dos tipos de análisis. El primero es un análisis de componentes principales (PCA), para el cual se creó una función propia con el fin de representar los datos en un diagrama de dispersión para los componentes principales (PC) seleccionados, diferenciando en el mismo las muestras por su sexo y por su diagnóstico (ME/CFS, MS o ND) e incluyendo el porcentaje de varianza que explica cada PC. Además, se añadió un diagrama de cargas con el paquete factextra (Kassambara & Mundt, 2020), el cual representa la contribución de cada variable (metabolitos) dentro de los dos componentes mostrados en el diagrama de dispersión. El segundo análisis consiste en un agrupamiento jerárquico acompañado de un mapa de calor de la interacción entre los metabolitos y las muestras, usando la plantilla del paquete ComplexHeatmap (Gu et al., 2016; Gu, 2022). Finalmente se interpretaron los resultados obtenidos.

Resultados

El objeto SummarizedExperiment fue creado de forma exitosa. Se trata de una extensión más moderna de la clase ExpressionSet, además de ser más flexible. Algunas ventajas que posee respecto a ExpressionSet es la posibilidad de incluir varios ensayos en el mismo objeto, capacidad de incluir coordenadas genómicas (muy útil para estudios de secuenciación) y mejor adaptación para estudios multiómicos.

Durante el preprocesado de los datos con el *workflow* del paquete POMA podemos observar la distribución de los datos normalizados en un diagrama de cajas para las variables del estudio (**Fig. 1**). En cuanto a la eliminación de los valores nulos, no se ha detectado ninguno en la base de datos. Por su parte, el diagnóstico de los valores extremos muestra que sólo hay una muestra que sería necesario eliminar debido a esos datos anómalos, la cual pertenece al grupo de muestras control de individuos sin enfermedad diagnosticada (**Fig. 2**).

El PCA realizado a los datos preprocesados muestra que los 4 primeros PC no son capaces de explicar demasiada varianza dentro de los datos. Concretamente el PC1 explica el 11,6%, el PC2 el 9,3%, el PC3 el 6,3% y el PC4 el 5,8% de varianza, sumando entre todos un porcentaje acumulado del 33% de la varianza. Los diagramas de dispersión de las muestras para dichos componentes principales (**Fig. 3 y 4**) muestran una nube de puntos bastante difusa. Sin grupos completamente separados. Además, los tres posibles diagnósticos de las muestras (ME/CFS) aparecen mezclados. Lo mismo ocurre con el sexo, apareciendo hombres y mujeres mezclados. Sin embargo, sí que es posible distinguir que, dentro del diagrama de dispersión entre el PC 3 y el 4 (**Fig. 4**), las muestras con ME/CFS

están ligeramente desplazadas hacia la parte superior izquierda del gráfico, teniendo mayores valores para el PC4 y menores para el PC3 respecto a otras muestras control. En el correspondiente diagrama de cargas vemos que los metabolitos que contribuyen positivamente en esa dirección son pocos y tienen poca influencia, destacando únicamente ME230651, ME230662 y ME230671, que corresponden a la hidroxilamina, al ácido glicérico y a la creatinina, respectivamente. Por su parte, los metabolitos que tienen una gran influencia en la dirección contraria son ME230624, ME230641, ME230637, ME230673 y ME230609, los cuales corresponden a la fosfoetanolamina, al ácido lignocérico, la manosa, el ácido cítrico y la treonina.

En cuanto al mapa de calor junto al agrupamiento jerárquico (**Fig. 5**) se observa que tampoco hay un agrupamiento claro de las muestras en función de su diagnóstico en base a todo el metaboloma estudiado. Se aprecian las muestras mezcladas sin seguir un patrón claro de ordenamiento. Además, dentro del propio mapa de calor tampoco parecen haber patrones claros. Tan sólo se aprecian un par de parches de color rojo o de color azul. El que más destaca es el parche rojo que incluye muestras control y los metabolitos ME230624, ME230641 y ME230613, correspondientes a la fosfoetanolamina, al ácido lignocérico y al ácido succínico, respectivamente. Además, por lo general parece que hay una mayor influencia positiva de estos metabolitos en las muestras y negativa en los pacientes con ME/CFS. Los otros dos parches azules que parecen claros en el gráfico pertenecen a los mismos metabolitos pero están asociados tanto a muestras control como a muestras diagnosticadas con ME/CFS. Finalmente, de los metabolitos que tenían una influencia positiva en la dirección de las muestras diagnosticadas con ME/CFS en el PCA, la creatinina y el ácido glicérico parece que muestran un patrón en el que el color rojo coincide generalmente con muestras con ME/CFS y el azul con muestras control, es decir, influyen positivamente en las muestras con la enfermedad y negativamente en muestras control.

