

**ESTUDIO PRELIMINAR PARA EL DESARROLLO DE UN MÉTODO DE
DIAGNÓSTICO CLÍNICO DE CHAGAS A TRAVÉS DE LA DIFERENCIACIÓN
METABÓLICA ANALIZADA POR ESPECTROSCOPIA DE RESONANCIA
MAGNÉTICA NUCLEAR**

KEYLA MAYDEE ORTIZ ORTIZ

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE CIENCIAS
ESCUELA DE QUÍMICA
MAESTRÍA EN QUÍMICA
BUCARAMANGA
2018**

**ESTUDIO PRELIMINAR PARA EL DESARROLLO DE UN MÉTODO DE
DIAGNÓSTICO CLÍNICO DE CHAGAS A TRAVÉS DE LA DIFERENCIACIÓN
METABÓLICA ANALIZADA POR ESPECTROSCOPIA DE RESONANCIA
MAGNÉTICA NUCLEAR**

KEYLA MAYDEE ORTIZ ORTIZ

Tesis de grado para optar al título de Magíster en Química

Director:

JUAN CARLOS POVEDA JARAMILLO, Ph.D en Química.

Codirectora:

CLARA ISABEL GONZÁLEZ, Ph.D en Ciencias Biológicas

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE CIENCIAS
ESCUELA DE QUÍMICA
MAESTRÍA EN QUÍMICA
BUCARAMANGA
2018**

...A mi familia y a Sergio, por su amor, confianza y apoyo.

*“La dicha de la vida consiste en tener siempre algo que hacer,
alguien a quien amar y alguna cosa que esperar”
Thomas Chalmers*

AGRADECIMIENTOS

A mis directores, Juan Carlos Poveda y Clara Isabel Gonzáles, por la confianza, enseñanzas y apoyo durante el desarrollo de esta investigación.

Al laboratorio de Resonancia Magnética Nuclear de la UIS por facilitar los espacios y equipos necesarios para desarrollar el trabajo; al profesor Daniel Molina y Mary Helena por toda su colaboración.

A mis compañeros y amigos quienes hicieron este proceso más ameno.

A mi familia, por su apoyo.

A Sergio, por ser mi fortaleza.

TABLA DE CONTENIDO

INTRODUCCIÓN	16
1 MARCO TEÓRICO.....	19
1.1 TRIPANOSOMIASIS AMERICANA O ENFERMEDAD DE CHAGAS.....	19
1.2 CIENCIAS “ÓMICAS”	21
1.2.1 Metabolómica.....	21
1.3 ANÁLISIS QUIMIOMÉTRICO	22
1.3.1 Pre-procesado de los datos	23
1.3.2 Procesado de los datos.....	24
1.3.2 Análisis de componentes principales (PCA)	25
1.3.3 Análisis discriminante por mínimos cuadrados parciales (PLS-DA).....	25
1.3.4 Construcción del modelo y optimización	26
2 OBJETIVOS	31
2.1 GENERAL.....	31
2.2 ESPECÍFICOS.....	31
3 AJUSTE DE LOS PARÁMETROS EXPERIMENTALES PARA LA TOMA DE ESPECTROS DE SUEROS SANGUÍNEOS	32
3.1 INTRODUCCIÓN.....	32
3.2 METODOLOGÍA	34
3.2.1 Pacientes y recolección de muestras.....	34
3.2.2 Experimentos de RMN	35
3.3 RESULTADOS Y DISCUSIÓN	36
3.3.1 Selección de la secuencia de pulsos	36
3.3.2 Optimización de la secuencia de pulsos	40
3.4 CONCLUSIONES	43
4 CONSTRUCCIÓN DE UN MODELO ESTADÍSTICO PARA DETERMINAR LA	

DIFERENCIACIÓN METABÓLICA	44
4.1 INTRODUCCIÓN	44
4.2 METODOLOGÍA	45
4.2.1 Población de estudio	45
4.2.2 Pre-procesado de los espectros	45
4.2.3 Análisis estadístico multivariado	47
4.3 RESULTADOS Y DISCUSIÓN	47
4.3.1 Selección de muestras	47
4.3.2 Análisis de componentes principales PCA	49
4.3.3. Análisis Discriminante por mínimos cuadrados parciales	54
4.4 CONCLUSIONES	64
5 IDENTIFICACIÓN DE LOS POSIBLES METABOLITOS CON POTENCIAL COMO BIOMARCADORES PARA MÉTODOS DE DIAGNÓSTICO CLÍNICO DE INFECCIÓN POR T. cruzi	65
5.1 INTRODUCCIÓN	65
5.2 METODOLOGÍA	66
5.3 RESULTADOS Y DISCUSIÓN	66
5.4 CONCLUSIÓN	75
6 CONCLUSIÓN GENERAL	76
7 RECOMENDACIONES	77
8 ASPECTOS ÉTICOS	78
REFERENCIAS BIBLIOGRÁFICAS	81
ANEXOS	93

LISTA DE FIGURAS

Figura 1. Secuencias de pulsos zgpr, NOESY y CPMG.	35
Figura 2. Espectros obtenidos con las secuencias zgpr, CPMG, NOESY. Elaboración propia.	37
Figura 3. Diagrama esquemático de la relación entre la relajación longitudinal T_1 , la relajación transversal T_2 y el tiempo de correlación molecular τ_c . Adaptado de: (KESHARI; WILSON, 2014).....	39
Figura 4. Diagrama de la secuencia de pulsos CPMG. Adaptado de (MEIBOOM; GILL, 1958).....	40
Figura 5. Ensayos de reproducibilidad utilizando la secuencia CPMG, con único suero.....	41
Figura 6. Espectros obtenidos con la secuencia CPMG con $n=100$ y para diferentes valores de τ	42
Figura 7. Espectros obtenidos con la secuencia CPMG con $\tau= 0.0005$ y diferentes valores de n	42
Figura 8. Región espectral de 0.5 -1.6 ppm antes del alineamiento. Elaboración propia.....	45
Figura 9. Región espectral de 0.5 -1.6 ppm después del alineamiento. Elaboración propia.....	46
Figura 10 Segmentación de los espectros utilizando Bucketing inteligente entre la región de 0.5-9 ppm. Elaboración propia.	46
Figura 11. Distribución normal. Izquierda: Antes del auto-escalado. Derecha: Después del escalado. Elaboración propia.....	47
Figura 12. Comparación espectros de ^1H -NMR de suero sanguíneo muestra 38 con posible contaminación y muestra 14 con señales características del suero sanguíneo. Elaboración propia.	48

