

Análisis de Metabolitos Identificados por Espectrometría de Masas de Resonancia Magnética (MRMS) utilizando equipo Bruker Solarix 7.0T de Espectrometría de Masas Avanzada.

Los metabolitos fueron evaluados mediante la técnica FT-ICR MS. Los resultados de este equipo son datos de intensidad contra m/z (masa sobre carga), estos son almacenados en un espectrograma por software propietario de Bruker, del mismo modo, la selección de picos (peak picking) y asignación de fórmulas se realiza también con software de esta empresa. Para el caso de la asignación de fórmulas, se toman en cuenta los valores de m/z y se comparan contra la HMDB (Human Metabolome Database). Estos resultados son exportados en una hoja de cálculo que contiene, la matriz de m/z (agrupada por picos) contra las intensidades, además del identificador de la muestra y el grupo experimental al que corresponde (calostro, meconio o copro) y los datos de asignación de las masas.

Esta matriz de datos es importada a R, en tres archivos csv, el primero contiene la matriz de m/z contra las intensidades; el segundo contiene los datos de la muestra (identificador y grupo experimental) y el tercero los datos de los picos (m/z , fórmula y nombre). Estos datos son transformados con el método VSN (variance stabilizing normalization), la gráfica de hexágonos (Figura 1) es un diagnóstico del ajuste de este modelo para el conjunto de datos trabajado; por otra parte, las gráficas de barras (Figura 2), muestran la variabilidad entre las muestras en términos de sumatoria de intensidades por muestra de picos por debajo del límite de detección, por muestra.

Previo a la normalización, los datos son visualizados con ayuda de un PCA (Principal Component Analysis), esto es un método de aprendizaje no supervisado que tiene como objetivo corroborar la clasificación de las muestras y ayudar en la clasificación de los picos por medio de la reducción de dimensiones en la matriz de datos a sus componentes principales. Por otra parte, se calcula el número óptimo de clústeres para buscar, en el caso de los picos, posibles agrupaciones entre

metabolitos por el método de k-means, los centroides son calculados a partir del promedio de las posiciones de la ordinación (PCA) en la gráfica (Figura 3).

Continuando con el problema de clasificación de las muestras y metabolitos, el heatmap es también una buena manera de observar la formación de clústeres en el conjunto de datos. A partir de la matriz de datos normalizados, se elaboró un mapa de calor (heatmap) de los metabolitos con una prevalencia en al menos el 60% de las muestras y se realizó una clusterización jerárquica, los datos fueron ordenados en el mapa de calor por el método OLO (Optimal Leaf Ordering) y se escalaron por columnas para observar el perfil de metabolitos (Figura 4).

Finalmente, se realizó un análisis diferencial, el modelo de distribución normal no representa un buen ajuste para la distribución de datos de metaboloma, debido a la abundancia de PMVs (Point Mass Values), es decir valores que son asignados como ceros. Estos valores pueden tener dos orígenes: BPMVs y TPMVs, es decir de origen biológico y de origen técnico y son responsables de la mayor parte de la variabilidad en la muestra. Las distribuciones que mejor funcionan para el tratamiento de estos datos son las distribuciones de la familia tweedie. Por otra parte, los valores que no son PMVs, se ajustan a la distribución normal debido a que fueron transformados por el método VSN. De manera que se aplicó un modelo llamado LIM-LRT (Left-Inflated Mixture model Likelihood Ratio Test), para realizar un análisis diferencial entre dos grupos, este modelo consigue diferenciar entre BPMVs y TPMVs y, a partir de eso, obtener los metabolitos presentes de manera diferencial (diffest) y sus valores de p. Estos valores de p, son corregidos por medio de Benjamini-Hochberg para la búsqueda de falsos positivos. Los resultados de los análisis diferenciales se ilustran en los gráficos de volcán entre los grupos HC vs MS, HC vs NM y MS vs NM (Figuras 5, 6 y 7).