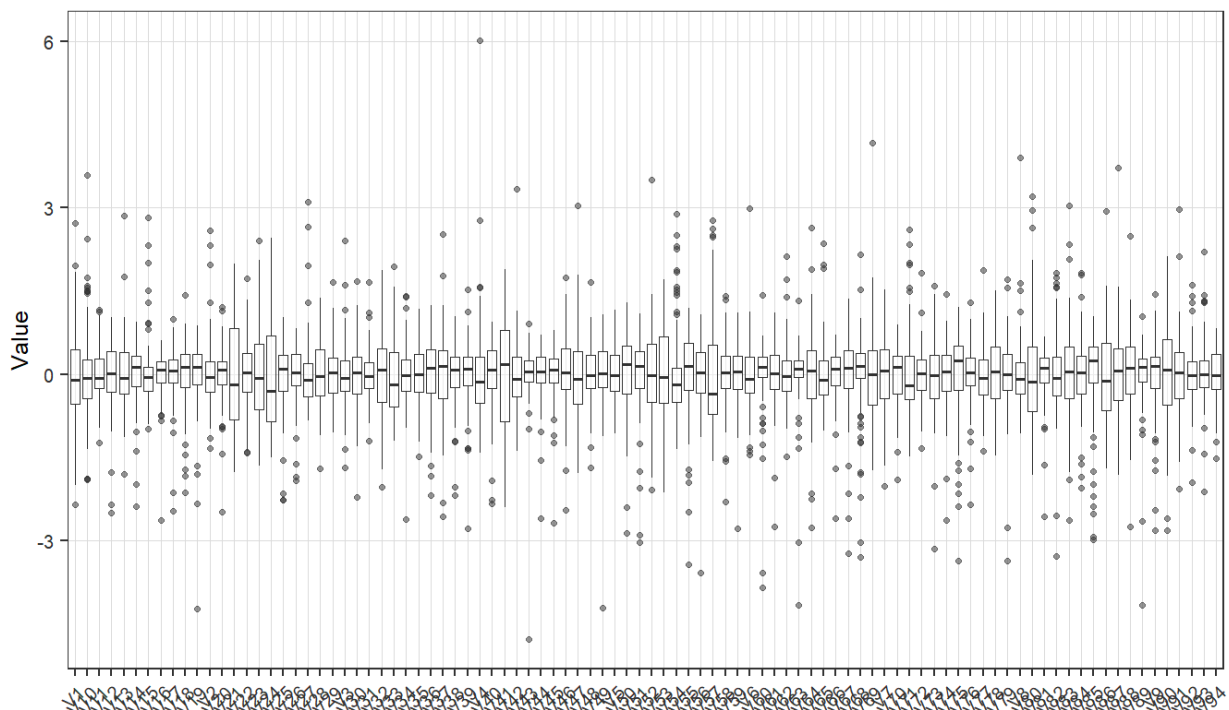


Fig. 1. Diagrama de cajas de los resultados para cada una de las variables en el estudio del metaboloma. Los puntos representan los posibles valores extremos.