Figura 13. Espectro de ^1H -NMR de suero sanguíneo de la muestra 164 con señales atípicas comparada con la muestra 14 que presenta las señales características. Elaboración propia.	49
Figura 14. Score plots. Izquierda: antes de realizar el filtrado de datos. Derecha: después de retirar los outliers. Elaboración propia.	50
Figura 15. Comparación espectros RMN-H de la muestra 80 identificada como outlier con dos muestras (14 y 2247) elegidas al azar que se encuentran dentro del intervalo de confianza. Elaboración propia.	51
Figura 16. Comparación espectros RMN- ^1H de la muestra 2808 identificada como outlier con dos muestras 154 y 59 elegidas al azar que se encuentran dentro del intervalo de confianza. Elaboración propia.	52
Figura 17. Score plot del análisis PCA utilizando las etiquetas de acuerdo con el género. Elaboración propia.	53
Figura 18. Score plot del análisis PCA utilizando las etiquetas de acuerdo con la edad. Elaboración propia.	53
Figura 19. Estadístico VIP obtenido para las variables predictor a partir del modelo Mssv. Elaboración propia.	56
Figura 20 Error de clasificación para el modelo Mssv obtenido a través de las 8 variables latentes. Elaboración propia.	56
Figura 21. Score plot obtenido para el modelo $M_{\text{SME-VIP}}$. Elaboración propia.	58
Figura 22. Evaluación del conjunto test set en el modelo $M_{\text{SME-VIP}}$. Elaboración propia.	60
Figura 23. Asignación de metabolitos para un espectro de RMN- ^1H obtenido para una muestra de suero sanguíneo con diagnóstico seropositivo. Izquierda) Región aromática. Derecha) Región alifática. Elaboración propia.	69
Figura 24 Asignación de metabolitos para un espectro de RMN- ^1H obtenido para una muestra de suero sanguíneo con diagnóstico seronegativo. Izquierda) Región	

aromática. Derecha) Región alifática	70
Figura 25. comparación score plots inicial. Arriba) R, factorRmine abajo) Metaboanalyst. Elaboración propia.....	96
Figura 26.. comparación score plots después de eliminar outliers. arriba) R, factorRmine abajo) Metaboanalyst. Elaboración propia.....	97
Figura 27. Comparación scree plots Arriba) Metaboanalyst Abajo) R, factorRmine abajo. Elaboración propia.	98
Figura 28. Test de permutación para el modelo $M_{SME-VIP}$ obtenido con la construcción de 1000 modelos a partir de la permutación aleatoria de la etiqueta de clase en la matriz de trabajo. Elaboración propia.	99
Figura 29. Curva ROC para el modelo $M_{SME-VIP}$. Elaboración propia.....	100

LISTA DE TABLAS

Tabla 1. Matriz de confusión. Elaboración propia	28
Tabla 2. Cálculos de T_2 para los experimentos realizados. Elaboración propia.	43
Tabla 3. Muestras empleadas para realizar el análisis supervisado PLS-DA. Elaboración propia.	54
Tabla 4. Matriz de confusión obtenida para el modelo Mssv. Elaboración propia.	55
Tabla 5. Parámetros de calidad obtenidos para el modelo Mssv. Elaboración propia.	55
Tabla 6. Parámetros de calidad para cada uno de los modelos evaluados. Elaboración propia.	57
Tabla 7. Regiones responsables de la discriminación en el modelo $M_{SME-VIP}$. Elaboración propia.	59
Tabla 8. Estudios de enfermedades a través de la diferenciación metabólica en suero sanguíneo utilizando RMN y modelado PLS-DA. Elaboración propia.	61
Tabla 9. Comparación con otros métodos de diagnóstico de la EC. Elaboración propia.	63
Tabla 10. Metabolitos identificados en los espectros de RMN-H de sueros sanguíneos. Elaboración propia.	68

LISTA DE ANEXOS

Anexo 1. Carta de autorización para el uso de las muestras por parte del investigador principal.	93
Anexo 2. Aval comité de ética de la Universidad Industrial de Santander	94
Anexo 3 Comparación de los resultados obtenidos para el análisis de componentes principales con el software Metaboanalyst y el software R.	96
Anexo 4. Test de permutación y curva ROC.....	99

RESUMEN

TITULO: ESTUDIO PRELIMINAR PARA EL DESARROLLO DE UN MÉTODO DE DIAGNÓSTICO CLÍNICO DE CHAGAS A TRAVÉS DE LA DIFERENCIACIÓN METABÓLICA ANALIZADA POR ESPECTROSCOPIA DE RESONANCIA MAGNÉTICA NUCLEAR*

AUTOR: KEYLA MAYDEE ORTIZ**

PALABRAS CLAVES: Enfermedad de Chagas, Metabólica, RMN.

La enfermedad de Chagas (EC) es causada por el parásito *Trypanosoma cruzi*. Esta enfermedad representa una causa importante de morbilidad y mortalidad en los países endémicos, afecta aproximadamente a 8 millones de personas en todo el mundo. En Colombia, la prevalencia de infección por *T. cruzi* es de aproximadamente 5%, que corresponde a 700,000 personas, y en algunas áreas del departamento de Santander la seroprevalencia es de aproximadamente 50%. La EC incluye una fase aguda y crónica, que presenta un amplio espectro de manifestaciones clínicas que incluyen formas cardíaca, digestiva y neurológica. El diagnóstico de la infección por *T. cruzi* se realiza mediante pruebas serológicas y parasitológicas, que se utilizan según la fase clínica. El diagnóstico es complejo, especialmente durante la fase crónica, debido a la falta de síntomas y la parasitemia baja o intermitente que conduce a métodos parasitológicos directos que tienen una baja sensibilidad. El objetivo de este trabajo fue realizar un estudio metabólico en suero sanguíneo de pacientes con la enfermedad de Chagas y su comparación con un grupo control, empleando Resonancia Magnética Nuclear de Hidrógeno como herramienta analítica. Posteriormente, se realizó un análisis multivariado PLS-DA con el propósito de desarrollar un modelo matemático que permitiera discriminar los pacientes seropositivos de pacientes seronegativos. El modelo obtenido permite clasificar los pacientes con un 94% de precisión, una sensibilidad del 97% y una especificidad del 91%. Finalmente, se asignaron los posibles metabolitos con potencial de biomarcadores, entre ellos, glucosa, glutamato, lactato, formiato y treonina.

* Tesis de maestría

** Universidad Industrial de Santander. Facultad de ciencias. Escuela de química. Director: Juan Carlos Poveda Jaramillo. Facultad de salud. Escuela de Microbiología. Codirectora: Clara Isabel González Rugeles.

ABSTRACT

TITLE: PRELIMINARY STUDY FOR THE DEVELOPMENT OF A CLINICAL DIAGNOSTIC METHOD OF CHAGAS THROUGH THE METABOLIC DIFFERENTIATION ANALYZED BY NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY *

AUTHOR: KEYLA MAYDEE ORTIZ **

KEY WORDS: Chagas disease, Metabolomics, NMR.

