



UNIVERSIDAD NACIONAL AUTÓNOMA DE MEXICO

PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS QUÍMICAS

**ESTUDIO DEL PERFIL METABOLÓMICO POR RESONANCIA MAGNÉTICA
NUCLEAR DE ORINA Y SANGRE DE NEONATOS PREMATUROS ENFERMOS
MEXICANOS**

**PROYECTO DE INVESTIGACIÓN
PARA OPTAR POR EL GRADO DE**

MAESTRO EN CIENCIAS

PRESENTA

Q.A. CIRCE CASANDRA HERNÁNDEZ ESPINO

DRA. NURIA ESTURAU ESCOFET
INSTITUTO DE QUÍMICA

CDMX, Agosto 2020



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS QUÍMICAS

ESTUDIO DEL PERFIL METABOLÓMICO POR RESONANCIA MAGNÉTICA
NUCLEAR DE ORINA Y SANGRE DE NEONATOS PREMATUROS ENFERMOS
MEXICANOS

PROYECTO DE INVESTIGACIÓN

PARA OPTAR POR EL GRADO DE

MAESTRO EN CIENCIAS

PRESENTA

Q.A. CIRCE CASANDRA HERNÁNDEZ ESPINO

DRA. NURIA ESTURAU ESCOFET
INSTITUTO DE QUÍMICA



Ciudad de México, Agosto 2020

AGRADECIMIENTOS

A la Universidad Nacional Autónoma de México, Instituto de Química y el Programa de Maestría y Doctorado en Ciencias Químicas, por la formación para este grado.

Al Laboratorio de Resonancia Magnética Nuclear (LURMN) por el equipo y tiempo de máquina necesario para la realización de esta tesis.

A mi tutora, la doctora Nuria Esturau, por la guía en este trabajo de maestría, las herramientas necesarias para su correcta realización y comentarios a lo largo de todo el trabajo realizado.

A la colaboración y apoyo en la realización del presente estudio de la Dra. Marcela Vela Amieva del Instituto Nacional de Pediatría, del Dr. Luis Rodrigo Macías Kauffer del Instituto Nacional de Medicina Genómica y de la Maestra Isabel Ibarra González del Instituto de Investigaciones Biomédicas UNAM-Instituto Nacional de Pediatría.

Al proyecto IQ-INP “APROXIMACIONES AL METABOLOMA DE LA PREMATUREZ MEDIANTE RESONANCIA MAGNÉTICA NUCLEAR Y ESPECTROMETRÍA DE MASAS EN TÁNDEM” por el financiamiento para su realización.

Al Consejo Nacional de Ciencia y Tecnología (CONACYT) por la beca otorgada dentro del Programa Nacional de Posgrados de Calidad, con número CVU 841880 y número de becario 629415.

PRESENTACIÓN EN CONGRESOS Y SIMPOSIOS

Los resultados de esta investigación fueron presentados bajo el nombre “*¹H-NMR metabolomic profiling of the feeding impact on the urinary metabolome. A preliminary study in preterm neonates*” en las siguientes reuniones académicas:

- *Colloquium Spectroscopicum Internationale XLV*, en la Unidad de Congresos de la Universidad Nacional Autónoma de México, Ciudad de México, acontecida del 10 al 14 de junio de 2019 bajo la modalidad de poster.
- Simposio Interno del Instituto de Química, acontecido del 19 al 21 de junio bajo la modalidad poster.

CONTENIDO

1.	INTRODUCCIÓN	1
2.	OBJETIVOS E HIPOTESIS.....	2
2.1.	HIPOTESIS	2
2.2.	OBJETIVO GENERAL	2
2.3.	OBJETIVOS PARTICULARES.....	2
3.	MARCO TEORICO.....	3
3.1.	INFANTES PREMATUROS	3
3.2.	METABOLÓMICA	4
3.3.	ABORDAJE EXPERIMENTAL	4
3.3.1.	PLATAFORMAS ANALÍTICAS	4
3.3.2.	RESONANCIA MAGNÉTICA NUCLEAR.....	5
3.3.3.	PROCESAMIENTO DE LOS ESPECTROS	6
3.3.4.	ANÁLISIS ESTADÍSTICO MULTIVARIANTE.	7
3.3.5.	IDENTIFICACIÓN DE METABOLITOS	9
4.	METODOLOGÍA.....	10
4.1.	PACIENTES DE ESTUDIO	10
4.1.1.	CRITERIOS DE INCLUSIÓN Y ELIMINACIÓN	10
4.1.2.	CONSENTIMIENTO INFORMADO Y APROBACIÓN DEL ESTUDIO	
	10	
4.2.	VARIABLES REGRESORAS Y CLASIFICATORIAS	10
4.3.	TOMA DE MUESTRAS	11
4.4.	PREPARACIÓN DE LA MUESTRA	11
4.5.	EXPERIMENTOS DE RMN.....	11
4.6.	PRE-PROCESAMIENTO DE LAS MUESTRAS.....	11
4.7.	CUANTIFICACION DE METABOLITOS EN ORINA	12
4.8.	ANÁLISIS ESTADÍSTICO MULTIVARIANTE.....	12
5.	RESULTADOS Y ANÁLISIS DE RESULTADOS	13
5.1.	ANÁLISIS DE SANGRE	13
5.1.1.	ESPECTROS DE RMN.....	13
5.1.2.	ANÁLISIS DE LOS ESPECTROS	13
5.2.	ANÁLISIS DE ORINA.....	14
5.2.1.	ESPECTROS DE RMN.....	14
5.2.2.	ANÁLISIS DE LOS ESPECTROS	15
5.2.3.	ANÁLISIS DE LA CONCENTRACIONES	17
6.	CONCLUSIONES.....	23

INDICE DE ACRONIMOS Y SIMBOLOGÍA

Acrónimo	Significado
1D-NOESY	Espectroscopia Monodimensional de Efecto Nuclear Overhauser (<i>One Dimensional Nuclear Overhauser Effect Spectroscopy</i>)
¹ H	Protón
¹ H-RMN	Resonancia Magnética Nuclear de Protón
AUROC	Área bajo la característica del operador receptor (<i>Area under the receiver operator characteristic</i>)
CPMG	Carr Purcel Meighboom Gill
DOSY	Espectroscopia de Difusion ordenada por RMN (<i>Diffusion-Ordered NMR Spectroscopy</i>)
FA	Análisis de Factores (<i>Factor Analysis</i>)
FID	Decaimiento libre inducido (<i>Free Induction Decay</i>)
HSQC	Espectroscopia de Coherencia Heteronuclear con Cuántica Simple (<i>Heteronuclear Single Quantum Coherence Spectroscopy</i>)
IR-FT	Espectroscopia de infrarrojo con transformada de Fourier (<i>Infra red-Fourier transform</i>)
MS-GC	Cromatografía de gases acoplada a espectrometría de masas (<i>Mass Spectrometry- Gas chromatography</i>)
MS-LC	Cromatografía de Líquidos acoplada a Espectroscopia de Masas (<i>Mass spectrometry- Liquid chromatography</i>)
NMC	Número de clasificaciones erradas (<i>Number of misclassifications</i>)
OPLS	Proyección Ortogonal a Estructuras Latentes (<i>Orthogonal Projection to Latent Structures</i>)
OPLS-DA	Análisis Discriminante de Proyección Ortogonal a Estructuras Latentes (<i>Orthogonal Projection to Latent Structures Discriminant Assay</i>)
PC	Componente Principal (<i>Principal Component</i>)
PCA	Análisis de Componentes Principales (<i>Principal Component Analysis</i>)
PLS	Mínimos Cuadrados Parciales (<i>Partial Least Squares</i>)
PLS-DA	Análisis Discriminante de Mínimos Cuadrados Parciales (<i>Partial Least Squares Discriminant Assay</i>)
Q ²	Coeficiente de predicción
R ² X	Varianza en la matriz X
R ² Y	Varianza asociada a la matriz Y
RMN	Resonancia Magnética Nuclear
SDG	Semanas de gestación
TOCSY	Espectroscopia de Correlación Total (<i>Total Correlation Spectroscopy</i>)

1. INTRODUCCIÓN

La prematuridad es considerada una forma de parto anormal en la que el nacimiento ocurre antes de las 37 semanas de gestación, con lo que su desarrollo puede estar comprometido, siendo una causa importante de morbilidad y mortalidad neonatal. A pesar de que hay múltiples estudios realizados desde distintos enfoques, la cantidad de muestra que se puede tomar de los recién nacidos prematuros juega en contra cuando se trata de investigación bioquímica, que se caracteriza por requerir cantidades de muestra por encima de lo que es posible obtener con esta clase de pacientes.

En este último punto la metabolómica, ciencia que se encarga del análisis de los metabolitos en un sistema biológico, permite identificar cambios sutiles en los perfiles metabólicos durante diferentes estados fisiológicos y/o patológicos. En específico la espectroscopia de resonancia magnética nuclear (RMN) es una técnica analítica que presenta muchas ventajas, siendo la principal la poca cantidad de muestra que requiere para realizar los análisis (μL).

Con el enfoque de metabolómica basada en RMN, el presente trabajo estudia el metabolismo de recién nacidos prematuros enfermos a lo largo de su estadía en la Unidad de Cuidados Intensivos. El presente estudio es resultado de una colaboración entre el Instituto de Química de la Universidad Nacional Autónoma de México y el Laboratorio de Errores Innatos del Metabolismo y Tamiz Neonatal del Instituto Nacional de Pediatría.