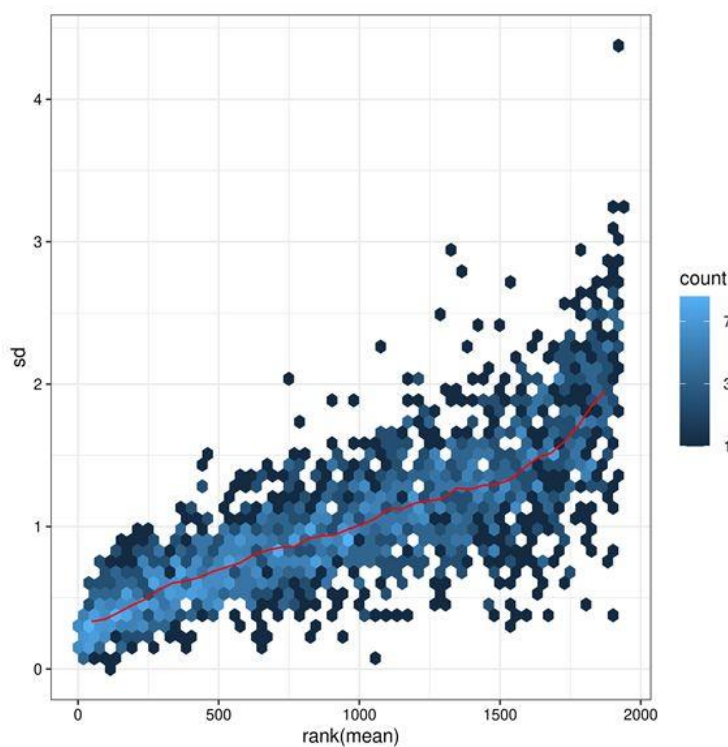


Figura 1. Gráfica de hexágonos donde se muestra el ajuste de la normalización de los datos mediante el método VSN (variance stabilizing normalization).

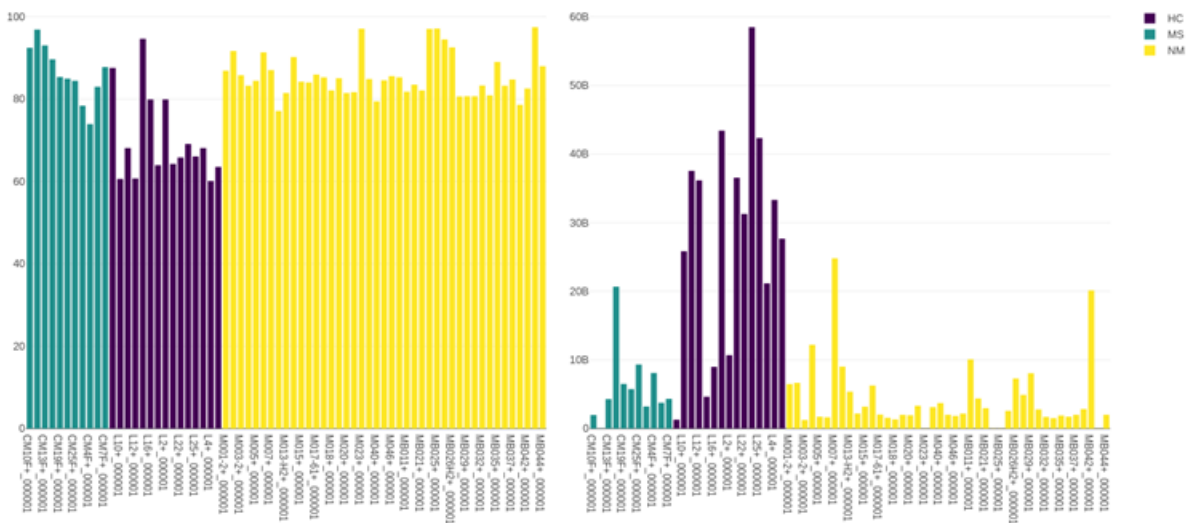


Figura 2. Gráficas de barras. Del lado izquierdo se observa el porcentaje de metabolitos por debajo del umbral de detección del equipo. Del lado derecho se observa la sumatoria de las intensidades detectadas con el equipo por muestra. HC: Calostro Humano, MS: Copro Madre, NM: Copro Neonato.