Chagas disease (CD) is caused by the *Trypanosoma cruzi* parasite. This disease represents an important cause of morbidity and mortality in endemic countries, affecting approximately 8 million people worldwide. In Colombia, the prevalence of *T. cruzi* infection is approximately 5%, which corresponds to 700,000 people, and in some areas of the department of Santander the seroprevalence is approximately 50%. The CD includes an acute and chronic phase, which presents a wide spectrum of clinical manifestations that include cardiac, digestive and neurological forms. The diagnosis of *T. cruzi* infection is made by serological and parasitological tests, which are used according to the clinical phase. The diagnosis is complex, especially during the chronic phase, due to the lack of symptoms and the low or intermittent parasitemia that leads to direct parasitological methods that have a low sensitivity. The aim of this work was to perform a metabolomic study in blood serum of patients with Chagas disease and its comparison with a control group, using Nuclear Magnetic Resonance of Hydrogen as an analytical tool. Subsequently, a multivariate analysis was carried out with the purpose of developing a mathematical model that would allow to discriminate seropositive patients from seronegative patients. The PLS-DA analysis allowed the development of a model which allows classifying patients with 94% accuracy, a sensitivity of 97% and a specificity of 91. Finally, possible metabolites with biomarker potential were assigned, among them glucose, glutamate, lactate, formate and threonine.

* Master's thesis

** Universidad Industrial de Santander. Facultad de ciencias. Escuela de química. Director: Juan Carlos Poveda Jaramillo. Facultad de salud. Escuela de Microbiología. Codirectora: Clara Isabel González Rugeles.

INTRODUCCIÓN

La tripanosomiasis americana o enfermedad de Chagas (EC) es una infección causada por el protozoo *Trypanosoma cruzi*. El *T. cruzi* es transmitido principalmente a través de deyecciones de insectos triatomíneos hematófagos, que contienen tripanosomas, los cuales entran en la piel por excoriaciones en el sitio de la picadura o a través de las mucosas (MANTILLA, 2011). Estos insectos se encuentran en ecosistemas naturales desde el sur de Estados Unidos hasta Sur América; también se puede transmitir durante intervenciones clínicas como transfusiones sanguíneas, donación de órganos y de la madre al niño al nacer (PEREIRA NUNES et al., 2013; ROCA SAUMELL et al., 2015). La EC es una de las enfermedades desatendidas más importantes de América Latina. Según las estimaciones de la Organización Mundial de la Salud (OMS) en el año 2010 cerca de 6 millones de personas se encontraban infectadas por *T. cruzi* en América Latina y el 13% de la población latinoamericana está en riesgo de adquirir la infección (MOLINA; SALVADOR, 2016; REYES et al., 2017). Se estima que en Colombia 4,7 millones de personas se encuentran en riesgo de adquirir la infección con una prevalencia del 5% (CASTELLANOS et al., 2014; MANRIQUE-ABRIL et al., 2013; OPS, 2006; ROSAS, 2011). Los departamentos con las mayores tasas de infestación son Arauca: 21,1%, Casanare: 10%, Santander: 6,3%, Norte de Santander: 5,2%, Boyacá: 3,7%, Cundinamarca: 1,9% y Meta: 1,7% (MONCAYO; SILVEIRA, 2009).

Actualmente, el diagnóstico de la EC se hace mediante pruebas serológicas y parasitológicas, dependiendo de la fase de la infección. En la fase aguda el diagnóstico se hace con pruebas parasitológicas. En la fase crónica, la parasitemia es baja o indetectable, por lo que el diagnóstico se lleva a cabo con ensayos basados en la detección de anticuerpos anti *T. cruzi* mediante ensayos serológicos. Generalmente, la prueba de ELISA (*Enzyme Linked Immunosorbent Assay*), la inmunofluorescencia indirecta (IFI) y hemoaglutinación indirecta. Estas pruebas

serológicas han demostrado ser sensibles, y prácticas para la detección de esta enfermedad, pero no reproducibles, y en algunas ocasiones presentan reactividad cruzada por anticuerpos producidos por otros patógenos, principalmente por *Leishmania sp* y *T. rangeli* (CASTELLANOS et al., 2014; FERRER et al., 2012; PONCE et al., 2005). La OMS recomienda realizar dos pruebas serológicas diferentes y los resultados deberán ser concordantes, en caso de discordancia deberán repetirse los ensayos con una nueva muestra de suero, si la discordancia persiste deberá hacerse una prueba de confirmación tal como PCR o “western blotting”. Sin embargo, un resultado positivo de PCR confirma la infección, pero un resultado negativo no excluye la infección (BRASIL et al., 2010; MOLINA; SALVADOR; SÁNCHEZ-MONTALVÁ, 2016; MOURE et al., 2016). Por otro lado, Western blotting puede reconocer diferentes fracciones de antígenos del parásito, pero es una técnica costosa, compleja, y requiere de personal calificado para realizarse (MOURE et al., 2016). Hasta ahora, no hay un método que proporcione un diagnóstico preciso y definitivo para pacientes con resultados serológicos discordantes (MOURE et al., 2016).

Por otra parte, la metabolómica permite comprender simultánea y sistemáticamente el perfil de la concentración de múltiples metabolitos y su fluctuación relacionados con perturbaciones genéticas, ambientales o en respuesta a un estímulo (BECKONERT et al., 2007; SMOLINSKA et al., 2012). La metabolómica es una disciplina emergente y de gran alcance que se ha convertido en una herramienta prometedora en el campo del diagnóstico clínico de diversas enfermedades (ZHANG et al., 2012). La metabolómica emplea diferentes técnicas de análisis químico para la obtención del perfil metabólico, entre ellas se destaca la espectroscopía de Resonancia Magnética Nuclear (RMN) ya que ofrece grandes ventajas dado que no requiere de una preparación rigurosa de la muestra, permite identificar y cuantificar diferentes analitos simultáneamente (SERKOVA; NIEMANN, 2006).

La metabolómica se ha empleado en diversas áreas de la investigación biomédica, tal como, estudios de toxicología, efectos nutricionales, consecuencias metabólicas de modificaciones genéticas, errores innatos del metabolismo, diagnóstico de enfermedades como el cáncer, enfermedades neurológicas, entre otras (MADSEN; LUNDSTEDT; TRYGG, 2010; SIMMLER et al., 2014; SMOLINSKA et al., 2012). En particular, la metabolómica basada en RMN como herramienta de diagnóstico clínico proporciona información para diferenciar entre pacientes que pueden padecer una enfermedad y aquellos que no la padecen, o entre grupos tratados con fármacos y no tratados (BLASCO et al., 2010; TIZIANI; LOPES; GÜNTHER, 2009). Principalmente, es posible el diagnóstico y seguimiento de individuos enfermos, a través de biofluidos, tales como, orina, plasma y suero sanguíneo dado que estas muestras son fáciles de obtener, no son clínicamente invasivas, y pueden ser analizadas sin etapas adicionales de derivatización o concentración (SINCLAIR et al., 2010). La RMN de biofluidos genera rápidamente un gran número de datos, el análisis de la información espectral obtenida se combina con métodos quimiométricos para extraer, describir las diferencias y evaluar los cambios globales en grandes conjuntos de metabolitos en grupos de individuos (SIMMLER et al., 2014; VIANT, 2003). Debido a la baja reproducibilidad de los métodos clínicos convencionales empleados en la detección de infección por *T. cruzi* y la necesidad de desarrollar métodos de diagnóstico efectivos, en este trabajo de investigación se realizó un estudio exploratorio en la diferenciación metabólica de suero sanguíneo a través del análisis metabolómico basado en RMN para identificar metabolitos como posibles bioindicadores de la EC.

