

Jácome-Navarrete-Juan-PEC1

Juan C. Jácome Navarrete

2025-04-02

Table of contents

1	Resumen	1
2	Objetivo	2
3	Métodos	2
3.1	Origen de los datos	2
3.2	Estructura de los datos	3
3.3	Procedimiento y Herramientas de análisis	3
4	Resultados	4
4.1	Construcción del objeto <i>SummarizedExperiment</i>	4
4.2	Análisis de abundancia de metabolitos	6
4.3	Análisis multivariante	7
5	Discusión	9
6	Conclusiones	10
7	Anexos	11
7.1	ANEXO 1	11
7.2	ANEXO 2	11
	Referencias	11

1 Resumen

Se analizó el perfil metabolómico de 12 muestras de leche humana, correspondientes a tres fracciones (entera, lipídica y acuosa) de cuatro madres, utilizando datos del estudio ST000957.

Se construyó un objeto `SummarizedExperiment` para proceder con la exploración de los datos. Las fracciones acuosa y entera presentaron mayor cantidad y concentración de metabolitos que la lipídica, especialmente compuestos hidrosolubles como phosphocholine y L-carnitine. Al agrupar por clase química, destacaron aminoácidos, péptidos y colinas. El análisis multivariante reveló que las diferencias entre madres fueron más marcadas que entre fracciones. Además, algunos metabolitos (como ácidos biliares conjugados) se concentraron en fracciones lipídicas, reflejando su afinidad hidrofóbica. Las limitaciones incluyeron el uso exclusivo del modo positivo, lo que pudo evitar la detección de otros metabolitos como carbohidratos.

2 Objetivo

El presente trabajo tiene como objetivo principal analizar las diferencias y patrones del perfil metabolómico de la leche humana recolectada en distintas fracciones (entera, lipídica y acuosa) y de diferentes madres.

3 Métodos

3.1 Origen de los datos

Para el análisis se seleccionó el dataset **ST000957 (Carney, n.d.)**, titulado “*Global metabolomics of human milk fractions*”, disponible en Metabolomics Workbench. Se trata de un estudio transversal con 12 muestras biológicas de leche humana, que analiza el perfil metabólico en tres fracciones de la leche: lipídica (fat), acuosa (skim) y entera (whole).

Este estudio fue realizado por la University of Florida utilizando espectrometría de masas acoplada a cromatografía líquida (LC-MS). Para simplificar el análisis solo se centró exclusivamente en metabolitos detectados bajo modo de ionización positiva.

Las muestras fueron recolectadas en estado de ayuno, aproximadamente 30 mL de leche y posterior fraccionamiento por centrifugación. No se especificó el estadio de lactancia al momento de la recolección.

Table 1: Resumen del estudio ST000957

Campo	Valor
study_id	ST000957
study_title	Global metabolomics of human milk fractions
species	Homo sapiens
institute	University of Florida
analysis_type	LC-MS
number_of_samples	12

3.2 Estructura de los datos

Los datos fueron obtenidos de tres archivos diferentes con los que se creó objeto **SummarizedExperiment**. Los archivos fueron:

1. Una matriz de abundancias (**assay**): esta matriz contiene en las columnas las muestras y las filas los metabolitos. Cada celda o combinación de fila y columna es un número entero que representa la cuantificación de metabolitos en cada muestra.
2. Una tabla de metadatos de muestra (**colData**): contiene información sobre las muestras y debe coincidir con las columnas de matriz del ensayo. Las variables de las muestras son:
 - **sujeto**: identifica a la madre donante (BLS001A, BLS002A, etc.).
 - **Milk_fraction**: indica si la muestra corresponde a la fracción lipídica, acuosa o entera.
3. Una tabla de anotaciones de metabolitos (**rowData**): contiene información sobre los metabolitos. En este caso se trata de la clasificación de los metabolitos según el la base de datos *Metabolics Workbench*. Debe coincidir con las filas de matriz del ensayo.
 - **Standardized_name**: nombre estandarizado según *Metabolomics Workbench*
 - **Formula**: fórmula química.
 - **Super_class**, **Main_class**, **Sub_class**: Diferentes niveles de clasificación química del metabolito.