En el estudio se realizó el seguimiento metabolómico de sangre y orina de recién nacidos prematuros desde su ingreso al Servicio de Neonatología hasta su egreso. Las metodologías de preparación de muestras fueron obtenidas de la literatura, que en caso de sangre fue la extracción con metanol de muestras de sangre impregnadas en papel filtro específico para tamiz neonatal, y en caso de orina fue el análisis directo previa adición de un buffer de fosfatos.

Después de analizar los resultados obtenidos con las muestras de sangre y realizar algunas pruebas adicionales decidimos que la extracción con metanol para muestras de sangre del tamiz neonatal no es la preparación ideal para estudios de RMN, por lo que, aunque estos datos se exponen en el informe, no se toman en cuenta para análisis metabolómico. Los datos de las muestras de orina mostraron correlación con la clasificación por edad gestacional, la edad gestacional corregida, el peso a la toma de muestra y con el tipo de nutrición administrada, siendo ésta última la correlación más evidente de todas.

La contribución principal del trabajo de investigación es el análisis del metaboloma por RMN de los recién nacidos mexicanos prematuros enfermos. De acuerdo con nuestro conocimiento es la primera vez que se hace este tipo de estudio en nuestro país.

2. OBJETIVOS E HIPOTESIS

2.1. HIPOTESIS

Diferentes características de recién nacidos prematuros enfermos hospitalizados pueden ser modeladas a partir de la información colectada en el metaboloma urinario y sanguíneo de estos durante su estadio en una Unidad de Cuidados Intensivos.

2.2. OBJETIVO GENERAL

Evaluar la progresión de 36 recién nacidos prematuros enfermos ingresados en el Servicio de Neonatología del Instituto Nacional de Pediatría, a través del análisis de su metaboloma sanguíneo y urinario mediante la técnica de espectroscopia de resonancia magnética nuclear.

2.3. OBJETIVOS PARTICULARES

- Adquirir los espectros de RMN de las muestras de orina y sangre a partir de la selección de un protocolo adecuado para cada fluido.
- Realizar el pre-procesamiento y la reducción de la dimensionalidad de los espectros para los análisis multivariados.
- Generar modelos de análisis multivariable para identificar patrones de comportamiento en las muestras sin supervisión y supervisados con la información de tipo de nutrición, edad gestacional; a la edad gestacional corregida; o a otros factores de interés clínico.
- Identificar y cuantificar los metabolitos en los espectros de las muestras de orina y sangre.
- Generar modelos de análisis multivariable para determinar patrones de comportamiento en las muestras sin supervisión y supervisados con la información de los metabolitos identificados y cuantificados, de manera análoga a lo realizado con los espectros de RMN.
- Identificar a los metabolitos responsables de la segregación entre los grupos estudiados, así como determinar la validez de los modelos.

3. MARCO TEORICO

En el presente capítulo se da una descripción general del estado de prematuridad y lo que se conoce de dicha condición en México, y se describen los estudios de metabólica por RMN.

3.1. INFANTES PREMATUROS

La Organización Mundial de la Salud (OMS) considera a un infante prematuro como un bebé nacido antes de las 37 semanas de gestación (SDG). La prematurez es la principal causa de morbilidad y mortalidad neonatal, siendo la responsable de la mayor parte de muertes en niños sin malformaciones y de la mitad de las discapacidades neurológicas congénitas.¹

Existe una subclasificación del tipo de prematuro según la edad gestacional, que está conformada por:

- Prematuros extremos, nacidos antes de 28 semanas de gestación.
- Muy prematuros, nacidos de 28 a 32 semanas de gestación.
- Prematuros moderados tardíos, nacidos de 32 a 37 semanas de gestación.

La frecuencia de nacimientos prematuros en México se estima del 5%, aunque no se cuenta con números bien definidos. El IMSS reportó en el 2017 el nacimiento de 41,664 infantes prematuros de un total de 425,516 nacimientos, lo que representa una incidencia del 9.8%.²

Hay muchos factores que se han ligado a la presencia de nacimientos prematuros, entre otros, patologías como la preeclampsia o diabetes gestacional, edad materna extrema (madres adolescentes o madres cercanas a la menopausia), hipertensión, infecciones vaginales, mala nutrición, toxicomanías, un nivel socioeconómico bajo, malformación fetal, gestación múltiple, entre otras.

La investigación bioquímica en infantes enfermos está limitada por las pequeñas cantidades de sangre y orina disponibles para el análisis, que muchas veces ofrecen una imagen incompleta de los padecimientos bioquímicos existentes.³

Existen algunos estudios acerca de la relación del perfil bioquímico de los prematuros con características como sexo y edad gestacional⁴, y con el desarrollo de patologías propias de los prematuros, como la enterocolitis necrotizante, sepsis, o la muerte. De igual manera, se tiene poca información de la evolución del perfil bioquímico de los prematuros.

Herramientas como la metabólica pueden ser de una gran ayuda para establecer el perfil de los infantes prematuros, debido a la posibilidad de identificar

¹ Organización Mundial de la Salud. <http://www.who.int/es/news-room/fact-sheets/detail/preterm-birth>

² Miranda, P. Nacimientos Prematuros: Cada Vez Más Frecuentes: IMSS. Sociedad, Nación. El Universal. 15 de julio de 2018, 19:05.

³ Brown, J. C.; Mills, G. A.; Sadler, P. J.; Walker, V., 1H NMR studies of urine from premature and sick babies. *Magnetic resonance in medicine* **1989**, 11 (2), 193-201.

⁴ Hawdon, J.; Platt, M. W.; Aynsley-Green, A., Patterns of metabolic adaptation for preterm and term infants in the first neonatal week. *Archives of disease in childhood* **1992**, 67 (4 Spec No), 357-365.

y cuantificar los metabolitos, con el fin de establecer biomarcadores para cada estado fisiológico o patológico que se desea investigar.

3.2. METABOLÓMICA

La metabolómica es la ciencia que se encarga del análisis de la totalidad de los metabolitos en un sistema biológico, denominado metaboloma. Permite identificar cambios sutiles en los perfiles metabólicos durante diferentes estados fisiopatológicos. Por tanto, permite la medición del impacto ambiental, la alimentación⁵, el ejercicio⁶ o el estrés⁷ sobre un organismo.

En la metabolómica, los metabolitos estudiados son moléculas de bajo peso molecular, menor a 1500 Dalton, que intervienen en procesos celulares, revelando el funcionamiento de un órgano determinado o de un ser vivo.

Debido a la complejidad y característica dinámica de los metabolomas, esta clase de investigaciones requiere de técnicas analíticas reproducibles, robustas, sensibles y que no requieran de un pretratamiento complejo. En este contexto, las técnicas más ampliamente utilizadas son la espectroscopia de infrarrojo con transformada de Fourier (IR-FT)⁸, la espectrometría de masas (EM) acoplada a sistemas cromatográficos (EM-CG y EM-CL)⁹ y la RMN.

3.3. ABORDAJE EXPERIMENTAL

3.3.1. PLATAFORMAS ANALÍTICAS

La EM, acoplada a técnicas cromatográficas GC y LC, permite analizar una gran cantidad de metabolitos a muy bajas concentraciones, y eligiendo el método de separación adecuado, generalmente se consigue una gran selectividad. Sin embargo, existen ciertas limitaciones en la falta de estandarización de los métodos y un tratamiento de muestra tedioso, que hacen los experimentos poco reproducibles, además de que es una técnica destructiva.¹⁰

La RMN es una herramienta versátil y robusta que permite la identificación de los metabolitos de forma confiable y repetitiva, con un nivel de automatización muy alto. Es un tipo de análisis no destructivo y requiere un pretratamiento mínimo o

⁵ Barron, R.; Bermingham, K.; Brennan, L.; Gibney, E. R.; Gibney, M. J.; Ryan, M. F.; O'Sullivan, A., Twin metabolomics: the key to unlocking complex phenotypes in nutrition research. *Nutrition Research* **2016**, 36 (4), 291-304.

⁶ Escolano, C. V., Efectos de la acupuntura sobre la fatiga inducida por ejercicios físicos exhaustivos: una investigación metabolómica. *Revista Internacional de Acupuntura* **2015**, 9 (4), 131-133.

⁷ Zhou, X.; Liu, L.; Zhang, Y.; Pu, J.; Yang, L.; Zhou, C.; Yuan, S.; Zhang, H.; Xie, P., Metabolomics identifies perturbations in amino acid metabolism in the prefrontal cortex of the learned helplessness rat model of depression. *Neuroscience* **2017**, 343, 1-9.

⁸ Kuligowski, J.; Pérez-Guaita, D.; Escobar, J.; Lliso, I.; de la Guardia, M.; Lendl, B.; Vento, M.; Quintás, G., Infrared biospectroscopy for a fast qualitative evaluation of sample preparation in metabolomics. *Talanta* **2014**, 127, 181-190.

⁹ Lu, X.; Zhao, X.; Bai, C.; Zhao, C.; Lu, G.; Xu, G., LC-MS-based metabolomics analysis. *Journal of Chromatography B* **2008**, 866 (1), 64-76.

¹⁰ Dunn, W.B., Ellis, D.I. Metabolomics: Current analytical platforms and methodologies. *Trends Anal. Chem.* **2005**, 24(4), 285-294.

nulo de las muestras. Hoy en día, la aparición de las sondas criogénicas ha permitido llegar a concentraciones de mg/mL. Entre sus desventajas se encuentra la superposición de señales provenientes de diferentes metabolitos, y una baja sensibilidad con respecto a la EM.