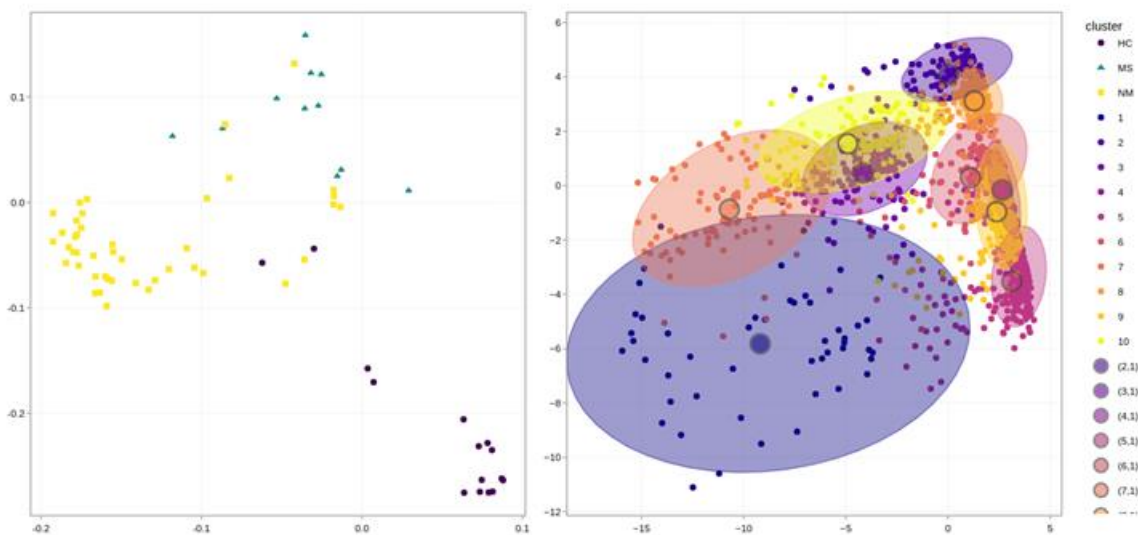


Figura 3. Gráficos de dispersión donde se observa el PCA (Principal Component Analysis). Del lado izquierdo se observan la distribución de las muestras. Del lado derecho, se agruparon los metabolitos en clústeres por el método *k*-means, se calculó un número óptimo de 10 clústeres para este conjunto de datos, los cuáles se muestran como un área sombreada de color, los puntos grandes representan el centroide de cada clúster.