3.3 Procedimiento y Herramientas de análisis

El proceso de análisis fue realizado en su totalidad en R (versión 4.4.2), y el presente informe fue generado con Quarto.

En primero lugar se busco los datos del proyecto con el paquete **metabolomicsWorkbenchR** utilizando la función **do_query()** para acceder al estudio ST000957. Luego se procedió a crear el objeto **SummarizedExperiment**, con el paquete del mismo nombre. Tras examinar su estructura, se identificó la presencia de datos faltantes, los cuales fueron imputados mediante el algoritmo KNN utilizando el paquete **POMA**. La normalización de los datos también se llevó a cabo con **POMA**, mediante la transformación log-pareto.

Desde este punto se realizó un análisis exploratorio. Para lo cual que usaron paquetes comunes de R y Bioconductor. Algunos de estos paquetes son **ggplot2**, junto con herramientas complementarias para mejorar la visualización y **patchwork**, que facilitó la organización y combinación de gráficos. Las tablas se generaron con los paquetes **knitr** y **kableExtra**.

Además se crearon tres funciones personalizadas:

- `calcular_abundancia_por_sujeto(objeto_se)`: para obtener las abundancias promedios de los metabolitos, llamada
- `seleccionar_top_y_otros(se, top_n = 20)`: para obtener un objeto `SummarizedExperiment` con los top-n metabolitos más abundantes.
- `agrupar_por_clase(se, clase)`: obtiene `SummarizedExperiment` agrupando los metabolitos según su clasificación química, sumando las abundancias.

En cuanto al análisis, en primer lugar se exploraron las intensidades absolutas de los metabolitos por fracción y por madre, tanto a nivel de metabolito individual como de clase funcional, empleando gráficos de barras apiladas. Esto permitió explorar el perfil metabólico de las muestras. Para analizar la abundancia de metabolitos individuales, se seleccionaron los 30 metabolitos con mayor intensidad o abundancia.

Posteriormente, tras la normalización, se procedió obtener el número de metabolitos detectados por fracción y madre (Anexo 1).

Siguiendo, se realizó heatmap con el paquete `pheatmap`. Para esto se calculó la varianza por metabolito, y se seleccionaron los 20 con mayor variabilidad, para comparar entre muestras. La función `pheatmap()` también me da la opción de realizar una clusterización entre variables de las muestras, que son la fracción de leche y la madre donadora.

- Si el metabolito se observa en rojo, tiene mayor abundancia en esa muestra (por encima del promedio);
- Si esta en azul, tiene menor abundancia (por debajo del promedio);
- Y si esta en blanco ese metabolito tiene un valor promedio en esa muestra.

Finalmente, con la función `prcomp()` se realizó un análisis de componentes principales (PCA) y se gráfico los primeros dos componentes en un diagrama de dispersión, para observar como se distribuyen las muestras en función de su perfil metabólico.

Los sriptis, el informe, datos y demas archivos del proyecto se pueden obtener en el repositorio de GitHub (Jácome, 2025): <https://github.com/juano3101/Jacome-Navarrete-Juan-PEC1>

4 Resultados

4.1 Construcción del objeto *SummarizedExperiment*

Lo primero se se realizó es el la construcción del *SummarizedExperiment*, para lo cual se importaron tres tablas necesarias: el ensayo, colData y el rowData. Además, se procedió a limpiar los datos, mantener solo las columnas deseadas, finalmente los valores faltantes fueron imputados con el método KNN, y no se eliminó ninguna muestra y ningún metabolito.

```
# CARGAR MATRICES DE DATOS
assay <- read.delim(params$assay, row.names = 1, check.names = FALSE)
coldata <- read.delim(params$coldata, row.names = 1, check.names = FALSE)
rowdata <- read.delim(params$rowdata, row.names = 1, check.names = FALSE)

# CREAR EL OBJETO SummarizedExperiment
se <- SummarizedExperiment(assays = list(count = assay),
  colData = coldata, rowData = rowdata)

# SE REALIZA UNA IMPUTACIÓN CON EL METODO KNN, SIN REMOVER MUESTRAS
se_imputed <- PomaImpute(se, method = "knn", zeros_as_na = TRUE, remove_na = FALSE)
```

```
class: SummarizedExperiment
dim: 158 12
metadata(0):
assays(1): count
rownames(158): 2-hydroxyglutarate+Na 2-Hydroxyphenylalanine ...
  Ursodeoxycholic Acid (UDCA)_R2 Xanthurenic Acid
rowData names(5): Standardized_name Formula Super_class Main_class
  Sub_class
colnames(12): BLS001A BLS010A ... BLS003A_3 BLS002A_3
colData names(2): Milk_fraction sujeto
```