Este documento se encuentra organizado por capítulos, el primer capítulo describe los fundamentos teóricos, el segundo capítulo los objetivos, los capítulos tercero, cuarto y quinto describe la importancia, metodología, resultados y conclusiones obtenidas a partir de cada objetivo específico. El sexto capítulo contiene las recomendaciones. Finalmente, el capítulo siete contiene la descripción de los aspectos éticos tenidos en cuenta durante el desarrollo de la investigación.

1 MARCO TEÓRICO

1.1 TRIPANOSOMIASIS AMERICANA O ENFERMEDAD DE CHAGAS

La tripanosomiasis americana o EC es causada por el protozoo *Trypanosoma cruzi* descubierto en 1909 por el médico brasileño Carlos Chagas. La EC ha sido un problema para la humanidad desde la antigüedad, y sigue siendo un problema social y económico en muchos países latinoamericanos, debido a que es una infección de por vida. La EC es transmitida al humano y otras especies animales a través del insecto vector, por transmisión vertical, por transmisión oral, por trasplante de órganos contaminados y/o transfusión de hemoderivados contaminados (MOLINA; SALVADOR, 2016; RASSI; RASSI; MARIN-NETO, 2010). *T. cruzi* tiene un ciclo de vida complejo caracterizado por varias formas de desarrollo presentes en huéspedes de vertebrados e invertebrados. En el huésped vertebrado el parásito presenta una forma “tripomastigote” no replicativa que infecta el huésped e ingresa a las células cerca al sitio de la infección. Posteriormente, los tripomastigotes se diferencian en la forma de amastigotes intracelulares. En consecuencia, se multiplican por fisión binaria en las células de los tejidos infectados, finalmente la célula se rompe, y los amastigotes se transforman en tripomastigotes sanguíneos que ingresan al torrente sanguíneo para infectar nuevas células. Con el desarrollo de la respuesta inmune, la parasitemia se reduce y el número de parásitos en los tejidos disminuye sustancialmente. Sin embargo, dado que el parásito no se elimina por completo, la infección de tejidos específicos, como los músculos o los ganglios entéricos, persiste indefinidamente durante la vida del huésped (RASSI; RASSI; MARIN-NETO, 2010).

La EC se presenta en dos fases la aguda y crónica. Se considera fase aguda desde la infección hasta que la parasitemia detectada microscópicamente es negativa. La fase aguda generalmente es asintomática y en algunos casos presenta síntomas leves e inespecíficos que se asemejan a un cuadro gripal con fiebre, malestar

general, hepatoesplenomegalia (RASSI; RASSI; MARIN-NETO, 2010). En la mayor parte de los casos la fase aguda pasa desapercibida. La muerte en la fase aguda es baja, ocurre principalmente por miocarditis o meningoencefalitis (MOLINA; SALVADOR; SÁNCHEZ-MONTALVÁ, 2016; RASSI; RASSI; MARIN-NETO, 2010).

La fase crónica se inicia cuando la parasitemia detectada microscópicamente es negativa y las pruebas serológicas son positivas, lo que ocurre aproximadamente 1-2 meses después de la infección. Durante la fase crónica la mayor parte de los pacientes están asintomáticos, lo que se conoce como forma indeterminada. Sin embargo, un porcentaje variable de pacientes desarrollan complicaciones décadas después de la infección, predominantemente con afectación cardíaca, digestiva o mixta (BERN, 2015; MOLINA; SALVADOR; SÁNCHEZ-MONTALVÁ, 2016).

La patogenia de la cardiopatía todavía no está bien esclarecida. Una respuesta inmunológica descontrolada y el constante daño directo del parásito producen un daño microvascular, en los miocardiocitos y en el tejido nervioso que terminan causando trastornos del ritmo, zonas de fibrosis y dilataciones cardíacas. Por otro lado, la EC puede producir una miocardiopatía dilatada que desencadenará una insuficiencia cardíaca progresiva. La mortalidad asociada a la cardiopatía chagásica es de aproximadamente el 3,9% al año, pero depende de la severidad de la cardiopatía. Por regla general, la mortalidad de la miocardiopatía dilatada de origen chagásico es mayor que la de la miocardiopatía dilatada de otra etiología (MOLINA; SALVADOR, 2016). A nivel gastrointestinal, la EC puede afectar a cualquier parte del tubo digestivo, siendo las localizaciones más afectadas el colon y el esófago. La frecuencia con la que produce afectación digestiva varía entre el 5 y 35%, al igual que la afectación cardíaca. Otra complicación asociada con la EC es el aumento del riesgo de padecer eventos tromboembólicos que ocurre principalmente por la formación de trombos en los pacientes con miocardiopatía dilatada, aneurismas o fibrilación auricular (MOLINA; SALVADOR, 2016).

1.2 CIENCIAS “ÓMICAS”

Las ciencias “ómicas” pretenden proporcionar una visión global e integradora de los procesos biológicos, encontrando relaciones causales entre los elementos que componen el sistema biológico como los genes, RNAs, proteínas y metabolitos, estas disciplinas se denominan genómica, transcriptómica, proteómica y metabolómica, respectivamente. Estas disciplinas permiten generar y analizar datos a gran escala con elevada precisión y resolución en donde se requiere de potentes técnicas computacionales para la evaluación e interpretación de los mismos (UNGER, 2014). Los estudios “ómicos” además de proporcionar una visión holística de los sistemas biológicos y sus alteraciones en respuesta a la enfermedad, facilitan el descubrimiento de biomarcadores de aplicación clínica y también contribuyen al conocimiento de los mecanismos moleculares implicados en la etiopatogenia de las enfermedades.

1.2.1 Metabolómica La metabolómica es la evaluación exhaustiva, la identificación sistemática y la cuantificación de metabolitos dentro de un sistema biológico (ENEA et al., 2010). La metabolómica es un campo interdisciplinar de la ciencia que combina técnicas de química analítica, plataformas tecnológicas y un sofisticado análisis de datos para la elucidación metabólica (WEN et al., 2010; ZHANG et al., 2013). Debido a que los metabolitos cumplen un rol importante en los organismos biológicos como indicadores de estados fisiológicos o patológicos, son candidatos atractivos para el estudio de enfermedades (ZHANG et al., 2013). Este enfoque abre un sin número de posibilidades de investigación en diversas áreas, incluyendo el descubrimiento de biomarcadores, estudios clínicos, eficacia de fármacos, toxicidad, métodos de diagnóstico y control de patologías.