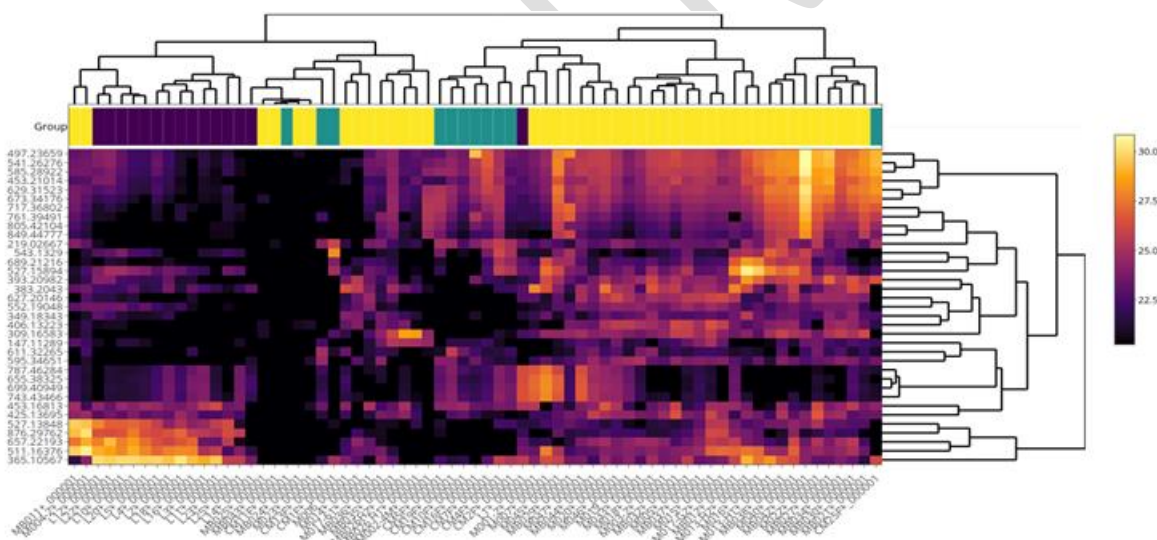


Figura 4. Mapa de calor de los metabolitos más prevalentes, detectados en al menos el 60% de las muestras, en el eje de las y se observa m/z medido con el equipo, el color representa la intensidad de la señal, en el eje de las x se ilustran las muestras. Las filas y las columnas del diagrama fueron clusterizadas de manera jerárquica por el método OLO (hierarchical clustering with Optimal Leaf Ordering). Los colores de la parte superior representan el tipo de muestra (HC: Calostro Humano, MS: Copro Madre, NM: Copro Neonato).

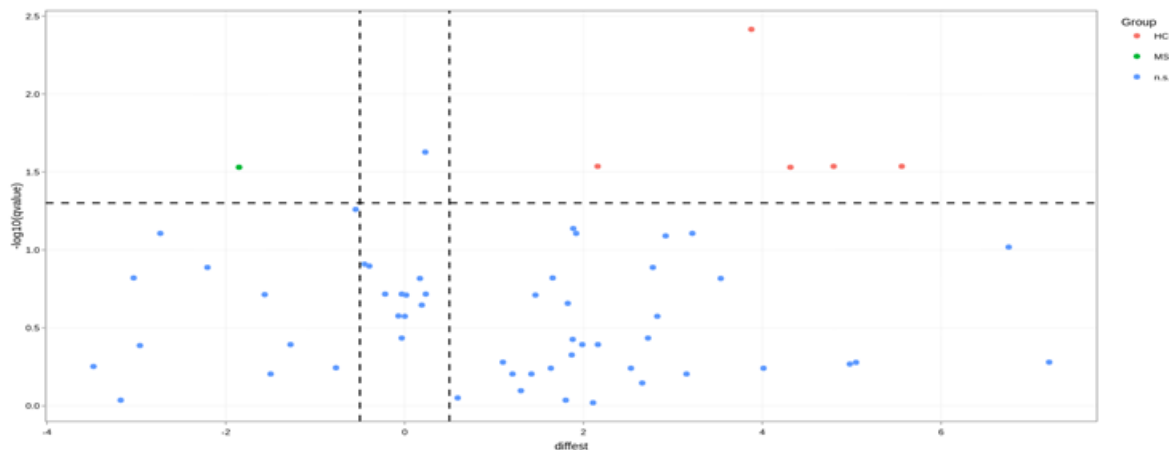


Figura 5. El análisis diferencial entre los grupos HC: Calostro Humano y MS: Copro Madre se ilustra mediante un gráfico de volcán. En el eje de las y: \log_{10} de los valores q , en el eje de las x: el estadístico que evalúa la magnitud de la diferencia entre ambos grupos.