El objeto `SummarizedExperiment` contiene un total de 158 metabolitos analizados en 12 muestras de leche humana. La matriz principal de datos, `count`, almacena las intensidades o abundancia de cada metabolito por muestra. Las columnas representan las muestras, etiquetadas como BLS001A, BLS010A, etc., y cuentan con metadatos asociados a través de `colData`, donde se especifica la fracción de leche (`Milk_fraction`) y la madre que dio la muestra (`sujeto`) del que proviene cada muestra.

Table 2: Distribución de muestras por madre y fracción de leche

Madre	Frecuencia_madre	Fracción	Frecuencia_fracción
BLS001A	3	fat	4
BLS002A	3	skim	4
BLS003A	3	whole	4
BLS010A	3		

La Tabla 2 muestra que cada uno de las cuatro madres (BLS001A, BLS002A, BLS003A y BLS010A) aportó exactamente tres muestras. Por su parte, la Tabla 3 que hay cuatro muestras por cada tipo de fracción: fat, skim y whole.

4.2 Análisis de abundancia de metabolitos

A continuación, se presentan diagramas de barras apiladas que muestran la distribución de los metabolitos individuales y su clasificación funcional en cada fracción y madre.

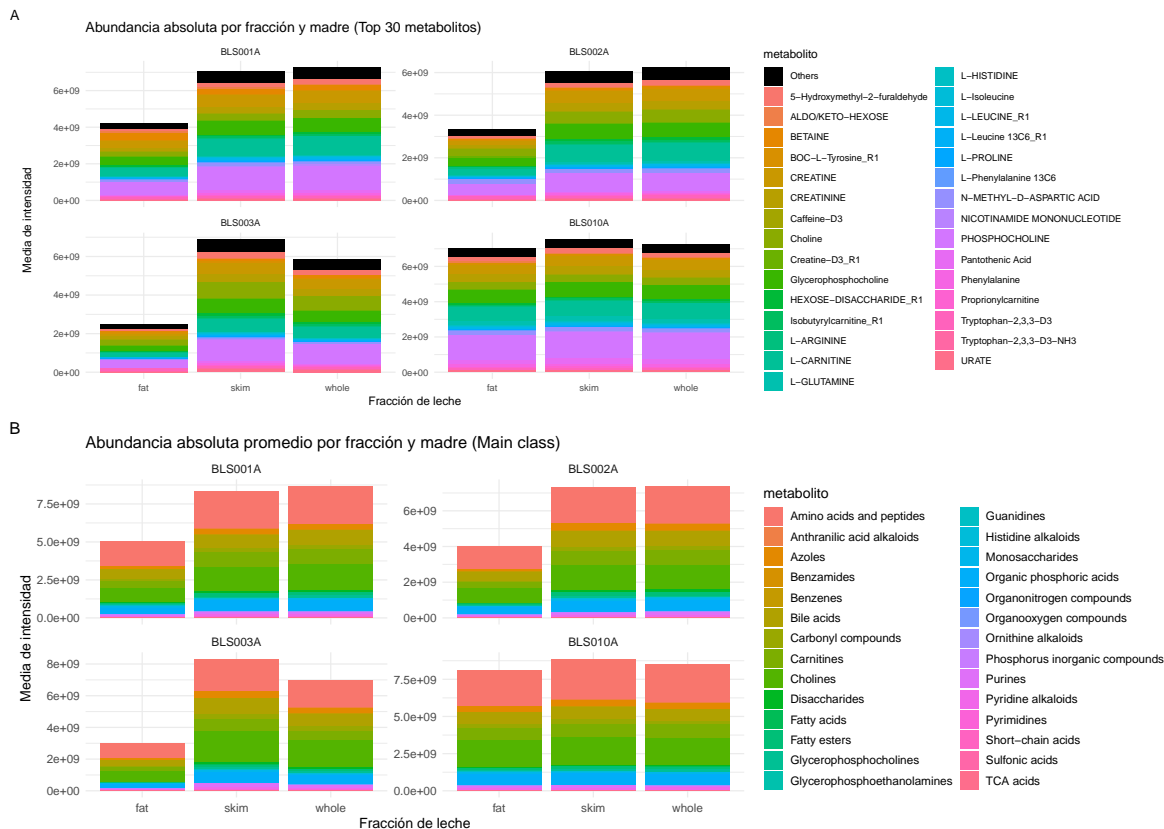


Figure 1: Abundancia de metabolitos por fracción y por sujeto. A: abundancia de metabolitos. B: abundancia por clasificación principal

La Figura 1 muestra la abundancia absoluta de metabolitos en leche humana, separada en tres fracciones: lipídica (fat), acuosa (skim) y entera (whole), provenientes de cuatro mujeres.