3.3.2. RESONANCIA MAGNÉTICA NUCLEAR

3.3.2.1. SECUENCIAS DE PULSOS PARA METABOLOMICA

Dado que el compuesto más abundante en estudios de metabolómica es el agua, y que presenta una señal intensa en los espectros de RMN que impide la observación de las señales correspondientes a los metabolitos, es necesario la supresión de dicha señal, convirtiéndose en un aspecto crucial para la observación de las señales más pequeñas. Hay numerosas secuencias de pulsos que se usan en RMN para minimizar esta intensa señal, de las cuales las más comunes son:¹⁹

1. Presaturación (presat): es el procedimiento más simple. Se basa en aplicar un pulso de baja potencia y largo (pulso suave) a la frecuencia de resonancia del agua (para así suprimir su señal). Posteriormente se aplica el pulso de alta potencia y corta duración (pulso duro, usualmente de 90°), no selectivo con el fin de excitar todas las señales del espectro (Figura 1).

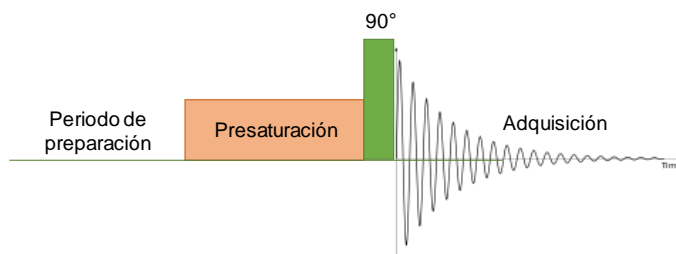


Figura 1. Secuencia de pulsos Presat

2. 1D NOESY-presat (*1 Dimension Nuclear Overhauser Effect Spectroscopy with Presaturation*)¹¹: muy utilizada en el análisis de muestras de orina, presentando una supresión más eficiente de la señal de agua con mejor línea base que el experimento *presat*. Toma ventaja de la diferencia de tiempo de relajación longitudinal (T_1) entre los núcleos de los metabolitos y el agua, de manera que con el segundo pulso de 90° cancela su componente magnética. La secuencia de pulsos se muestra en la Figura 2. Su uso se ha popularizado en metabolómica debido a una mínima necesidad de optimización, aunado a que se evita la pérdida de protones intercambiables.¹²

¹¹ McKay, R.T. How the 1D-NOESY suppresses solvent signal in metabolomics NMR spectroscopy: An examination of the pulse sequence. *Concepts Magn. Reson.* **2011**, 38A (5), 197-220.

¹² Teng, Q. Chapter 9. NMR-based metabolomics. *Structural biology: NMR Applications*. 2nd ed. Springer. 2013. 337, 339-340.

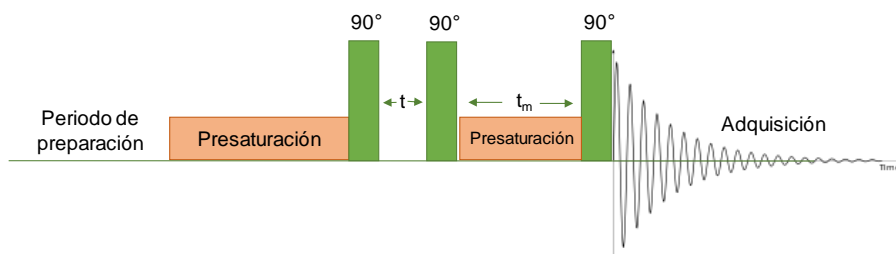


Figura 2. Secuencia de pulsos 1D NOESY.

3. CPMG (“Car Purcel Meiboom Gill”) con pre-saturación de agua^{13, 14} Es la secuencia más utilizada en el análisis de muestras de plasma y suero, ya que elimina las señales anchas producidas por las proteínas y los lípidos que distorsionan parcialmente la línea base y ocultan las señales de los metabolitos. Hace uso de las diferencias en los T_2 entre las moléculas de alto peso molecular, como las proteínas, y los metabolitos de bajo peso molecular. En la Figura 3 se muestra la secuencia de pulsos.

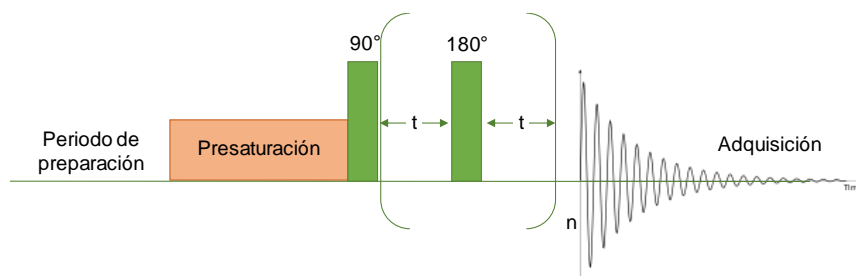


Figura 3. Secuencia de pulsos CPMG.

3.3.3. PROCESAMIENTO DE LOS ESPECTROS

Para extraer la información del espectro de RMN en metabolómica es necesario aplicar una serie de correcciones para que la información sea útil. Éstas incluyen la corrección de fase y línea base, la apodización del espectro, la referencia del espectro a un compuesto estándar de concentración conocida para la corrección del desplazamiento químico (DSS= 0.00 ppm o TSP= -0.015 ppm a 25°C y pH=7¹⁵), *binning* con un tamaño de caja de 0.04 ppm y normalización del espectro a una señal de referencia o al total de área.¹⁶ Este procesamiento ayuda a corregir los errores por integración, por desplazamiento químico y por concentración relativa en las muestras.

¹³ Car, H.Y.; Purcell, E.M. Effects of Diffusion on Free Precession In Nuclear Magnetic Resonance Experiments. *Phys. Rev.* **1954**, 94, 630-638.

¹⁴ Meighboom, S.; Gill, D. Modified Spin Echo Method for Measuring Nuclear Relaxation Times. *Rev. Sci. Instrum.* **1958**, 9 (8), 688-691.

¹⁵ Wishart, D. S.; Bigam, C. G.; Yao, J.; Abildgaard, F.; Dyson, H. J.; Oldfield, E.; Markley, J. L.; Sykes, B. D., 1 H, 13 C and 15 N chemical shift referencing in biomolecular NMR. *Journal of biomolecular NMR* **1995**, 6 (2), 135-140.

¹⁶ Larive, C. K.; Barding Jr, G. A.; Dinges, M. M., NMR spectroscopy for metabolomics and metabolic profiling. *Analytical chemistry* **2014**, 87 (1), 133-146.

En muestras que tengan metabolitos con comportamiento ácido-base, es posible encontrar variaciones de desplazamiento químico importantes entre muestras, por lo que también es frecuente el uso de algoritmos de alineación que permitan disminuir el error biológico para que los espectros puedan ser comparables entre sí.

3.3.4. ANALISIS ESTADÍSTICO MULTIVARIANTE.

En estudios metabolómicos es común la presencia de una gran cantidad de variables para la descripción de un número determinado de muestras, por lo que se hace indispensable el uso de técnicas de análisis multivariante para poder extraer la información presente en el estudio.

La adaptación de la quimiometría (análisis multivariable para datos químicos) a la espectroscopia de RMN fue probada en los inicios de los 80 en el análisis de mezclas complejas y demostró tener un amplio potencial en la metabolómica.

Dependiendo del objetivo que se persiga, los métodos quimiométricos pueden ser descriptivos o predictivos. En los primeros, las propiedades de compuestos químicos se modelan, con la intención de aprender las relaciones subyacentes y las estructuras de los mismos, mientras que, en los segundos, las propiedades de los sistemas químicos se modelan con la intención de predecir propiedades o comportamientos de interés.

De manera común, en primer lugar, se emplean los métodos de análisis no supervisados (que permiten evaluar la variabilidad de los datos sin tener en cuenta su clasificación dentro de ningún grupo o modelos de regresión), seguido de los modelos supervisados (que utilizan información de las muestras para ajustar la varianza a la clasificación o regresión que se busca).

3.3.4.1. ANÁLISIS DE COMPONENTES PRINCIPALES

El PCA es una técnica exploratoria no supervisada para la descripción de patrones y/o tendencias a partir de la descomposición espectral de la totalidad de los datos, y la retención de las denominadas componentes principales para la descripción del sistema en un número de dimensiones menor a la original. Dichas componentes principales son combinaciones lineales de las variables originales que presentan una alta correlación, y permiten visualizar la información de manera similar a la que se muestra en la Figura 4.

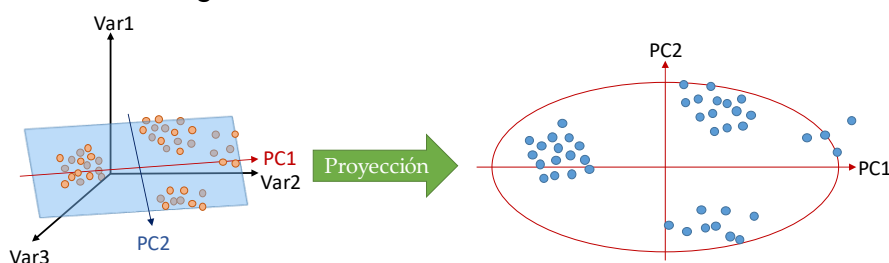


Figura 4. Representación gráfica de la reducción dimensional en un PCA con dos componentes principales a partir de un espacio inicial de 3 dimensiones.

Previo el uso de estas técnicas, es requerimiento común el estandarizar los datos hacia un comportamiento de distribución normal con media igual a cero y varianza

igual a uno. El objetivo es la minimización de la contribución de ruido del espectro y varianza experimental no requerida.

La reducción de variables se analiza generalmente con los gráficos de *scores* y *loadings*. El primero es una proyección en los dos componentes principales de mayor varianza, en la que los objetos cercanos en distancia muestran características similares mientras que objetos con distancias lejanas son diferentes. El segundo explica cuáles son las variables que marcan el comportamiento de dichas tendencias o agrupaciones.