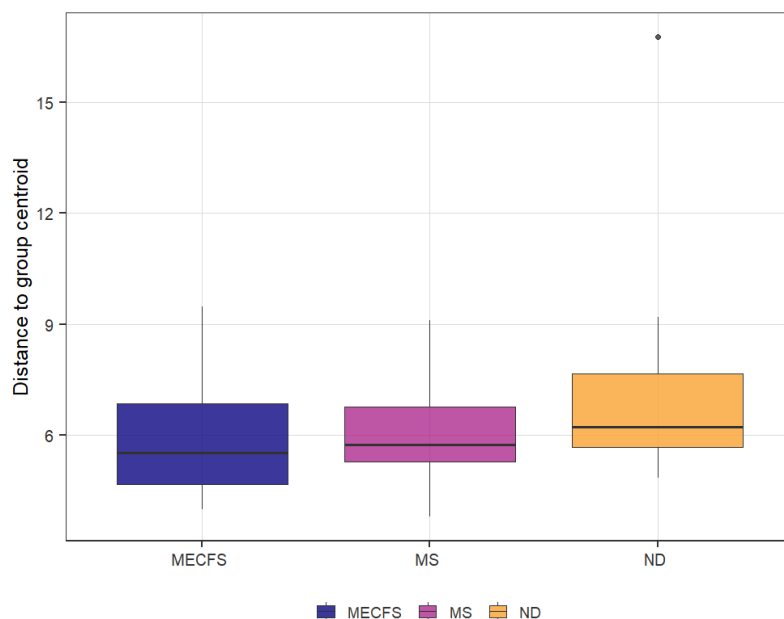


Fig. 2. Diagrama de cajas para las muestras en el estudio de los valores extremos mediante el método del centroide. Los puntos representan los valores extremos demasiado lejos del centroide de su propio grupo. Las muestras se agrupan según el diagnóstico de enfermedad mental (MECFS = encefalomiелitis miálgica/síndrome de fatiga crónica; MS = esclerosis múltiple; ND = sin enfermedad neurológica).

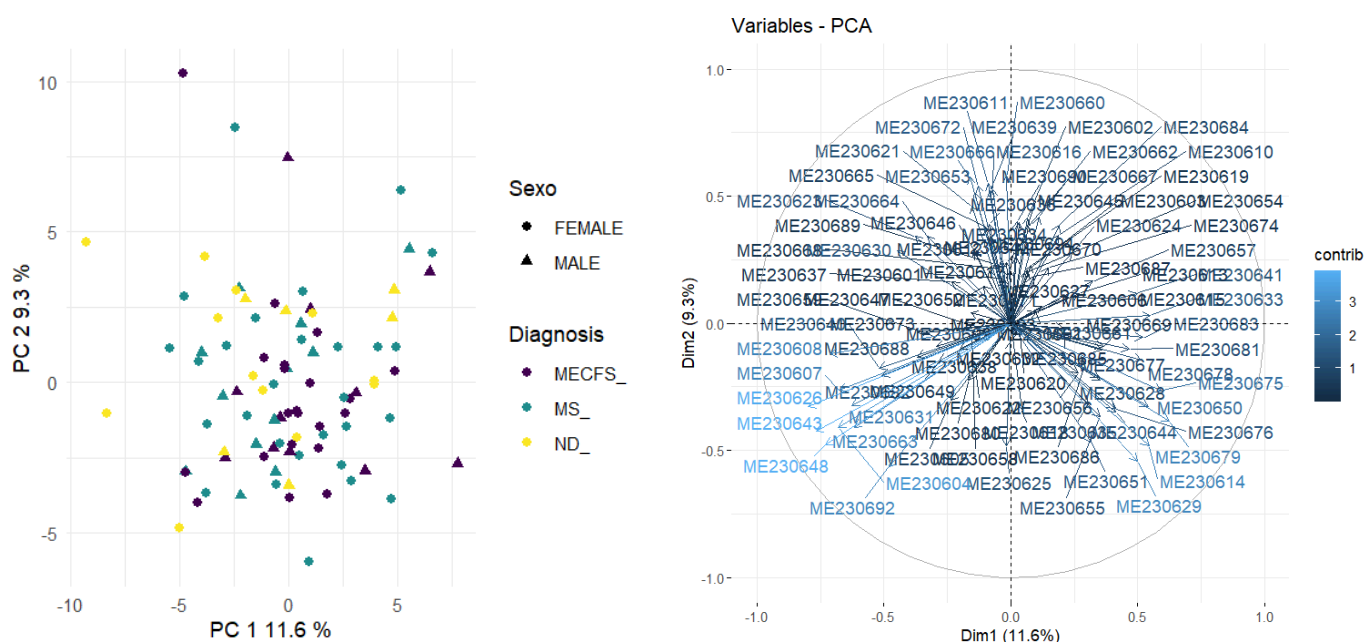


Fig. 3. Análisis de componentes principales para los resultados metabólicos. En la izquierda, diagrama de dispersión de las muestras para los componentes principales 1 y 2. La forma de cada punto depende del factor sexo de cada muestra y el color del diagnóstico de enfermedad neuronal (MECFS = encefalomiелitis miálgica/síndrome de fatiga crónica; MS = esclerosis múltiple; ND = sin enfermedad neurológica). En la derecha, el diagrama de cargas de cada una de las variables (metabolitos nombrados con sus ID de WorkBench)

para los componentes principales 1 y 2. El color depende de la cantidad de carga. En los ejes se muestra el porcentaje de varianza que explica cada componente/dimensión.