La Figura 1.A presenta los 30 metabolitos más abundantes. Las fracciones acuosa y entera tienden a tener mayores concentraciones de metabolitos que la fracción lipídica en todas las mujeres, lo que sugiere una mayor concentración de compuestos hidrosolubles en la leche materna. Entre los metabolitos individuales, phosphocholine y L-carnitine fueron los más abundantes en las tres fracciones. En la fracción lipídica, también se observaron altos niveles de glycerophosphocholine y creatine. Estos resultados evidencian similitudes en los metabolitos más abundantes entre individuos, aunque con diferencias en la magnitud de sus intensidades.

La Figura 1.B agrupa los metabolitos según su clasificación funcional principal (Main class), permitiendo una visualización más general de la composición química. Se destaca que los aminoácidos y péptidos, así como las colinas, son la clase más abundante en todas las fracciones y madres. Es importante mencionar que los Carbohidratos (mono y disacáridos) representaron entre el 1 y 2% de la composición de la leche en todas las muestras.

4.3 Análisis multivariante

Para poder tener más confianza en el análisis multivariante y reducir el efecto de variaciones debidos a la técnica, los datos fueron normalizados utilizando el método log-pareto con la función `PomaNorm()` del paquete POMA.

Una vez normalizado se observó que la mayoría de curvas de densidad están centradas en torno a cero, lo que indica que la normalización fue efectiva. También se obtuvo el número total de metabolitos detectados en cada fracción de leche, para las cuatro madres, utilizando los datos normalizados. Se ve claramente que las fracciones acuosa y entera presentan un mayor número de metabolitos en comparación con la fracción lipídica, lo cual concuerda con los patrones observados previamente en las gráficas de abundancia. (ANEXOS 1 y 2)