Un aspecto clave a resaltar es que los datos pueden seguir cualquier tendencia de variación, y ésta no necesariamente tiene que reflejar aquello que se busca dentro del estudio.

3.3.4.2. REGRESION Y ANALISIS DISCRIMINANTE POR MÍNIMOS CUADRADOS

El método PLS es una técnica supervisada que puede ser usada para mejorar la separación entre los grupos de observaciones, rotando los componentes del PCA hasta que se obtenga el máximo grado de separación con base en la información previamente conocida acerca de las muestras, como se representa en la Figura 5.¹⁷

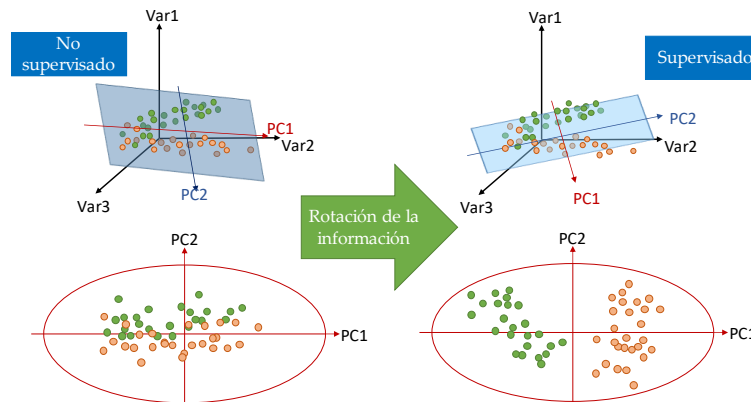


Figura 5. Rotación de la varianza en las componentes principales a partir de la relación con información conocida (respuesta) de las observaciones.

El modelo PLS-DA (*Partial Least Squares Discriminant Assay*) es una extensión del modelo PLS y se ocupa típicamente en situaciones de clasificación, para conocer las diferencias entre grupos nominales de dos o más tipos de objetos. El uso de PLS-DA se complica cuando hay más de dos grupos en el análisis, por lo que la forma de la asignación de los grupos es crucial, y puede incluir una matriz con una sola columna, o N-1 columnas, donde N es el número de clases.¹⁸

El modelo debe ser capaz de predecir razonablemente a qué grupo corresponde una muestra desconocida, conocido como la validación cruzada del modelo. De manera general, hay cuatro herramientas para evaluar la validez de los modelos de clasificación: el parámetro Q^2 evalúa el error entre la predicción categórica de

¹⁷ Brereton, R.G. 10. Biological and Medical Applications of Chemometrics. Applied Chemometrics for Scientist. Wiley. 2010, 287-318.

¹⁸ Szymańska, E.; Saccenti, E.; Smilde, A. K.; Westerhuis, J. A., Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies. *Metabolomics* 2012, 8 (1), 3-16.

la variable y la clase a la que pertenece, el parámetro DQ^2 se basa en el hecho de que el error de predicción es ignorado cuando la predicción es entre las clases, el número de errores de clasificación NMC (*Number of Misclassification*, por sus siglas en inglés) que evalúa la cantidad de elementos que son categorizados por predicción en otro grupo, y el Área bajo la característica del operador receptor (AUROC por sus siglas en inglés, *Area under the receiver operator characteristic*) que obtiene los parámetros de especificidad y sensibilidad, relacionadas con la tasa de verdaderos positivos y verdaderos negativos.¹⁸

Al igual que PCA, este tipo de modelos generan gráficos de *scores* y *loadings*, fácilmente interpretables. Los *scores* en este caso pueden ayudar a determinar qué variables químicas caracterizan cada clase, permitiendo la identificación de biomarcadores.¹⁷

Además del uso del PLS, el modelo OPLS es una extensión introducida por Trygg y Wold en 2002¹⁹ que ha sido muy importante para la interpretación de datos en las ciencias ómicas. Al igual que PLS busca la mayor separación entre las clases, pero a diferencia de éste, lo hace a partir de la búsqueda de componentes que varían sistemáticamente en las variables originales, y no estén correlacionadas con las variables regresoras, como problemas instrumentales, de muestreo o errores biológicos.

3.3.5. IDENTIFICACIÓN DE METABOLITOS

Una vez obtenido el espectro, se comparan los desplazamientos químicos y acoplamientos escalares de las señales con las de las referencias encontradas en las bases de datos de metabolitos. Entre las más conocidas están la *Human Metabolome Database* (hmdb.ca), el *Biological Magnetic Resonance Data Bank for Metabolomics* (bmrw.wisc.edu/metabolomics) o el *Maddison Metabolomics Consortium Database* (<http://mmcd.nmr.fam.wisc.edu/>). La implementación de algoritmos de reconocimiento de metabolitos (BATMAN^{20, 21}, Chenomx²¹), ha reducido considerablemente el tiempo de búsqueda para los investigadores, siendo recomendable en dichos casos que más de un espectroscopista identifique todos los metabolitos posibles en la muestra a su criterio, de manera que no esté sesgado por la visión de un solo investigador.

En caso de duda, se recurre a la comprobación del metabolito a partir de la comparación con estándares puros, o con la realización de experimentos de RMN selectivos y/o de correlación en 2 dimensiones, siendo los más usados el experimento TOCSY (por sus siglas en inglés *Total Correlation Spectroscopy*), y el experimento HSQC (por sus siglas en inglés *Heteronuclear Single Quantum Spectroscopy*).

¹⁹ Trygg, J.; Wold, S. Orthogonal Projections to Latent Structures (O-PLS). *J. Chemom.* **2002**, *16*, 119-128.

²⁰ Hao, J.; Astle, W.; De Iorio, M.; Ebbels, T. M., BATMAN—an R package for the automated quantification of metabolites from nuclear magnetic resonance spectra using a Bayesian model. *Bioinformatics* **2012**, *28* (15), 2088-2090.

²¹ Ellinger, J. J.; Chylla, R. A.; Ulrich, E. L.; Markley, J. L., Databases and software for NMR-based metabolomics. *Current Metabolomics* **2013**, *1* (1), 28-40.

4. METODOLOGÍA

4.1. PACIENTES DE ESTUDIO

4.1.1. CRITERIOS DE INCLUSIÓN Y ELIMINACIÓN

Fueron incluidos en el presente estudio neonatos prematuros enfermos derivados al Servicio de Neonatología del Instituto Nacional de Pediatría. A todos se les tomó muestra durante su estancia en la unidad pediátrica.

Para orina, aquellas muestras de infantes prematuros que no alcanzarán como mínimo los 400 μ L fueron eliminadas del estudio. Para sangre, aquellas muestras que no alcanzaran como mínimo 4 *bloodspots* (círculos de 2 mm de diámetro) fueron eliminadas del estudio.

Hubo muestras que no se pudieron asociar a ningún tipo de nutrición, debido a que el infante estaba sometido a ayuno. Dichas muestras fueron eliminadas de los correspondientes análisis, tanto de sangre como de orina.

Las muestras de orina cuyos espectros presentaron características anómalas fueron descartadas del análisis.

4.1.2. CONSENTIMIENTO INFORMADO Y APROBACIÓN DEL ESTUDIO

La firma de los padres de los recién nacidos fue recolectada en consentimientos informados donde se les brindó información acerca de los procedimientos de toma de muestra y de tratamiento y confidencialidad de los datos. Este estudio fue aprobado por los consejos de ética y de investigación de las instituciones colaboradoras.

4.2. VARIABLES REGRESORAS Y CLASIFICATORIAS

Las clasificaciones se tomaron de las definiciones de la OMS para la prematuridad. Para cada factor, la tabla 1 muestra el número de niveles existentes en el análisis, el código de clasificación ocupado, así como su significado.

Tabla 1. Relación de los factores utilizados en el análisis multivariable.

Factor	Niveles	Significado del nivel
Sexo	Femenino	NA*
	Masculino	NA*
Edad gestacional	Extremo	<28 SDG
	Muy prematuro	28<SDG<32
	Tardío	32<SDG<37
Peso	A	<1000 g
	B	1000-1500 g
	C	1500-2500 g
	D	2500-4000 g
	E	>4000 g
Tipo de nutrición	Enteral	A través de sonda nasogástrica
	Parenteral	A través de infusión arterial

NA: No aplica

4.3. TOMA DE MUESTRAS

A cada recién nacido se le tomarán 3 gotas de sangre en papel filtro y 5 ml de orina en los siguientes tiempos: ingreso, egreso y semanalmente durante su estancia hospitalaria.

Las tarjetas fueron clasificadas y almacenadas hasta el momento de su análisis. Las muestras de orina fueron colocadas en tubos Falcón de 15 mL y guardadas a -70°C hasta el momento de su análisis.

4.4. PREPARACIÓN DE LA MUESTRA

Las muestras de sangre se prepararon cortando 4 círculos de sangre (de 2.2 mm de diámetro cada uno) y se extrajeron con 600 μL de metanol deuterado (MeOD) durante una hora en un *shaker*. A cada muestra se le adicionó 10 μL de agua deuterada (D_2O) con 0.02 mM de TSP como estándar interno.

Las muestras de orina se descongelaron a temperatura ambiente, y 500 μL se transfirieron a tubos Eppendorf. Se sometieron a centrifugación en frío (4°C) a 3500 RPM por 5 minutos. Pasado este tiempo, 400 μL del sobrenadante se transfirió a otro Eppendorf, donde se agregaron 200 μL de buffer de fosfatos deuterado (1.5 M, $\text{pH}=7.4$) con TSP como estándar interno (0.01%). Se agito, se sometió nuevamente a centrifugación en frío (4°C) a 3500 RPM por cinco minutos, y finalmente 600 μL de la mezcla se llevaron a tubos de RMN de 5 mm.