La metabolómica presenta dos tipos de enfoques analíticos diferentes, “*Metabolomic profiling*” o “perfil metabólico” donde se realiza un análisis (identificación y cuantificación) de una clase de metabolitos predefinidos

(metabolitos químicamente relacionados o relacionados con una vía metabólica en particular) y “*Metabolomic fingerprinting*” o “huella dactilar” que se basa en el análisis del conjunto total de metabolitos, evitando prejuicios contra ciertas clases de compuestos, para la clasificación rápida de muestras. Este análisis ofrece una clasificación sin identificar y cuantificar metabolitos individuales, para la clasificación rápida de muestras. Ambos enfoques metabólicos se utilizan en la búsqueda de nuevos biomarcadores que cambian como un indicador de la presencia de una enfermedad (DUNN; ELLIS, 2005).

Metabolomic fingerprinting es el método más comúnmente utilizado en la metabolómica de RMN y se basa en el análisis multivariante de grupos de espectros de RMN completos, donde cada espectro se puede considerar una 'huella dactilar' de señales no asignadas. El objetivo típico de dicho análisis es clasificar grupos de muestras de acuerdo con similitudes y / o diferencias entre sus huellas dactilares de RMN. Una vez identificadas las diferencias entre los espectros se usan métodos adicionales en un intento para asignar esas líneas dentro de la huella digital de RMN que son responsables de la discriminación. La toma de huellas dactilares por RMN se puede dividir en tres fases:

- Adquisición y procesamiento espectral
- Análisis multivariable
- Identificación de biomarcadores.

1.3 ANÁLISIS QUIMIOMÉTRICO

La quimiometría se define como el uso de métodos matemáticos y estadísticos para mejorar la comprensión de la información química contenida en los espectros de mezclas complejas. Generalmente, un análisis quimiométrico involucra la medición de múltiples variables sobre un conjunto de sujetos de estudio, lo que genera un conjunto complejo de datos que requiere de un adecuado tratamiento estadístico para obtener información útil y significativa a partir de los mismos. Para ello se

realiza un pre-procesado de los datos y la elección de los métodos de análisis multivariado apropiado para los datos que desean analizar.

1.3.1 Pre-procesado de los datos Previo al análisis de datos, los espectros de suero sanguíneo adquiridos con RMN reciben un pre-tratamiento con el objetivo de eliminar y/o reducir las variaciones no deseadas que puedan interferir en el análisis posterior. Durante el pre-procesado se realiza: la corrección de la fase y línea base, para eliminar las distorsiones que puedan afectar el análisis de los datos; la calibración de los espectros con respecto al patrón interno (TSP) para corregir las diferencias intrínseca de cada una de las muestras (SMOLINSKA et al., 2012).

Dado que el desplazamiento químico de las señales es muy sensible al entorno químico de los núcleos (^1H), su localización en el espectro se ve afectado por cambios de pH, temperatura, concentración salina y homogeneidad del campo (*"shimming"*), que se traduce en variaciones significativas en los desplazamientos químicos de las señales. Si bien durante el pre-procesado se realiza un alineamiento global calibrando todo el espectro con respecto a una señal de referencia (TSP), no todas las señales se ven afectadas de la misma manera, ni fluctúan o se desplazan en el mismo sentido, por tanto, se hace necesario realizar un mejor alineamiento. Existen diferentes algoritmos para realizar el alineamiento, entre ellos, el alineamiento por regiones, este método permite alinear regiones aisladas del espectro tomando como referencia un espectro representativo. El alineamiento de las regiones seleccionadas se lleva a cabo sin modificar la posición del resto de las señales del espectro. Posteriormente, se lleva a cabo la exclusión de las regiones que no formarán parte del análisis, tal como la señal residual del agua cuya información no es reproducible y la señal del TSP utilizado como patrón de campo, debido a que esta señal no contiene señales correspondientes a metabolitos

1.3.2 Procesado de los datos El procesado de los datos consiste en obtener la matriz de datos y tratamiento de los mismos. Para ello se realiza una segmentación del espectro o “*Bucketing*” que consiste en subdividir los espectros en pequeñas regiones llamadas “*buckets*” las cuales se integran. Como resultado del *bucketing* se obtiene la matriz de datos $n \times p$, donde n es el número de muestras analizadas y p es el número de *buckets* en los que se han dividido los espectros y que constituyen las variables descriptoras. El método de *bucketing* puede ser de ancho fijo o de ancho variable, en los métodos de anchura fija todos los *buckets* tienen el mismo ancho, generalmente 0,04 ppm, por otra parte los métodos de anchura variable, donde las regiones se definen de acuerdo a la señal de interés a la cual se le está realizando el seguimiento. Uno de los métodos más usados es el método adaptivo e inteligente que consiste en realizar una cubicación inteligente, en dichos métodos, el ancho de cada cubo se determina luego por la diferencia máxima de desplazamientos químicos entre todos los espectros (De Meyer et al., 2008).

Debido a que se pretende identificar cambios metabólicos relativos como consecuencia de una determinada condición fisiopatológica independiente del estado de dilución de la muestra, se realiza una normalización de los datos con el objeto de compensar las diferencias de concentración entre las muestras (DIETERLE et al., 2006). Entre los métodos de normalización se encuentra la normalización integral, que es la más utilizada. En este método se asume que la concentración de cada metabolito es función de la concentración global de la muestra y se divide cada región de integración (*bucket*) entre el área total del espectro, con esta matriz de datos normalizada se procede a realizar el análisis quimiométrico. Sin embargo, incluso cuando todas las variables tengan unidades idénticas, las intensidades y variaciones muy dispares entre las variables obligarán a la mayoría de las formas de análisis multivariante a centrarse en un pequeño conjunto de señales intensas. Por estas razón, las variables pueden escalarse automáticamente con el objeto de que las variables sean comparables (WORLEY; POWERS, 2013), el autoescalado consiste en centrar la variable y dividirla por la

desviación estándar, es decir, hacer que el promedio sea cero y la desviación estándar 1, de acuerdo con la Ecuación 1:

$$\frac{x_i - \text{promedio}(x)}{sd(x)}$$

Ecuación 1

Posteriormente, una vez las variables han sido normalizadas y escaladas se procede con el análisis multivariado.