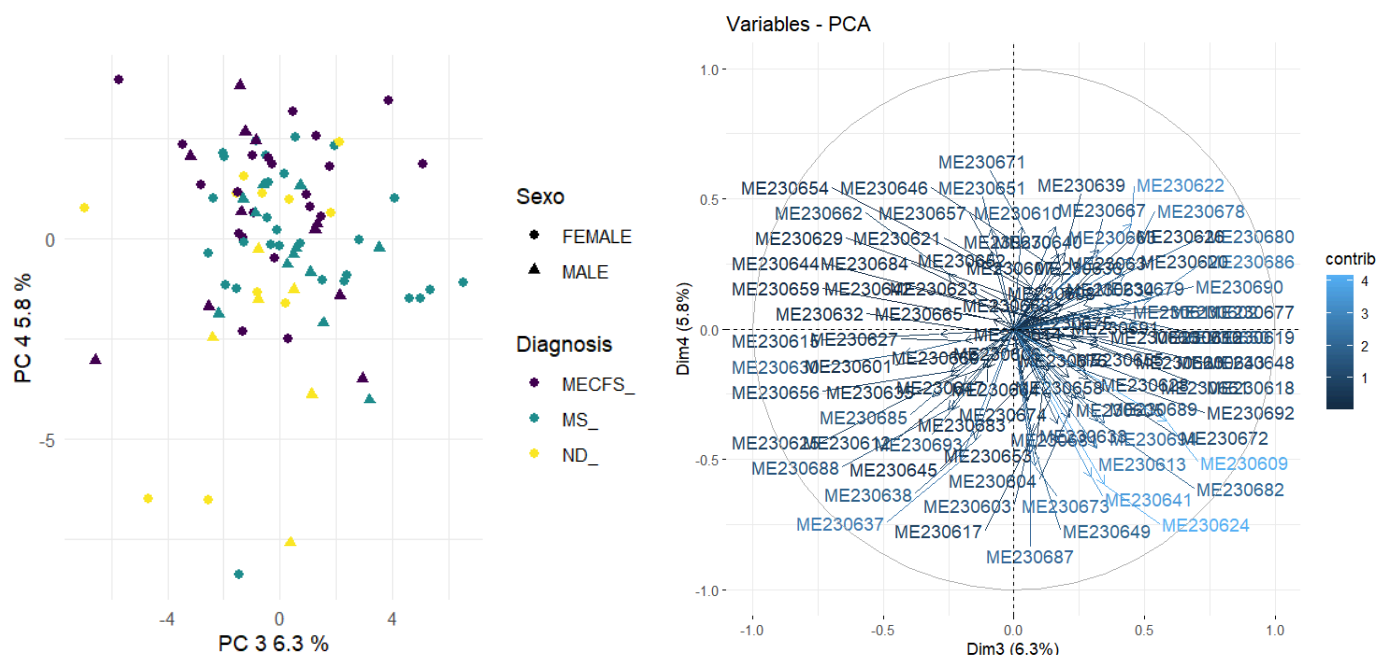


Fig. 4. Análisis de componentes principales para los resultados metabólicos. En la izquierda, diagrama de dispersión de las muestras para los componentes principales 3 y 4. La forma de cada punto depende del factor sexo de cada muestra y el color del diagnóstico de enfermedad neuronal (MECFS = encefalomielitis miálgica/síndrome de fatiga crónica; MS = esclerosis múltiple; ND = sin enfermedad neurológica). En la derecha, el diagrama de cargas de cada una de las variables (metabolitos nombrados con sus ID de WorkBench) para los componentes principales 3 y 4. El color depende de la cantidad de carga. En los ejes se muestra el porcentaje de varianza que explica cada componente/dimensión.

Discusión

Los resultados del estudio en cuestión parece que han sido tomados de forma muy precisa ya que durante el preprocesamiento tan sólo ha habido que eliminar una muestra del total de 91.