4.5. EXPERIMENTOS DE RMN

Los espectros de RMN se obtuvieron en un equipo Bruker Avance III HD de 700 MHz (frecuencia de Larmor del protón: 699.97 MHz), equipado con una criosonda enfriada por Helio y un automuestreador Sample Jet.

Las muestras de extracto de sangre se adquirieron con la secuencia CPMG-presat (*cpmgpr1d*), 256 scans, 64k puntos complejos del dominio del tiempo con un ancho de ventana de 12 ppm (8403.361 Hz), un tiempo de reciclaje D1 de 4 s.

Las muestras de orina se adquirieron con la secuencia protón con pre-saturación (*noesygppr1d*), 256 scans, 64K puntos complejos del dominio del tiempo con un ancho de ventana de 21 ppm (14705.883 Hz) y un tiempo de reciclaje D1 de 4 s.

4.6. PRE-PROCESAMIENTO DE LAS MUESTRAS

Al decaimiento libre inducido (FID, por sus siglas en inglés *Free Induction Decay*) se les aplicó la transformada de Fourier (FT, por sus siglas en inglés *Fourier Transform*) en el programa Chenomx 12.8 (Chenomx NMR Suite; Alberta, Canadá). A los espectros de sangre se aplicó una función de apodización exponencial de 0.1, corrección automática de la fase y línea base. A los espectros de muestras de orina se aplicó una función de apodización exponencial de 0.3, corrección automática de la fase y línea base. La corrección del desplazamiento químico se realiza de manera automática, señalando como referencia al simple de TSP, localizado en -0.016 ppm.

A los espectros de sangre, se cortó la zona correspondiente al agua (4.50-4.99 ppm), al metanol (3.16-3.44 ppm) y a un polímero desconocido proveniente de la extracción del papel con metanol (1.00-1.30 ppm). A los espectros de orina se cortó las zonas correspondientes al agua (4.62-4.94 ppm) y urea (5.65-6.02 ppm).

Las matrices se construyeron con un ancho de caja de 0.04 ppm (aproximadamente 28 Hz) usando la paquetería *Automatic Statistical Identification in Complex Spectra* (ASICS) del programa R (*The R Project* v. 16.0.0) y una normalización al total de área para reducir el

efecto de la diferencia en concentraciones a lo largo de las muestras. Fueron exportadas en Excel 16.0 (Microsoft Office, Microsoft Corp.; Estados Unidos) donde se agregaron las variables clasificatorias que se evaluarían, y finalmente se llevaron al programa SIMCA 16.0 (Umetrics, Sartorius; Umea, Sweden) para su análisis.

4.7. CUANTIFICACION DE METABOLITOS EN ORINA

Se ubicó la mayor cantidad de metabolitos posibles existentes en la base de datos de programa Chenomx. Una vez que se establecieron los metabolitos presentes, se procedió a la cuantificación de estos a través de la relación de área con el estándar interno de concentración conocida.

Conocida la concentración, se construyó la matriz en Excel y la concentración de cada metabolito se normalizó a la concentración de creatinina. La matriz se exportó a SIMCA para su análisis estadístico.

4.8. ANALISIS ESTADÍSTICO MULTIVARIANTE

Las matrices de datos espectrales fueron centradas y escaladas previo a la construcción de los modelos. Primero se realizó la reducción dimensional PCA con dos componentes principales para conocer las variaciones intrínsecas a las muestras, y se evaluó si dicha variación estaba marcada en tendencias por alguna de las características que se querían evaluar. Se identificaron los anómalos estadísticos como aquellas muestras que salían del límite de confianza creado por la prueba hotellings T2 default del programa SIMCA y se eliminaron del análisis después de un censo acerca de su distancia a las demás muestras.

Posteriormente se realizó el modelo de clasificación PLS-DA y OPLS-DA contra las variables características que se desea evaluar: sexo, edad gestacional, peso a la toma de la muestra y tipo de nutrición. También se realizó el modelo de regresión OPLS para calibrar los datos de edad gestacional corregida y peso a la toma de la muestra.

La matriz de concentraciones normalizadas a la concentración de creatinina se centró y escaló con la función Varianza Unitaria. Al igual que las matrices espectrales, se realizó la reducción dimensional PCA, el modelo de clasificación OPLS-DA y el modelo de regresión OPLS.

Cuando alguno de los modelos de clasificación mostró tendencias, se procedió a realizar cálculos de permutación ($n=100$) para validación, y el número de clasificación erróneas (MCN) se utilizó para conocer el número de clasificados correctamente.

5. RESULTADOS Y ANÁLISIS DE RESULTADOS

5.1. ANÁLISIS DE SANGRE

5.1.1. ESPECTROS DE RMN

La Figura 6 expone seis espectros de extracto de sangre de muestras de los pacientes, tomadas al azar, las zonas donde aparecen las señales de: agua, metanol y una impureza fueron eliminadas. Se observa una ligera diferencia a lo largo del espectro, pero sobresale la ausencia de carbohidratos (región de 2.0 a 4.5 ppm) en todas las muestras.

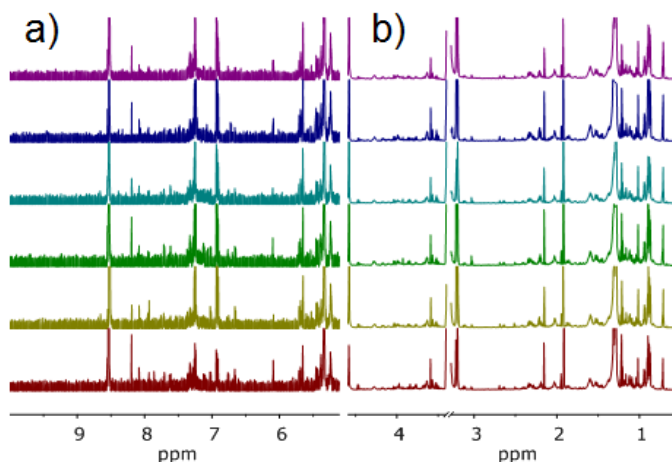


Figura 6. Espectros CPMG-presat 700 MHz de 10 muestras de extracto se sangre en metanol al azar. La zona aromática a) posee una ampliación de 20 veces comparada con la zona de azúcares y aminoácidos b).

5.1.2. ANALISIS DE LOS ESPECTROS

La Figura 8A muestra el *score plot* del PCA con un valor R^2X de 54.4%, y se coloreó según el tipo de nutrición que se había administrado al infante prematuro, Figura 8B. Se observó una tendencia en los datos: las muestras provenientes de nutrición parenteral obtienen valores positivos en el PC2 mientras que las muestras provenientes de nutrición enteral obtienen valores negativos. Por ello se decidió realizar el modelo de clasificación OPLS-DA para optimizar dicha diferencia e identificar los metabolitos causantes de esta variación.

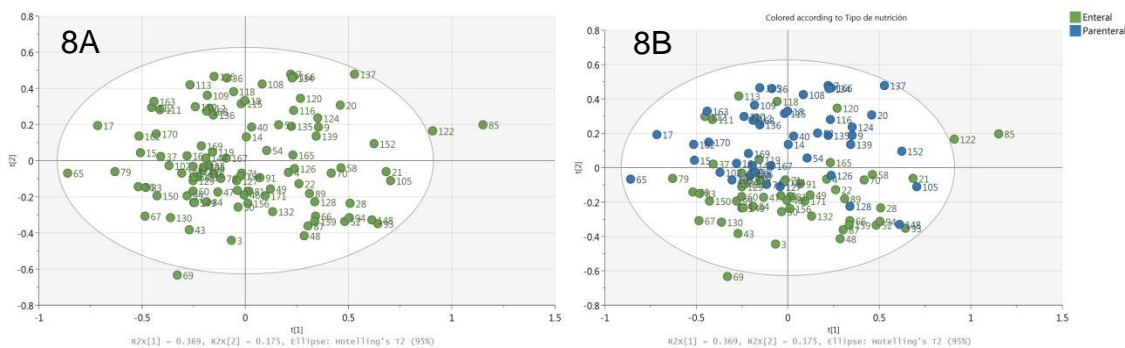


Figura 7. Score plot del PCA modelado con los datos espectrales de sangre. A) Sin colorear, B) Coloreado conforme a la nutrición brindada: verde- nutrición enteral y azul- nutrición parenteral.

El modelo OPLS-DA en la Figura 8 reduce la información útil a una variable latente y una variable ortogonal, que contienen el 74.8% de R^2Y , junto con un coeficiente de predicción Q^2 de 49.0%. Como se observa en la figura 8, las muestras tienden a separarse en la variable latente. Como prueba de validación se usó la prueba de permutaciones con 100 repeticiones que arroja valores de $R^2Y= 0.0, 0.379$ y de $Q^2= 0.0, -0.646$, y de acuerdo con las especificaciones del programa SIMCA de Umetrics-Sartorius, un valor de ordenada al origen en R^2Y menor a 0.45 y un valor de ordenada de Q^2 menor a 0.05 confirma la validez del modelo. Los modelos supervisados a las demás variables no mostraron valores de clasificación estadísticamente significativos.