1.3.2 Análisis de componentes principales (PCA) convierte los datos de un espacio multidimensional a un espacio de menor dimensión. Esta técnica expresa la mayor parte de la varianza dentro de un conjunto de datos utilizando un número menor de factores, así llamado componentes principales. Cada componente principal es una combinación lineal de los variables originales, y se van construyendo según el orden de importancia en cuanto a la variabilidad total que recogen de la muestra. Los componentes principales son ortogonales entre sí, por tanto, exhiben diferente información. La variación espectral es descrita por pocos componentes principales, comparado con el número de variables originales. PCA permite encontrar tendencias, agrupaciones y encontrar valores atípicos (NAGANA GOWDA; RAFTERY, 2015).

1.3.3 Análisis discriminante por mínimos cuadrados parciales (PLS-DA) Es un método de regresión multivariable que considera dos grupos de variables, independientes (matriz X) y dependientes (matriz Y), permitiendo correlacionar los dos conjuntos de datos entre sí, o la predicción de valor de un conjunto mediante el uso de los datos desde el otro conjunto que es más fácil de medir.

La regresión PLS consta de dos procesos, inicialmente se transforma la matriz de predictores X de orden $n \times p$, con ayuda del vector de respuestas Y de orden $n \times 1$,

en una matriz de componentes o variables latentes no correlacionados, $T = (T_1, \dots, T_p)$ de orden $n \times p$, llamados componentes PLS o variables latentes; esto contrasta con el análisis de componentes principales en el cual los componentes son obtenidos usando sólo la matriz de variables predictoras X . Luego, se calcula el modelo de regresión estimado usando el vector de respuestas original y la matriz de componentes PLS. La reducción de la dimensionalidad puede ser aplicada directamente sobre los componentes ya que estos son ortogonales. El número de componentes necesarios para el análisis de regresión debe ser mucho menor que el número de predictores (WOLD; SJÖSTRÖM; ERIKSSON, 2001).

1.3.4 Construcción del modelo y optimización El modelado PLS-DA intenta encontrar una relación lineal entre una matriz de predictores X , que en este estudio se obtuvo a partir de las regiones de integración o *buckets*, y una matriz de respuestas Y que corresponde a la clasificación del individuo con respecto a su estado fisiopatológico. Debido al alto número de predictores, los modelos pueden ser fácilmente sobreajustados, es decir, un modelo con poco error residual pero poca capacidad predictiva. La única forma de estimar de manera confiable los valores de Y es utilizar un nuevo conjunto de individuos. Esto se puede lograr dividiendo el conjunto de datos en dos subconjuntos: un conjunto de calibración o “*Calibration set*” y un conjunto de prueba “*Test set*”. El conjunto de calibración es utilizado para construir el modelo, y el conjunto de prueba se usa para estimar la capacidad predictiva del modelo (TRIBA et al., 2015; WORLEY; POWERS, 2013). Idealmente, el subconjunto de calibración debería ser dividido nuevamente en dos subconjuntos: un conjunto de calibración “*training set*” y un conjunto de validación “*validation set*”. Como la separación del conjunto de calibración entre el conjunto entrenamiento y el conjunto de validación a menudo no es posible por el poco número de muestras disponibles, debido a la naturaleza costosa de la adquisición de muestras y adquisición de datos, una alternativa es utilizar la validación cruzada interna donde el método “*Leave one out*” es el más común (WESTERHUIS et al., 2008; WORLEY; POWERS, 2013). Este método consiste en

separar los datos de forma que para cada iteración se deja un dato afuera para la validación y los demás datos son utilizados para el entrenamiento del modelo, se realizan tantas iteraciones como datos se tengan, el error se calcula como el promedio de los errores cometidos a través de todas las iteraciones.

La evaluación del modelo se rige por diferentes parámetros de calidad, los más utilizados en los estudios de metabolómica son el parámetro R^2 y Q^2 (también llamado R^2 de validación cruzada). R^2 mide la bondad del ajuste, mientras que Q^2 mide la capacidad predictiva del modelo, $R^2=1$ la descripción perfecta de los datos por el modelo y $Q^2=1$ indica la capacidad predictiva perfecta del modelo. R^2 aumenta progresivamente con el aumento de variables latentes, y se acerca a 1 cuando el número de variables latentes utilizadas se acerca al rango de la matriz de datos. Q^2 no necesariamente se acerca a uno al aumentar el número de variables latentes, sino que alcanzará una meseta y disminuirá, lo que indica que a cierto grado de complejidad la capacidad predictiva de modelo se deteriora (TRIBA et al., 2015; WESTERHUIS et al., 2008).

Otros parámetros de calidad empleados para evaluar el modelo pueden ser obtenidos a partir de la matriz de confusión (SZYMAŃSKA et al., 2012), esta herramienta permite visualizar el desempeño del modelo en términos de la clasificación de los objetos en cada clase, en este caso la clasificación del paciente como seropositivo o seronegativo, la diagonal contiene los objetos correctamente clasificados: Verdaderos Positivos (*True Positives*, *TP*) y Verdaderos Negativos (*True Negatives*, *TN*). Fuera de la diagonal se encuentran los pacientes que son clasificados incorrectamente: falso negativo (*False Negative*, *FN*) y falso positivos (*False Positive*, *FP*) (

Tabla 1).

Tabla 1. Matriz de confusión. Elaboración propia

		Clase real	
		Seropositivo	Seronegativo
Clase estimada	Seropositivo	Verdaderos	Falso
		Positivos	positivo
	Seronegativo	Falso	Verdadero
		Negativo	Negativo

Error de clasificación o “*misclassifications*”: Da la fracción de pacientes incorrectamente clasificados, se calcula mediante la Ecuación 2

$$Error\ de\ clasificación = \frac{\sum(\tilde{y}_i \neq y_i)}{N} = \frac{FN + FP}{TP + FN + FP + TN}$$

Ecuación 2

Precisión: Mide que tan bien el modelo puede asignar la clase de un paciente desconocido, se calcula mediante la Ecuación 3.

$$Precisión = \frac{\sum(\tilde{y}_i = y_i)}{N} = \frac{TP + TN}{TP + FN + FP + TN}$$

Ecuación 3

La suma de la precisión con el error de clasificación debe ser igual a 1.