En cuanto a los resultados estadísticos obtenidos tras el análisis multivariante se observa que, por lo general, no hay patrones claros en el metaboloma. Las muestras no se agrupan únicamente entre sí dentro del PCA o el agrupamiento jerárquico, apareciendo mezcladas con los grupos control. Esto sugiere que el uso del metaboloma procedente del líquido cerebro espinal de los pacientes para diagnosticar la ME/CFS no es óptimo. A pesar de detectarse algunos patrones como el observado en diagrama de dispersión para el PCA con los PC 3 y 4, los datos se solapan de tal manera que un diagnóstico de la enfermedad llevaría a numerosos falsos positivos provocados por la variabilidad intrapoblacional. Lo mismo ocurriría con un mapa de calor, donde también se aprecia esa variabilidad.

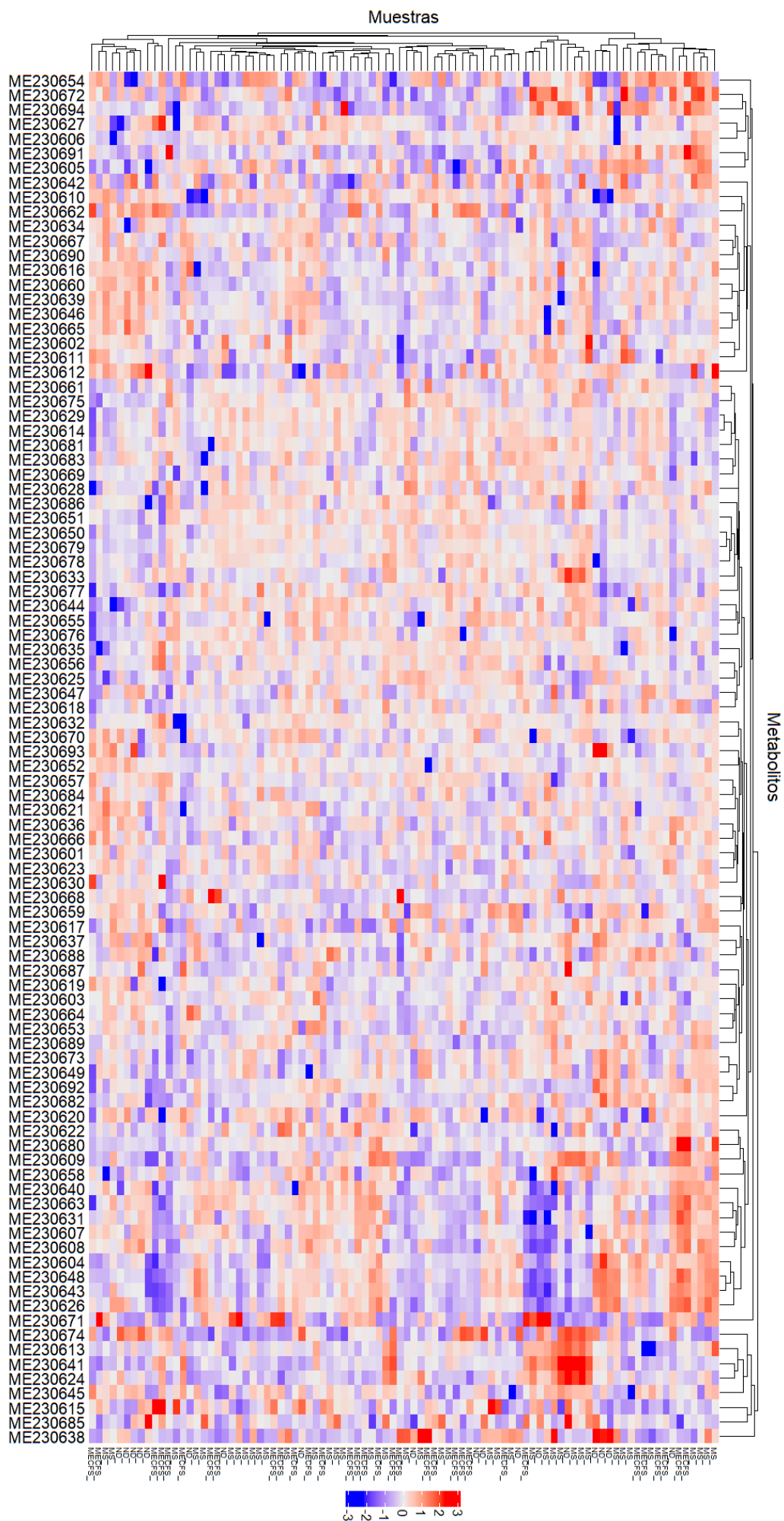


Fig. 5. Mapa de calor y dendrogramas jerárquicos generados con los resultados del metaboloma. Uno de los ejes representa las muestras, nombradas en función del factor de diagnóstico de enfermedad neuronal (ME/CFS = encefalomiелitis miálgica/síndrome de fatiga crónica; MS = esclerosis múltiple; ND = sin enfermedad neurológica), y el otro los metabolitos medidos, nombrados con sus ID de WorkBench. Los colores van de azul a rojo en función del efecto de los metabolitos sobre las muestras, desde negativo a positivo respectivamente.