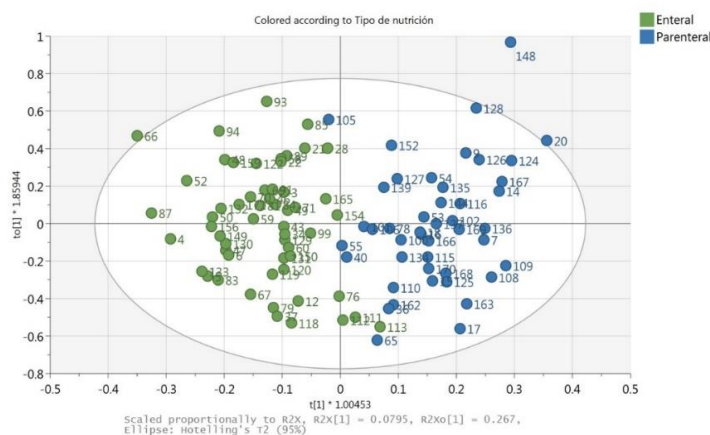


Figura 8. Score plot del OPLS-DA modelado con los datos espectrales de sangre. Los colores representan el tipo de nutrición brindada: verde- nutrición enteral y azul- nutrición parenteral.

Las variables importantes en el modelado de datos conforme el tipo de nutrición son: 1.9 y 1.3 en nutrición enteral, y 1.94, 0.86, 2.22, 1.5, 1.46, 1.82 y 8.5 en la nutrición parenteral. Sin embargo, debido a que los análisis se realizaron en metanol, y las bases de datos están realizadas en agua no se pudo realizar la identificación de dichos metabolitos.

5.2. ANALISIS DE ORINA

5.2.1. ESPECTROS DE RMN

La Figura 9 muestra 7 espectros correspondientes a muestras de 7 semanas consecutivas de un mismo paciente. Se observa como la concentración de los compuestos varían con el tiempo.

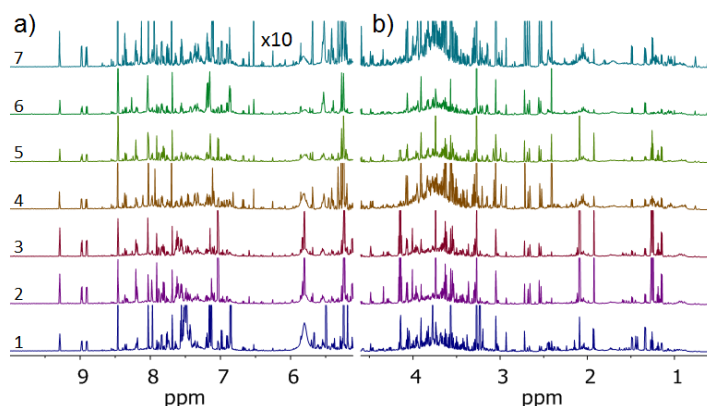


Figura 9. Espectros 1D NOESY 700 MHz de 10 muestras de orina al azar. a) Zona aromática (10.0 - 6.0 ppm). b) Zona de azúcares y aminoácidos (5.5 - 1.0 ppm).

5.2.2. ANALISIS DE LOS ESPECTROS

Los espectros centrados y escalados se analizaron por PCA, mostrado en la Figura 10. En primera instancia se observa una separación a lo largo de los PC's 1 y 2 en un grupo bien definido del lado derecho correspondiente a los pacientes con nutrición enteral y a un grupo con mayor dispersión correspondiente a los pacientes con nutrición parenteral. El porcentaje de varianza R^2X con dos componentes es de 39.41%.

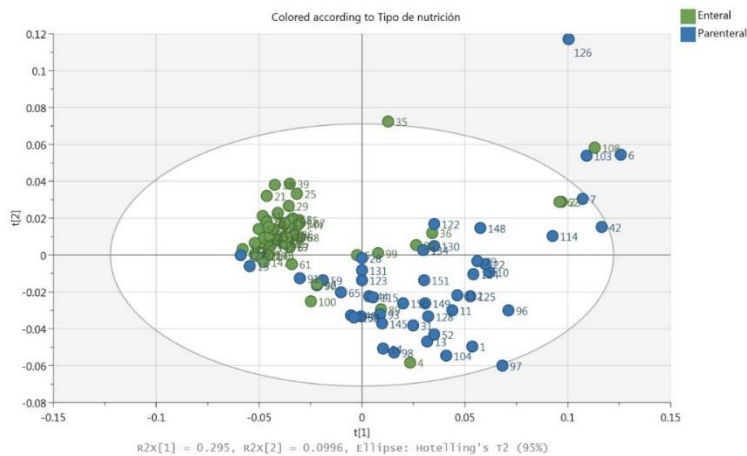


Figura 10. Score plot del PCA modelado con los datos espectrales de orina.

Las muestras 126, 108, 103 y 6 presentan la presencia de una señal ancha en 3.72 ppm, que no fue reconocida con el programa Chenomx. Se revisaron los datos para buscar alguna relación entre las muestras y se determinó que poseen comportamiento anómalo, por lo que después de una detenida revisión de la información, se decidió excluirlas del análisis. La Figura 11 muestra el score plot sin estos datos anómalos.

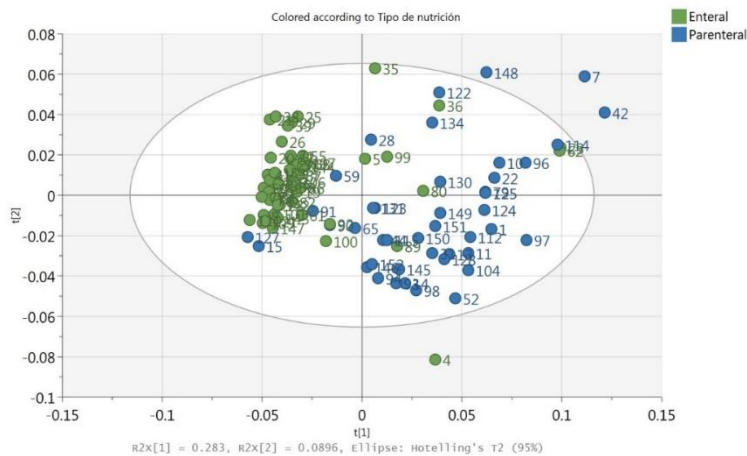


Figura 11. Score plot del PCA modelado con los datos espectrales de orina sin los anomalos 126, 108, 103 y 6.

La tabla 2 muestra los valores de R^2Y y Q^2 para cada modelo generado con las variables clasificatorias. Solo el tipo de nutrición brindó una clasificación aceptable.

Tabla 2. Varianza y predicción de los modelos de clasificación y regresión OPLS generados para cada factor.

Modelo	Tipo de modelo	No de componentes	R ² Y	Q ²
Edad gestacional	Clasificación	2+1+0	0.283	0.048
Peso	Clasificación	1+1+0	0.013	0.023
Tipo de nutrición	Clasificación	1+1+0	0.713	0.616
Edad gestacional corregida	Regresión	1+1+0	0.428	0.271
Peso a la toma de muestra	Regresión	1+1+0	0.407	0.225

El *score plot* del modelo OPLS-DA generado con el tipo de nutrición se muestra en la Figura 12. Se aprecia que el modelo genera una variable latente y una variable ortogonal, en la que las muestras provenientes de nutrición enteral se orientan a valores positivos sobre la variable latente, mientras que las muestras provenientes de nutrición parenteral se orientan en valores negativos sobre la misma variable.

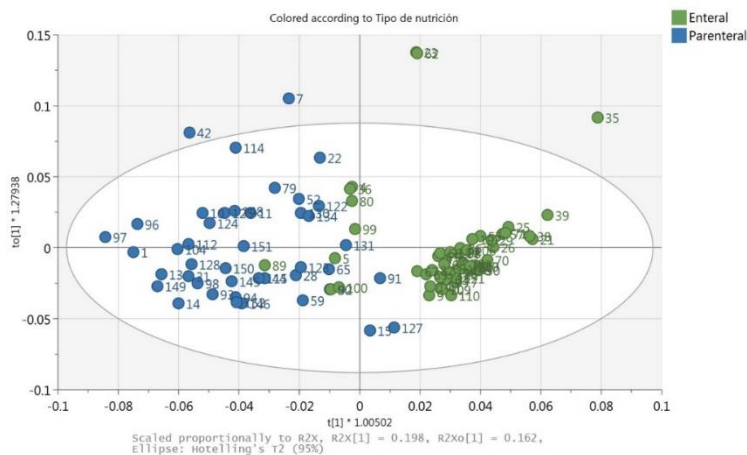


Figura 12. *Score plot* del modelo OPLS-DA clasificado contra el tipo de nutrición.

Tomando el MCN, se obtiene 90.7% de clasificaciones correctas al tipo de nutrición parenteral y 92.7% de clasificaciones correctas al tipo de nutrición enteral. Las variables que impactan en el modelo están asociadas a los metabolitos en la tabla 4. No todos los picos pudieron ser identificados y se indican como “NI” (No identificado).

Tabla 4. Metabolitos identificados con el programa Chenomx que generan diferencia entre los tipos de alimentación.