Sensibilidad: Mide la capacidad del modelo para clasificar correctamente los pacientes seropositivos, se calcula mediante la Ecuación 4

$$Sensibilidad = \frac{TP}{TP + FN}$$

Ecuación 4

Especificidad: Mide la capacidad del modelo para clasificar correctamente los pacientes seronegativos, se calcula mediante la Ecuación 5

$$Especificidad = \frac{TN}{FP + TN}$$

Ecuación 5

A menudo el modelo está entrenado con toda la región espectral, la mayoría de estas variables son poco relevantes para el problema investigado ya que presentan una variación no relacionada con la respuesta a modelar, por tanto, las variables pueden reducirse drásticamente con poca pérdida de la información. Los métodos de selección de variables ayudan a seleccionar un pequeño conjunto de variables predictoras muy relevantes que están correlacionadas con una variable de respuesta particular. La selección de variables puede mejorar la estimación e interpretabilidad del modelo identificando efectivamente el subconjunto de predictores importantes (ANDERSEN; BRO, 2010; FARRÉS et al., 2015). Uno de los métodos ampliamente usados para la selección de variables es el estadístico VIP “*Variable Influence on Projection or Variable Importance in Projection*”, permite comprender en orden de importancia las variables predictoras X que mejor explican la varianza Y. El estadístico VIP es calculado de acuerdo con la siguiente fórmula, (Ecuación 6):

$$VIP_j = \sqrt{\frac{\sum_{f=1}^F w_{jf}^2 \cdot SSY_f \cdot J}{SSY_{total} \cdot F}}$$

Ecuación 6

Donde VIP_j es la contribución global de la variable j en el modelo, w_{ij} es el peso de

la variable j en la componente f , SSY_f es la suma de los cuadrados de la varianza explicada por la f^{th} componente, J es el número de variables predictoras, SSY_{total} es la suma total de los cuadrados de la varianza explicada por la variable dependiente y F el total de componentes (FARRÉS et al., 2015).

Por otra parte, un problema común con PLS-DA es su propensión al sobreajuste de datos. Esto ocurre cuando el algoritmo parece lograr una buena separación, pero lo ha hecho recogiendo ruido aleatorio en lugar de señales reales. Se ha demostrado que este problema no siempre se puede detectar a través de la validación cruzada, pero se puede detectar mediante pruebas de permutación (BIJLSMA et al., 2006). La prueba de permutación es una estrategia necesaria que puede evaluar si la clasificación específica de los individuos de los dos grupos es significativamente mejor que cualquier otra clasificación al azar en dos grupos arbitrarios. En una prueba de permutación, se calcula un nuevo modelo de clasificación después de permutar las etiquetas de las clases, es decir, las variables de clasificación se asignan aleatoriamente entre las distintas muestras del modelo, siendo el resultado esperado que la capacidad predictiva de los modelos aleatorios sea significativamente inferior a la obtenida para el modelo original. Además, la calidad puede ser evaluada a partir de Característica Operativa del Receptor (ROC; por sus siglas en inglés). La curva ROC es una representación gráfica de la sensibilidad frente a la especificidad para un sistema clasificador binario según se varía el umbral de discriminación, es útil para organizar clasificadores y visualizar su rendimiento. El valor del Área bajo la curva ROC está entre 0 y 1, donde 1 describe un valor de clasificación perfecto y 0,5 una modelo sin capacidad discriminatoria (FAWCETT, 2006).

2 OBJETIVOS

2.1 GENERAL

Analizar las fluctuaciones metabólicas en suero sanguíneo asociadas a la Tripanosomiasis americana mediante Resonancia Magnética Nuclear de hidrógeno.

2.2 ESPECÍFICOS

- ✓ Ajustar los parámetros experimentales para la toma de espectros de sueros sanguíneos de pacientes con Chagas en diferentes estados infecciosos mediante espectroscopia de Resonancia Magnética Nuclear.
- ✓ Construir un modelo estadístico para determinar la diferenciación metabólica.
- ✓ Identificar los posibles metabolitos con potencial como biomarcadores para métodos de diagnóstico clínico de Chagas.

3 AJUSTE DE LOS PARÁMETROS EXPERIMENTALES PARA LA TOMA DE ESPECTROS DE SUEROS SANGUÍNEOS

3.1 INTRODUCCIÓN

Una de las técnicas analíticas espectroscópicas más comunes utilizadas en estudios metabolómicos es la RMN, ya que permite identificar y cuantificar simultáneamente una amplia gama de compuestos orgánicos en el rango micromolar. La RMN es sencilla y en gran medida automatizada y no destructiva, por lo que las muestras pueden utilizarse para otros análisis químicos (Serkova & Niemann 2006, Zhang et al., 2013). Una de las dificultades en los estudios metabolómicos de biofluidos utilizando RMN se presenta debido a que estos tienen gran número de compuestos con variedad de pesos moleculares, concentración y movilidad molecular, donde la gran mayoría contienen hidrógeno que son detectados simultáneamente causando superposición de señales en el espectro. La presencia de moléculas de alto peso molecular como las proteínas y/o lípidos dominan las señales del espectro generando señales anchas que afectan la identificación de las moléculas más pequeñas (KAY, 1995; LIU; NICHOLSON; LINDON, 1996; RASTRELLI; JHA; MANCIN, 2009).

Se ha demostrado una separación de las señales en los biofluidos utilizando los tiempos de relajación y los coeficientes de difusión molecular. Los dos principales fenómenos de relajación de RMN son: relajación longitudinal que se caracteriza por un tiempo T_1 , y la relajación transversal caracterizada por un tiempo T_2 . A su vez, las moléculas con alto peso molecular tienen tiempos de correlación rotacional (τ_c) más largos que las moléculas más pequeñas dando como resultado tiempos de relajación más cortos. Por tanto, los espectros de RMN pueden ser editados aprovechando las diferencias en el tiempo de difusión y relajación de las moléculas para conseguir una mejor resolución (LIU; NICHOLSON; LINDON, 1996; RASTRELLI; JHA; MANCIN, 2009; TANG et al., 2004). La edición espectral puede

realizarse insertando un “filtro de relajación” dependiendo de los tiempos de relajación para atenuar o eliminar las señales de las moléculas con tiempos de relajación relativamente cortos. En el caso de la edición de T_1 se puede emplear una secuencia de pulsos de inversión-recuperación ($180^\circ_x - \tau_1 - 90^\circ_x$) de tal manera que la señal en el espectro tenga intensidades dependiendo del τ_1 relativo a T_1 . Análogamente, la edición de T_2 se puede realizar usando una secuencia spin-eco, típicamente la secuencia CPMG [$90^\circ_x - (\tau - 180^\circ_y - \tau)_n$], seleccionando adecuadamente τ y n las señales pueden ser separadas de acuerdo al T_2 . En general, las señales anchas de las macromoléculas son atenuadas o eliminadas completamente dejando las señales de resonancia más agudas de las moléculas pequeñas no afectadas (TANG et al., 2004; TANG; WANG, 2006).

Por otra parte, la concentración de ^1H presentes en $^1\text{H}_2\text{O}$ es mucho más alta que la concentración de ^1H en otras moléculas, con una relación relativa a protones del soluto del orden del 10^5 , hecho que dificulta la resolución de señales para compuestos que están en muy bajas concentraciones, como los metabolitos (LIU; NICHOLSON; LINDON, 1996; RASTRELLI; JHA; MANCIN, 2009). La edición espectral para la supresión de la señal de agua puede realizarse utilizando el método de “pre-saturación” en donde se aplica un pulso relativamente largo de baja potencia que irradia sólo la región espectral del solvente (H_2O) de tal forma que el solvente alcanza una condición de saturación y la señal es atenuada y/o eliminada del espectro.