Sin embargo, los pocos patrones que se han detectado son de gran interés ya que, a pesar de no poder diagnosticar la enfermedad, se pueden obtener pistas sobre su origen o causas. La influencia positiva o negativa de los metabolitos sobre las muestras con ME/CFS en los análisis sugiere que pueden estar relacionados con la enfermedad, encontrándose, respectivamente, en altas o bajas concentraciones dentro del metaboloma de los enfermos. Teniendo en cuenta todos los análisis, los metabolitos relacionados positivamente con la ME/CFS son la creatinina, la hidroxilamina y el ácido glicérico. Los relacionados negativamente con la enfermedad son la fosfoetanolamina, el ácido lignocérico, la manosa, el ácido cítrico, el ácido succínico y la treonina. La implicación metabólica de muchos de los compuestos encontrados en baja concentración en el metaboloma de los enfermos están relacionadas con la obtención de energía en el cuerpo humano. Por ejemplo, la manosa es un monosacárido relacionado con la obtención rápida de energía, el ácido cítrico es uno de los intermediarios en el ciclo de ácidos tricarboxílicos, la treonina participa en el metabolismo de las grasas, el ácido lignocérico es un ácido graso y la fosfoetanolamina participa en el metabolismo mitocondrial. Por todo ello, la baja concentración de estos compuestos en el metaboloma de los enfermos respecto a los grupos control puede estar estrechamente relacionado con los síntomas de fatiga extrema o disfunción cognitiva, reflejando la incapacidad de los afectados para metabolizar la energía de forma tan eficiente. De hecho, en otros estudios ya se han relacionado anomalías metabólicas con la ME/CFS relacionadas con el ciclo de ácidos tricarboxílicos, el metabolismo de azúcares o el metabolismo de grasas (Nagy-Szakal et al., 2018; Yamano et al., 2021).

Por su parte, entre los metabolitos encontrados en mayor concentración en los diagnosticados con la enfermedad destaca la creatinina, un compuesto de desecho derivado de la creatina que se produce con la degradación muscular a causa del uso del propio músculo (Kashani et al., 2020). Esta mayor concentración de creatinina también se ha encontrado en otros estudios (Yamano et al., 2021). Sin embargo, en algunos trabajos se ha detectado una menor concentración en sangre de creatinina en los enfermos de ME/CFS (Nacul et al., 2019). De hecho, la creatina ha sido sugerida como un tratamiento para los enfermos (Godlewska et al., 2024). Esto puede estar explicado, en nuestro caso, por el uso muestras de líquido cerebro espinal o por la existencia de diferentes cuadros clínicos para esta enfermedad.

Como se puede comprobar, los resultados de este estudio no son del todo concluyentes por lo que es necesario seguir investigando sobre esta enfermedad. En concreto, otros estudios metabolómicos con muestras procedentes de otros tejidos podrían complementar los trabajos ya publicados y ofrecer una visión global del metabolismo de enfermos de ME/CFS.

Conclusiones

Este informe concluye que:

- Las técnicas bioinformáticas son necesarias para estudiar los datos ómicos.
- Los objetos de clase SummarizedExperiment son muy útiles para bases de datos ómicas como la usada en el informe.
- El preprocesamiento de los datos metabolómicos ha permitido hacer un posterior análisis más exhaustivo.
- Los análisis multivariantes de PCA y agrupamiento jerárquico sobre los datos metabolómicos de muestras de líquido cerebro espinal no parecen permitir diagnosticar la encefalomielitis miálgica/síndrome de fatiga crónica.
- Los análisis multivariantes muestran que los enfermos de encefalomielitis miálgica/síndrome de fatiga crónica poseen menores niveles de metabolitos relacionados con el metabolismo de la energía como el ciclo de Krebs, el metabolismo de los azúcares o el metabolismo de las grasas.
- Estos resultados tienen limitaciones al solo haber desarrollado unos pocos análisis multivariantes y no haber tenido en cuenta los metabolomas de otros tejidos.
- Es necesario continuar con los estudios sobre esta enfermedad para arrojar luz sobre sus causas y poder desarrollar curas, tratamientos y sistemas de diagnóstico.