Bin (ppm)	Mayor concentración en	Metabolito
7.16, 6.84, 2.84, 1.92, 7.76, 3.76, 4.40	Parenteral	N-Acetil tirosina
4.96	Parenteral	NI
3.76	Parenteral	Manitol
4.04, 3.84	Parenteral	Gluconato
3.07	Parenteral	Creatinina
3.16	Enteral	Etanolamina
4.52	Enteral	Ascorbato
8.48	Enteral	Formiato
8.52	Enteral	NI
1.80	Enteral	NI
1.60	Enteral	NI
4.2	Enteral	4-Hidroxifenillactato
3.92	Enteral	Glucosa
3.32	Enteral	1,3-Dimetilurato
8.40	Enteral	NI
2.00	Enteral	NI
2.72	Enteral	Dimetilamina
2.92	Enteral	N,N-Dimetilglicina
8.44	Enteral	Formiato
2.36	Enteral	NI
2.28	Enteral	NI
7.86	Enteral	3-Indoxilsulfato
3.64	Enteral	NI
2.04	Enteral	NI
2.52, 2.64, 2.68, 2.56	Enteral	Citrato
3.28	Enteral	Betaína
3.60, 4.08	Enteral	Mio-Inositol
3.96, 4.44	Enteral	Lactosa
4.60	Enteral	Galactosa
4.32	Enteral	Prolina
2.40	Enteral	NI
2.32	Enteral	Succinato

NI No identificado

Los metabolitos cuya concentración impacta el comportamiento de las muestras de niños sometidos a nutrición parenteral en el modelo son la N-acetil tirosina, un metabolito no identificado (4.96 ppm) y el gluconato. Por otro lado, los metabolitos cuya concentración impacta el comportamiento de las muestras de niños sometidos a nutrición enteral son el succinato, un metabolito no identificado (2.40 ppm) y la prolina. Posteriormente se realizó cuantificación de los metabolitos para así comparar las concentraciones.

5.2.3. ANALISIS DE LA CONCENTRACIONES

Para comparar las concentraciones es necesario realizar normalización a un metabolito cuya excreción sea considerara una constante. En este caso se usó normalización a la

concentración de creatinina, que cumple la característica previamente mencionada. Para que esta normalización sea válida, es necesario justificar que no existe daño renal. Una de las muestras tuvo un área bajo la curva en las señales de creatinina mayor al doble que las demás (después de realizar normalización al total de área en el espectro). Es posible que los riñones de dicho paciente no estuvieran completamente desarrollados, viéndose afectada la excreción de creatinina, por lo que dicha muestra se eliminó del análisis.

Para explorar los datos se realizó un PCA, del que se retuvieron 2 PC's. La Figura 13.A muestra el *score plot* del modelo sin colorear. Coloreando las muestras en función del tipo de nutrición se apreció que la tendencia natural del agrupamiento en este gráfico está dada en función de dicha variable, como se muestra en la Figura 22.B.

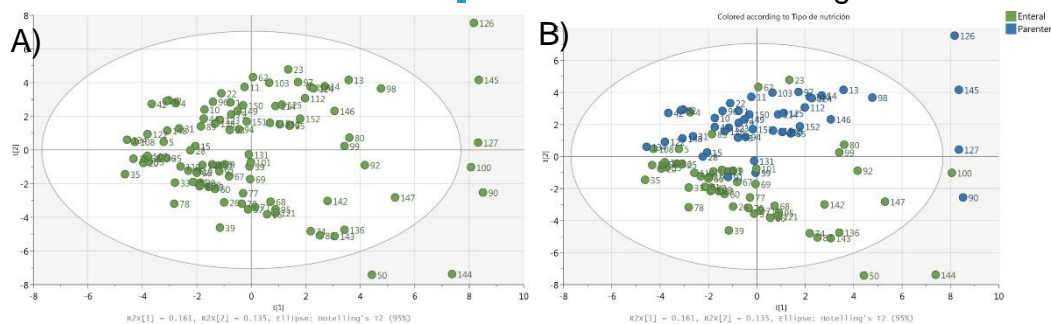


Figura 13. Score plot del PCA modelado con las cuantificaciones de los metabolitos en orina, sin colorear (A) y coloreado con el tipo de nutrición (B). Se nota que a la izquierda preferentemente se ubican las muestras de nutrición parenteral, mientras que a la derecha se ubican las muestras de nutrición enteral.

5.2.3.1. CLASIFICACIÓN POR TIPO DE NUTRICIÓN

La clasificación OPLS-DA por tipo de nutrición genera una variable latente y se presentan en la Figura 14, donde las muestras que provienen de nutrición enteral se muestran en color verde, mientras que las muestras que provienen de nutrición parenteral se muestran en color azul.

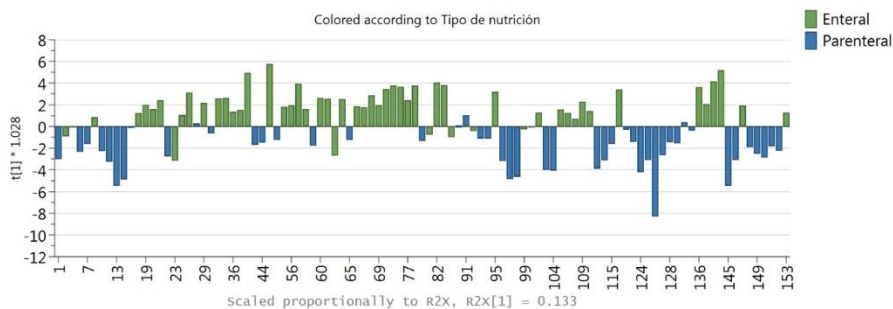


Figura 14. Score plot del modelo OPLS-DA clasificado por el tipo de alimentación

El MCN muestra 89.1% de clasificaciones correctas al tipo de nutrición enteral y 87.0% al tipo de nutrición parenteral. Al reducir la información, de un espectro completo a solo los metabolitos, se reduce también la capacidad predictiva. El análisis por permutaciones muestra que este modelo es válido tomando en cuenta las especificaciones anteriores ($R^2X= 0.0, 0.14$ y $Q^2= 0.0, -0.19$).

Los compuestos que correlacionan con la administración de nutrición enteral son succinato, treonato, succinato, 3-indoxilsulfato, trans-aconitato y galactosa, mientras que los que correlacionan con la administración de nutrición parenteral son N-

acetiltirosina, gluconato, 3-hidroxiisovalerato, manitol y glicina. Se realizó la prueba U de Mann-Whitney para comparar los valores de los cocientes de metabolitos a creatinina. El resumen estadístico se da en la tabla 5.

Tabla 5. Principales metabolitos que hacen diferentes las muestras de orina de niños con nutrición enteral o parenteral. Media aritmética y error estándar de los valores del cociente de las concentraciones del metabolito/creatinina.

Compuesto	Enteral (n=56)	Parenteral (n=46)	P<
Galactosa	0.6843 ± 0.0865	0.1539 ± 0.0333	0.000001
Treonato	0.5786 ± 0.0610	0.1299 ± 0.0327	0.000001
Succinato	0.6017 ± 0.0552	0.2445 ± 0.0354	0.000001
Trans-Aconitato	0.0298 ± 0.0039	0.0049 ± 0.0010	0.000001
3-indoxilsulfato	0.0919 ± 0.0154	0.0056 ± 0.0021	0.000001
Gluconato	1.0931 ± 0.3129	7.9903 ± 0.8432	0.000001
N-acetiltirosina	0.2213 ± 0.1059	1.8475 ± 0.2049	0.000001
Glicina	1.9032 ± 0.1490	3.9601 ± 0.3540	0.000001
3-Hidroxiisovalerato	0.0444 ± 0.0051	0.1160 ± 0.0150	0.000001
Manitol	1.4231 ± 0.5906	5.9745 ± 0.9958	0.01

5.2.3.2. OTROS CLASIFICADORES

Se decidió probar si existía correlación de los datos con los otros clasificadores, y se obtuvieron valores estadísticos multivariantes que se reportan en la tabla 6. En todos los casos hubo mejora en los porcentajes de regresión R²Y y Q² conforme a los valores obtenidos cuando se analizó la totalidad del espectro en forma de *bins*.

Tabla 6. Varianza y predicción de los modelos de clasificación y regresión OPLS generados para cada variable

Modelo	Tipo de modelo	No de componentes	R ² Y	Q ²
Edad gestacional	Clasificación	2+0+0	0.385	0.212
Peso	Clasificación	1+1+0	0.190	0.076
Tipo de nutrición	Clasificación	1+0+0	0.594	0.532
Edad gestacional corregida	Regresión	1+1+0	0.581	0.371
Peso a la toma de muestra	Regresión	1+1+0	0.584	0.392

5.2.3.3. CLASIFICACIÓN POR EDAD GESTACIONAL

El score plot en la Figura 24 muestra que el perfil bioquímico urinario de los recién nacidos prematuros extremos tiene una tendencia diferente del comportamiento de los otros grupos en la información de las variables latente 1 y 2, y a su vez los tardíos muestran una tendencia a orientarse a valores negativos de la variable latente 1, mientras que los muy prematuros se distribuyen a lo largo de la misma variable latente.

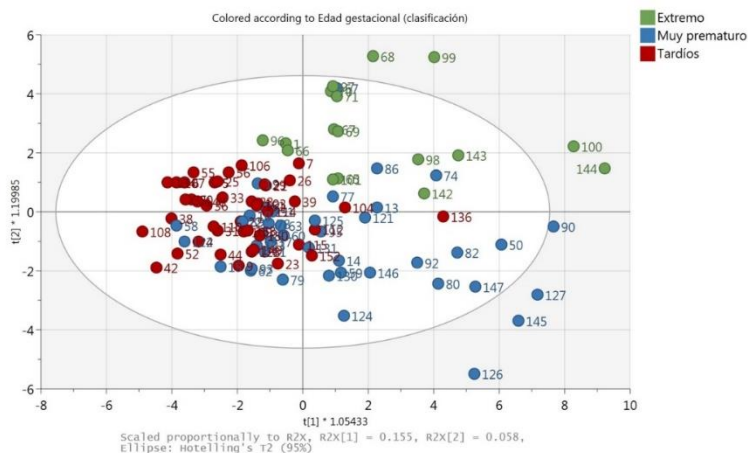


Figura 24. Score plot del modelo OPLS-DA clasificado por edad gestacional.