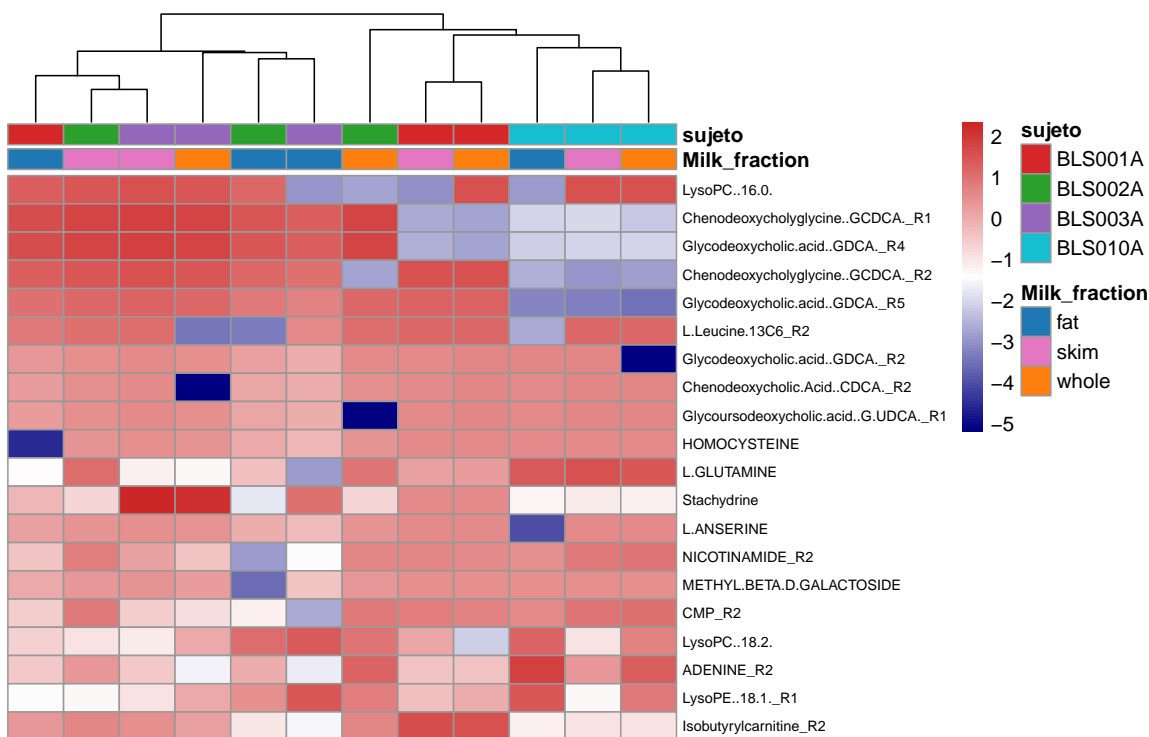


Figure 2: Heatmap de los 20 metabolitos más variables entre muestras

La Figura 2 muestra un heatmap de las intensidades normalizadas de los 20 metabolitos con mayor variabilidad entre las muestras. Se observan agrupamientos de muestras por madre, especialmente en BLS001A y BLS003A, así como diferencias claras entre fracciones de leche en metabolitos como los ácidos biliares conjugados (por ejemplo, Glycodeoxycholic.acid.GDCA_R2).

Además, compuestos como Stachydrine (un alcaloide natural), METHYL.BETA.D.GALACTOSIDE (un derivado de galactosa), péptidos como L-Anserine y lípidos como LysoPC y LysoPE presentan una alta variabilidad tanto entre madres como entre fracciones. Esto sugiere que el perfil metabolómico de la leche humana está influenciado tanto por características individuales como por el tipo de fracción analizada.

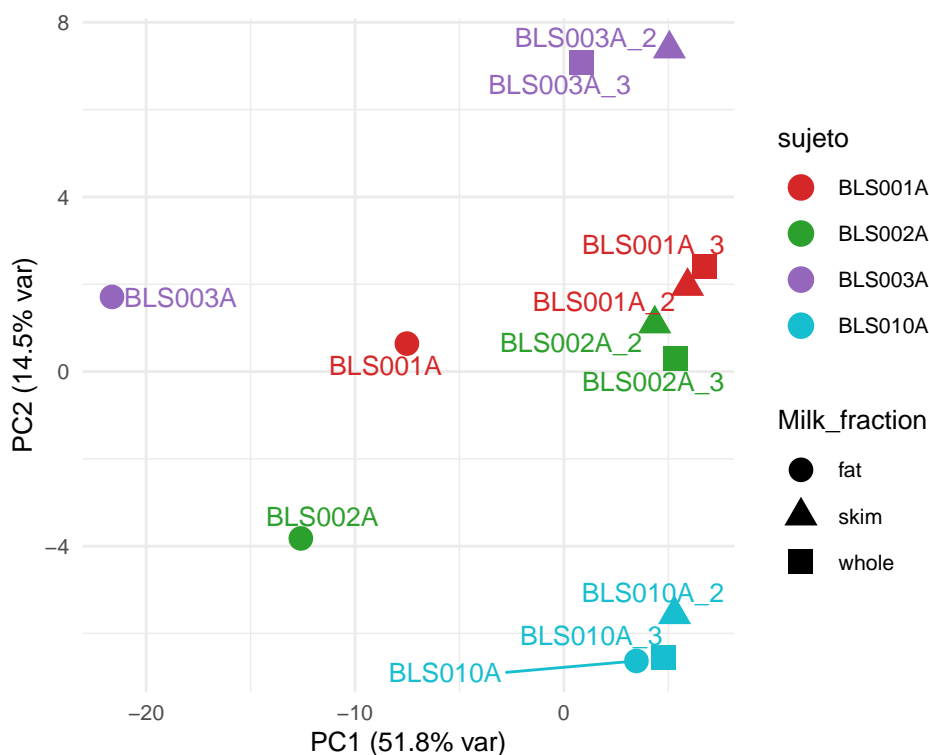


Figure 3: Análisis de componentes principales (PCA) de muestras de leche humana según metabolitos