Referencias

1. Castellano-Escuder, P., González-Domínguez, R., Carmona-Pontaque, F., Andrés-Lacueva, C., Sánchez-Pla, A. (2021). "POMAShiny: A user-friendly web-based workflow for metabolomics and proteomics data analysis." *PLOS Computational Biology*, 17(7), 1-15. <https://doi.org/10.1371/journal.pcbi.1009148>
2. Godlewska, B. R., Sylvester, A. L., Emir, U. E., Sharpley, A. L., Clarke, W. T., Martens, M. A. G., & Cowen, P. J. (2024). Six-Week Supplementation with Creatine in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS): A Magnetic Resonance Spectroscopy Feasibility Study at 3 Tesla. *Nutrients*, 16(19), 3308. <https://doi.org/10.3390/nu16193308>
3. Gu, Z., Eils, R., & Schlesner, M. (2016). "Complex heatmaps reveal patterns and correlations in multidimensional genomic data." *Bioinformatics*. doi:10.1093/bioinformatics/btw313.
4. Gu, Z. (2022). "Complex Heatmap Visualization." *iMeta*. doi:10.1002/imt2.43.
5. Huber, W., Carey, V. J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B. S., Bravo, H. C., Davis, S., Gatto, L., Girke, T., Gottardo, R., Hahne, F., Hansen, K. D., Irizarry, R. A., Lawrence, M., Love, M. I., MacDonald, J., Obenchain, V., Oleś, A. K., Pagès, H., Reyes, A., Shannon, P., Smyth, G. K., Tenenbaum, D., Waldron, L., & Morgan, M. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. *Nature Methods*, 12(2), 115-121. <https://doi.org/10.1038/nmeth.3252>
6. Kashani, K., Rosner, M. H., & Ostermann, M. (2020). Creatinine: From physiology to clinical application. *European Journal of Internal Medicine*, 72, 9-14. <https://doi.org/10.1016/j.ejim.2019.10.025>
7. Kassambara, A., & Mundt, F. (2020). factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R package version 1.0.7, <https://CRAN.R-project.org/package=factoextra>
8. Lloyd, G. R., & Weber, R. J. M. (2024). *metabolomicsWorkbenchR: Metabolomics Workbench in R*. R package version 1.16.0.

9. Nacul, L., De Barros, B., Kingdon, C. C., Cliff, J. M., Clark, T. G., Mudie, K., Dockrell, H. M., & Lacerda, E. M. (2019). Evidence of Clinical Pathology Abnormalities in People with Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) from an Analytic Cross-Sectional Study. *Diagnostics*, 9(2), 41. <https://doi.org/10.3390/diagnostics9020041>
10. Nagy-Szakal, D., Barupal, D. K., Lee, B., Che, X., Williams, B. L., Kahn, E. J. R., Ukaigwe, J. E., Bateman, L., Klimas, N. G., Komaroff, A. L., Levine, S., Montoya, J. G., Peterson, D. L., Levin, B., Hornig, M., Fiehn, O., & Lipkin, W. I. (2018). Insights into myalgic encephalomyelitis/chronic fatigue syndrome phenotypes through comprehensive metabolomics. *Scientific Reports*, 8(1), 10056. <https://doi.org/10.1038/s41598-018-28477-9>
11. R Core Team (2025). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
12. Sud, M., Fahy, E., Cotter, D., Azam, K., Vadivelu, I., Burant, C., Edison, A., Fiehn, O., Higashi, R., Nair, K. S., Sumner, S., & Subramaniam, S. (2016). Metabolomics Workbench: An international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. *Nucleic Acids Research*, 44(D1), D463-D470. <https://doi.org/10.1093/nar/gkv1042>
13. Yamano, E., Watanabe, Y., & Kataoka, Y. (2021). Insights into Metabolite Diagnostic Biomarkers for Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *International Journal of Molecular Sciences*, 22(7), 3423. <https://doi.org/10.3390/ijms22073423>

Enlace de GitHub: <https://github.com/AlvarManjon/Manjon-Garcia-Alvar-PEC1.git>