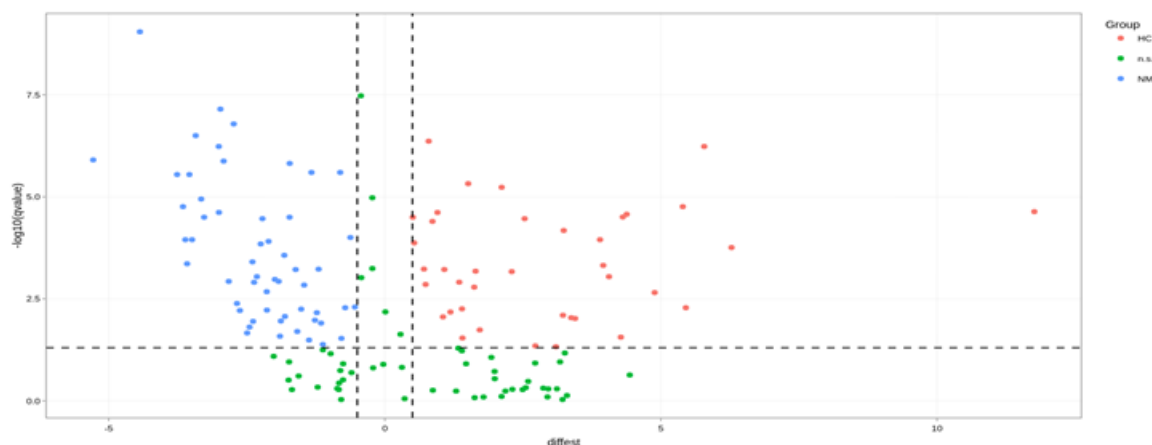


Figura 6. El análisis diferencial entre los grupos HC: Calostro Humano y NM: Copro Neonato se ilustra mediante un gráfico de volcán. En el eje de las y: \log_{10} de los valores q , en el eje de las x: el estadístico que evalúa la magnitud de la diferencia entre ambos grupos.

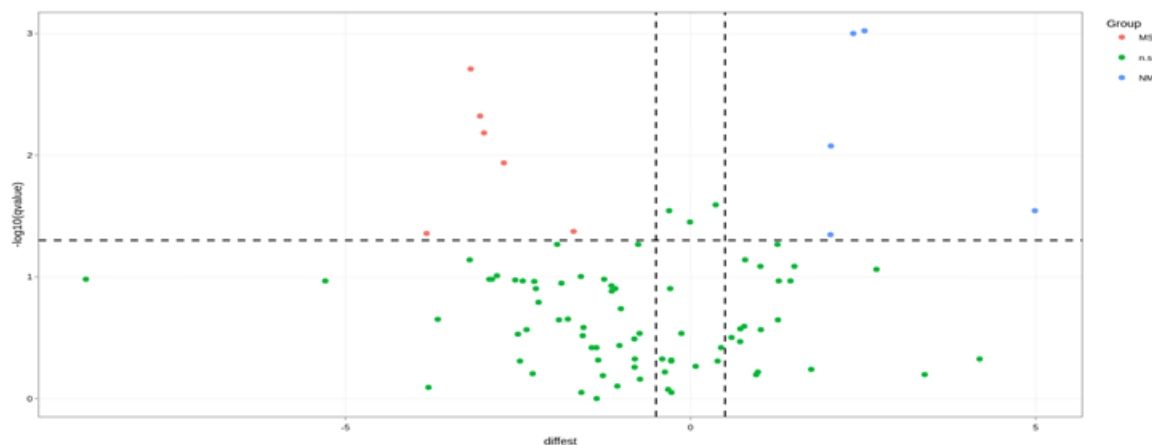


Figura 7. El análisis diferencial entre los grupos MS: Copro Madre y NM: Copro Neonato se ilustra mediante un gráfico de volcán. En el eje de las y: \log_{10} de los valores q , en el eje de las x: el estadístico que evalúa la magnitud de la diferencia entre ambos grupos.