La Figura 3 muestra un análisis de componentes principales (PCA) basado en las intensidades de metabolitos en muestras de leche humana. Los dos primeros componentes explican el 66.3% de la variabilidad total. Se observan agrupamientos por madre, con las muestras de cada madre ubicadas próximas entre sí. Por lo contrario, no se evidencia una agrupación por fracción de leche, aunque la fracción grasa (fat) parece separarse de las demás. Esto sugiere

que la composición metabolómica varía principalmente entre individuos, y en menor medida según la fracción de leche.

5 Discusión

El presente análisis permitió caracterizar el perfil metabolómico de muestras de leche humana en las fracciones lipídica, acuosa y entera, procedentes de cuatro madres distintas, utilizando LC-MS. Para este estudio se utilizaron exclusivamente los datos obtenidos en modo de ionización positiva.

Entre los metabolitos más abundantes se identificaron compuestos como phosphocholine y L-carnitine, presentes en todas las fracciones, en concordancia con otros estudios (Hampel et al., 2021; Torrez Lamberti et al., 2020). La phosphocholine es una forma abundante de colina en la leche materna, esencial para la síntesis de fosfolípidos y el desarrollo cerebral del lactante, particularmente en los primeros meses de vida. (Hampel et al., 2021; Poulsen et al., 2022). La carnitina, por otro lado, facilita el transporte de ácidos grasos hacia las mitocondrias para la producción de energía (Maheshwari, 2022; Poulsen et al., 2022).

Al agrupar los metabolitos por su clase funcional, se observó un predominio de aminoácidos y péptidos, a parte de colinas y glicerofosocolinas que ya se menciono anteriormente. Algunas aproximaciones indican que la leche porcentadene un porcentaje de aminoácidos de entre el 5 y el 10 % del contenido total de nitrógeno (Moughan et al., 2024).

La limitada detección de carbohidratos y ácidos grasos de cadena corta en este estudio puede atribuirse al uso exclusivo del modo de ionización positiva. Se ha demostrado que metabolitos de naturaleza ácida y polar presentan mayor eficiencia de ionización en modo negativo (Liigand et al., 2017). Asimismo, las condiciones cromatográficas y el tipo de ionización seleccionados influyen directamente en la detección de metabolitos (Lei et al., 2011).

Respecto a la variabilidad entre muestras, los análisis de abundancia, heatmap y PCA coincidieron que las diferencias entre madres fueron más marcadas que entre fracciones. Esta variabilidad podría deberse a diferencias fisiológicas, genéticas o ambientales, como la dieta materna, el perfil hormonal o el estilo de vida (Cortes-Macías et al., 2021; Golan & Assaraf, 2020). Estudios recientes también sugieren que variables como el tipo de parto o el nivel de estrés materno pueden afectar la composición metabólica de la leche (Juncker et al., 2025; Matharu et al., 2024).

El heatmap de los 20 metabolitos más variables reveló que algunos compuestos, como ciertos ácidos biliares conjugados (por ejemplo, *Glycodeoxycholic.acid.GDCA_R2*), se concentran preferentemente en las fracciones lipídica o entera, en línea con su naturaleza hidrofóbica. Aunque sus concentraciones suelen ser bajas en condiciones fisiológicas, se ha documentado que ciertos fármacos pueden incrementar sus niveles en la leche materna, sin que esto represente un riesgo clínico significativo para el lactante (Šimják et al., 2022). Este hallazgo sugiere

que tanto el tipo de fracción como las condiciones fisiológicas de la madre pueden modular el perfil de algunos metabolitos.