Este tipo de comportamiento sugirió que se puede construir una clasificación con dos clases donde no se diferencie entre los prematuros clasificados como muy prematuros y los tardíos. Los resultados se muestran en la Figura 25, donde la información ahora esta segregada a lo largo de la variable latente 1. Los valores de regresión y de predicción para este modelo son $R^2Y = 0.709$ y $Q^2 = 0.375$. Aunque no nos da fiabilidad de una predicción entre el tipo de prematuros, si nos habla acerca de la diferencia del desarrollo bioquímico entre los recién nacidos prematuros menores a 28 SDG y los mayores a 28 SDG.

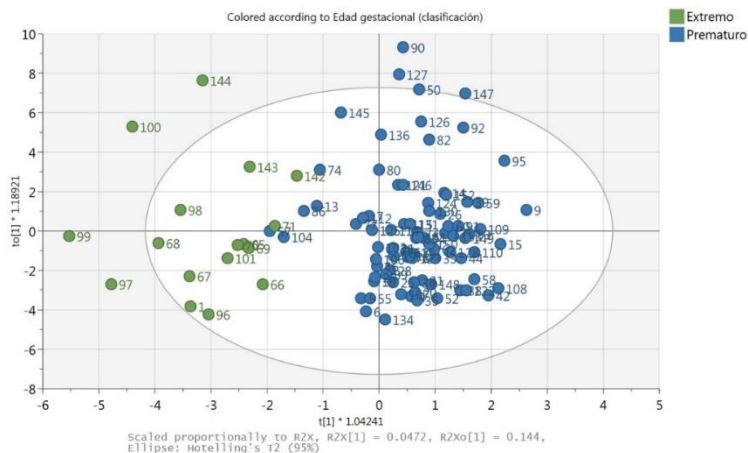


Figura 2515. Score plot del OPLS-DA clasificado por edad gestacional considerando dos clases: extremos y muy prematuros.

La validación obtenida por método de permutaciones evidencia que ambos modelos son válidos.

Por su parte, el valor de clasificaciones correctas a la clase extremos para el modelo con tres grupos es de 82.35%, el de muy prematuro es de 51.22% y el de tardíos es de 86.05%. Mientras, el modelo que clasifica a dos grupos tiene un valor de clasificados correctos a los extremos es de 94.12% y el de no extremos es de 98.81%.

Los metabolitos que hacen diferentes a los prematuros extremos de los prematuros no extremos son: dimetilamina, etanolamina, glucosa, transaconitato y betaína, que están presentes en mayor cantidad en la orina de los extremos, mientras que el 2-oxoglutarato, el 3-aminoisobutirato y el citrato se encuentran mayormente en la orina de

los no extremos. Estos metabolitos fueron sometidos a la prueba U de Mann-Whitney y el resumen se muestra en la tabla 8.

Tabla 8. Principales metabolitos que hacen diferentes las muestras de orina de niños prematuros extremos y no extremos. Media aritmética y error estándar de los valores del cociente de las concentraciones del metabolito/creatinina.

Compuesto	Extremos (n=17)	No extremos (n=85)	P<
Dimetilamina	0.4275 ± 0.0235	0.3039 ± 0.0091	0.00001
Etanolamina	0.7682 ± 4.1231	0.3643 ± 0.0323	0.0001
Glucosa	6.6057 ± 0.0087	2.0928 ± 0.3754	0.01
trans-Aconitato	0.0421 ± 0.0082	0.0139 ± 0.0022	0.001
Betaína	1.4648 ± 0.1439	0.9903 ± 0.0516	0.01
2-Oxoglutarato	0.0454 ± 0.0313	0.1134 ± 0.0194	0.05
3-Aminoisobutirato	0.0000 ± 0.0000	0.1459 ± 0.0317	0.05
Citrato	1.4854 ± 0.2459	1.7100 ± 0.1136	

Lo más notorio de estos resultados es que el 3-aminoisobutirato se encuentra ausente en la orina de los prematuros extremos. El isómero R es el principalmente excretado en la orina debido al metabolismo de la timina.²² También hay indicios de deficiencias en el metabolismo energético en los prematuros extremos, además de grandes cantidades de glucosa excretada, pudiendo ser síntoma de un exceso de glucosa en sangre, o un indicio de falla renal que permite la permeabilidad de éste compuesto.

Un intento de encontrar posibles diferencias entre los grupos Muy prematuros y Tardíos arroja que no se puede construir el modelo de clasificación con la información existente en la matriz, por lo que es mejor la consideración de estos dos grupos como uno solo en su desarrollo bioquímico.

5.2.3.4. CLASIFICACIÓN POR EDAD GESTACIONAL AJUSTADA Y PESO A LA TOMA DE MUESTRA

El primer modelo de correlación evaluó un parámetro obtenido y nombrado “edad gestacional ajustada”, que suma la edad gestacional a las semanas de vida extrauterina. El gráfico de variables observadas contra variables predichas se muestra en la Figura 26, y lo que se busca es que el valor de regresión sea lo más alto posible, aunado a que el comportamiento esté ajustado por una gráfica de correlación lineal. Para dicho modelo, el valor de correlación R es de 0.762.

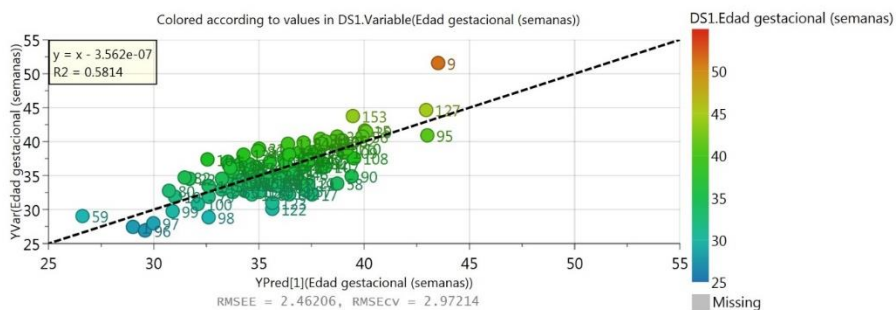


Figura 26. Gráfico de objetos observados contra objetos predichos conforme a edad gestacional ajustada (semanas).

²² Tanaguchi, K. Deficiency of D-b-aminoisobutyrate-pyruvate aminotransferase in the liver of genetic high excretors of D-b-aminoisobutyrate. *Biochem. Biophys. Acta.* **1972**, 279, 475-480.

Un segundo modelo de correlación evaluó el peso a la toma de muestra, y el valor de correlación R es de 0.764. Su gráfico se observaciones predichas contra observadas se muestra en la Figura 2716.

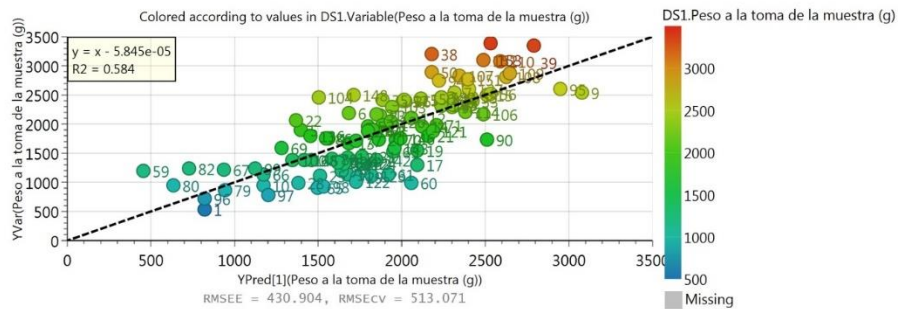


Figura 2716. Gráfico de objetos observados contra objetos predichos conforme al peso a la toma de muestra (g).

Estos resultados muestran que es posible realizar una estimación acerca de la edad gestacional corregida y del peso del infante a través de la totalidad de los metabolitos identificados y cuantificados en el espectro de RMN.

6. CONCLUSIONES

Se realizó el estudio del metaboloma urinario y sanguíneo del prematuro enfermo mexicano, identificando que el tipo de nutrición suministrada es el factor que mayoritariamente afecta el metaboloma. Otros clasificadores utilizados no tuvieron una correlación tan grande como la generada por el tipo de nutrición, pero si nos dan indicios del comportamiento bioquímico en el desarrollo de los prematuros.

En el caso del metaboloma del espectro sanguíneo se puede observar desde el análisis de componentes principales (PCA) que existe la tendencia marcada por el tipo de alimentación. Sin embargo, se considera que la metodología de extracción realizada no es la mejor para estudios de metabolómica donde la plataforma a utilizar sea la RMN.

En el caso del metaboloma urinario, los datos, tanto desplazamientos químicos como concentraciones normalizadas a la creatinina muestran fuerte correlación con el tipo de nutrición.

Con la identificación de los metabolitos y la determinación de sus concentraciones, es posible generar modelos de correlación que muestran tendencias en el *score plot*, aunque los coeficientes de predicción no mejoran respecto de lo que se obtuvieron con la información espectral.

En el metaboloma urinario también encontramos la correlación en el OPLS a los datos de peso a la toma de muestra (g) y edad gestacional ajustada (semanas) con coeficientes de correlación R de aproximadamente 0.70.

De esta manera podemos probar que características como el tipo de nutrición brindada al infante, el peso a la toma de muestra y la edad gestacional ajustada pueden ser modeladas a partir del metaboloma urinario. Es probable que el metaboloma sanguíneo también pueda correlacionar con las características estudiadas en este trabajo de investigación, por lo que repetir el ensayo, esta vez con extracción de los metabolitos en agua deuterada, es un interés particular del grupo.