Este estudio presenta algunas limitaciones. En primer lugar, el uso exclusivo del modo de ionización positiva probablemente limitó la detección de metabolitos relevantes como azúcares y ácidos orgánicos. En segundo lugar, el reducido número de sujetos ($n = 4$) impide generalizar los hallazgos a una población más amplia. En futuras investigaciones se recomienda aplicar modo dual de ionización (positivo y negativo) para lograr una cobertura metabolómica más completa. Asimismo, incluir un mayor número de madres y considerar variables como el estadio de lactancia, la dieta materna y el estado de salud podría enriquecer el análisis y permitir asociaciones más profundas con factores fisiológicos.

En conjunto, los resultados indican que no hay un único factor que explique la variabilidad metabolómica, sino una interacción compleja entre el sujeto y la fracción de leche. Este hallazgo es consistente con estudios previos que destacan la alta complejidad y variabilidad de la composición de la leche humana, subrayando la importancia de analizar múltiples niveles para comprender su perfil funcional.

6 Conclusiones

El presente análisis permitió caracterizar el perfil metabolómico de leche humana en tres fracciones, revelando una mayor abundancia de metabolitos en las fracciones acuosa y entera. Se identificaron metabolitos dominantes como phosphocholine y L-carnitine, presentes en todas las fracciones y sujetos. Al agrupar por clase funcional, se observó un predominio de aminoácidos, péptidos y colinas. El análisis multivariante evidenció que las diferencias entre madres fueron más relevantes que las diferencias entre fracciones de leche, lo que sugiere una fuerte influencia relacionada posiblemente con factores fisiológicos, genéticos o ambientales.

Asimismo, se identificaron metabolitos con afinidad específica por ciertas fracciones, como los ácidos biliares conjugados, que se concentraron en la fracción grasa. El uso exclusivo del modo de ionización positiva limitó la detección de metabolitos polares y ácidos, lo que resalta la necesidad de aplicar enfoques analíticos complementarios en futuros estudios. Estos hallazgos subrayan la complejidad de la leche humana y la importancia de un análisis integral para comprender su composición y función.

7 Anexos

7.1 ANEXO 1

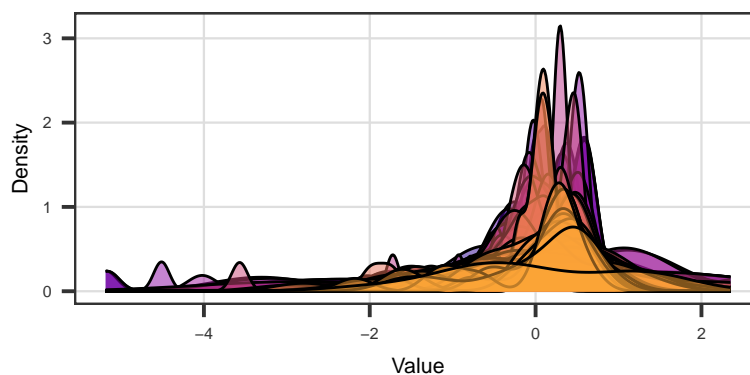


Figure 4: Distribución de intensidades normalizadas (log-pareto)

7.2 ANEXO 2

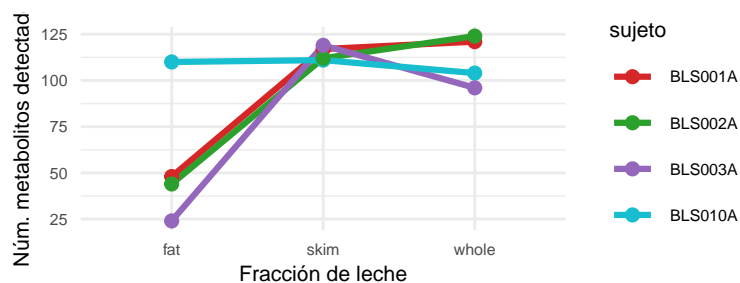


Figure 5: Número de metabolitos detectados por fracción de leche y madre

Referencias

- Carney, O. (n.d.). *Global metabolomics of human milk fractions - ST000957*. <https://www.metabolomicsworkbench.org/data/DRCCMetadata.php?Mode=Study&StudyID=ST000957>
- Cortes-Macías, E., Selma-Royo, M., García-Mantrana, I., Calatayud, M., González, S., Martínez-Costa, C., & Collado, M. C. (2021). Maternal Diet Shapes the Breast Milk Microbiota Composition and Diversity: Impact of Mode of Delivery and Antibiotic Exposure. *The Journal of Nutrition*, 151(2), 330–340. <https://doi.org/10.1093/jn/nxaa310>

- Golan, Y., & Assaraf, Y. G. (2020). Genetic and Physiological Factors Affecting Human Milk Production and Composition. *Nutrients*, 12(5), 1500. <https://doi.org/10.3390/nu12051500>
- Hampel, D., Shahab-Ferdows, S., Nguyen, N., Kac, G., & Allen, L. H. (2021). High-throughput analysis of water-soluble forms of choline and related metabolites in human milk by UPLC-MS/MS and its application. *Frontiers in Nutrition*, 7, 604570. <https://doi.org/10.3389/fnut.2020.604570>
- Jácome, J. (2025). *Repositorio PEC1*. <https://github.com/juano3101/Jacome-Navarrete-Juan-PEC1>
- Juncker, H. G., Jakobsen, R. R., Naninck, E. F. G., Davids, M., Herrema, H., Van Goudoever, J. B., De Rooij, S. R., & Korosi, A. (2025). Maternal stress in the early postpartum period is associated with alterations in human milk microbiome composition. *Brain, Behavior, and Immunity*, 124, 74–84. <https://doi.org/10.1016/j.bbi.2024.11.020>
- Lei, Z., Huhman, D. V., & Sumner, L. W. (2011). Mass Spectrometry Strategies in Metabolomics. *Journal of Biological Chemistry*, 286(29), 25435–25442. <https://doi.org/10.1074/jbc.R111.238691>
- Liigand, P., Kaupmees, K., Haav, K., Liigand, J., Leito, I., Girod, M., Antoine, R., & Krueve, A. (2017). Think Negative: Finding the Best Electrospray Ionization/MS Mode for Your Analyte. *Analytical Chemistry*, 89(11), 5665–5668. <https://doi.org/10.1021/acs.analchem.7b00096>
- Maheshwari, A. (2022). Fats in Human Milk: 2022 Updates on Chemical Composition. *Newborn*, 1(4), 384–396. <https://doi.org/10.5005/jp-journals-11002-0050>
- Matharu, D., Ponsero, A. J., Lengyel, M., Meszaros-Matwiejuk, A., Kolho, K.-L., De Vos, W. M., Molnar-Gabor, D., & Salonen, A. (2024). Human milk oligosaccharide composition is affected by season and parity and associates with infant gut microbiota in a birth mode dependent manner in a Finnish birth cohort. *eBioMedicine*, 104, 105182. <https://doi.org/10.1016/j.ebiom.2024.105182>
- Moughan, P. J., Deglaire, A., Yan, Y., Wescombe, P., Lim, W. X. J., Stroebinger, N., Duan, S., Szeto, I. M.-Y., & Hodgkinson, S. (2024). Amino acid requirements of the infant: The amino acid composition of human breast milk. *Frontiers in Nutrition*, 11, 1446565. <https://doi.org/10.3389/fnut.2024.1446565>
- Poulsen, K. O., Meng, F., Lanfranchi, E., Young, J. F., Stanton, C., Ryan, C. A., Kelly, A. L., & Sundekilde, U. K. (2022). Dynamic changes in the human milk metabolome over 25 weeks of lactation. *Frontiers in Nutrition*, 9, 917659. <https://doi.org/10.3389/fnut.2022.917659>
- Šimják, P., Petr, T., Kaslová, B., Fejfar, T., Hůlek, P., Pařízek, A., & Vitek, L. (2022). Ursodeoxycholic acid use in lactating female patients is associated with clinically negligible concentrations of this bile acid in breast milk. *Scientific Reports*, 12(1), 19543. <https://doi.org/10.1038/s41598-022-24253-y>
- Torrez Lamberti, M. F., DeBose-Scarlett, E., Garret, T., Parker, L. A., Neu, J., & Lorca, G. L. (2020). Metabolomic Profile of Personalized Donor Human Milk. *Molecules*, 25(24), 5783. <https://doi.org/10.3390/molecules25245783>