El método de pre-saturación puede ser implementado fácilmente en la mayoría de las secuencias de pulsos y no afecta la cuantificación excepto para las señales cercanas al agua o que interactúan con ella (HORE, 1983; TANG; WANG, 2006). Una de las variantes más utilizadas es NOESY1D ($90^\circ - t_1 - 90^\circ - t_{\text{mix}} - 90^\circ - \text{AQ-RD}$). Este método ha demostrado funcionar bien para sistemas donde el intercambio de hidrógenos entre los metabolitos y el agua es lento o no ocurre, eliminando por completo la señal residual del agua (NICHOLSON et al., 1995; TANG; WANG,

2006), donde la energía de transferencia NOE es posible, lo que puede incrementar las áreas relativas de las señales procedentes de la fuerte interacción agua-molécula.

En general, dependiendo de la secuencia de pulsos empleada y el ajuste de los parámetros de la misma se puede editar el espectro en función de los pesos moleculares de los compuestos para seleccionar el tipo de analitos que se desea observar, en este caso los metabolitos. En este trabajo, se evaluaron los parámetros experimentales en muestras de suero sanguíneo para la eliminación de la señal de agua y el filtrado T_2 en las secuencias pulsos NOESY y CPMG, respectivamente.

3.2 METODOLOGÍA

3.2.1 Pacientes y recolección de muestras En este trabajo de investigación *“Estudio preliminar para el desarrollo de un método de diagnóstico clínico de Chagas a través de la diferenciación metabólica analizada por espectroscopia de Resonancia Magnética Nuclear”*, se utilizaron muestras de suero sanguíneo de pacientes en fase crónica de la infección por *T. cruzi* y de pacientes sanos. Las muestras fueron previamente recolectadas por el Grupo de Inmunología y Epidemiología Molecular (GIEM) de la Universidad Industrial de Santander en el marco del proyecto *“Estudio multicéntrico para la evaluación de marcadores genéticos asociados con el desarrollo de la enfermedad de Chagas en Colombia”* Código Colciencias: 38974, y se cuenta con autorización del investigador del estudio primario para uso de las mismas (**Anexo 1**). Para garantizar la confidencialidad de la información se cumplió con lo reglamentado en la Ley estatutaria 1581 de 2012 sobre el tratamiento de datos personales, y descritas en el presente documento en la sección de aspectos éticos. Este estudio cuenta con el aval del Comité de Ética en Investigación Científica, CEINCI de la Universidad Industrial de Santander (**Anexo 2**).

3.2.2 Experimentos de RMN Los espectros de los sueros sanguíneos fueron obtenidos en un espectrómetro Bruker Avance III de 9.4 Tesla (400MHz) equipado con una sonda directa H / F-X de 5 mm, disponible en el laboratorio de Resonancia Magnética Nuclear de la Universidad Industrial de Santander, Parque Tecnológico de Guatiguará. Para llevar a cabo el ajuste de los parámetros experimentales se trabajó únicamente con suero sanguíneo de pacientes seronegativos. Para cada una de las muestras se tomaron 500 μ L los cuales fueron colocados en un tubo de RMN, se insertó de forma coaxial un capilar con D₂O para la fijación de lock de campo el cual contenía el patrón de campo ácido trimetilsililpropanoico (TSP, 0,05 mg/mL). El TSP fue usado como referencia a $\delta=0.0$ ppm. Se evaluaron las secuencias de pulsos NOESY y CPMG con pre-saturación del solvente comparadas con un experimento one pulse (zgpr) con presaturación de solvente (**Figura 1**), todos los espectros fueron obtenidos con 128 *scans* a una temperatura de 298 °K, con una ventana espectral de 12 ppm.

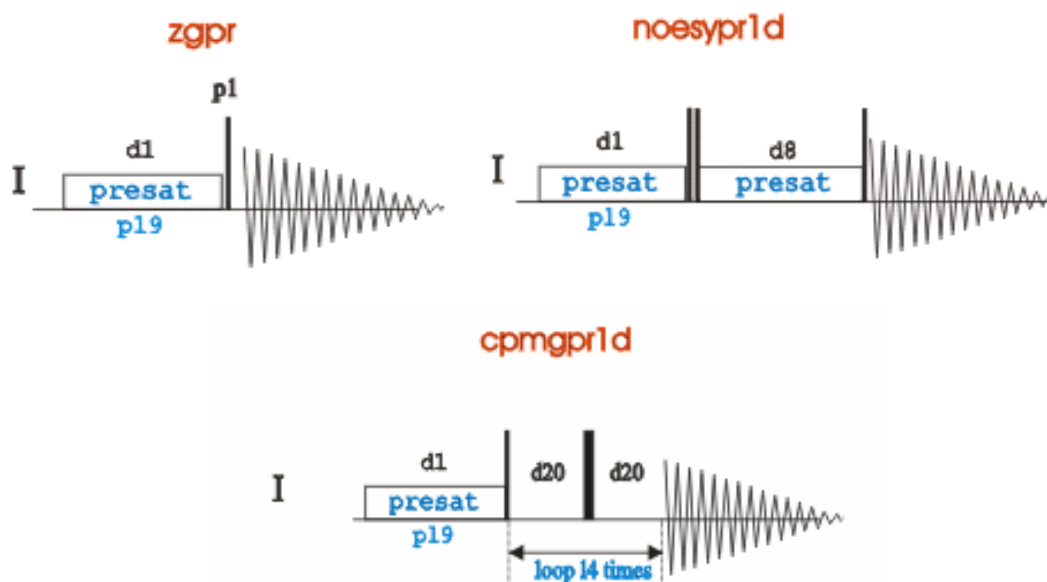


Figura 1. Secuencias de pulsos zgpr, NOESY y CPMG.

Fuente: Bruker BioSpin, Rheinstetten, Alemania.

3.3 RESULTADOS Y DISCUSIÓN

3.3.1 Selección de la secuencia de pulsos De acuerdo con las secuencias de pulso evaluadas, los resultados experimentales evidenciaron que con la secuencia NOESY se obtuvo una mejor pre-saturación del solvente comparado con la secuencia convencional zgpr y CPMG (**Figura 2**). Esto es debido a que el agua tiene un tiempo de relajación T_1 diferente al de los metabolitos, usualmente más largo. Aprovechando esta característica, durante el pre-*delay* el pulso de presaturación suprime la mayoría de la intensidad de la señal de agua. Los dos primeros pulsos de 90° invierten la magnetización al eje -z, cuando los metabolitos se relajan hacia el eje z, la señal residual del agua lo hace más lentamente, y allí durante el tiempo de mezclado, se puede aplicar un nuevo pulso de pre-saturación para eliminar la señal residual de agua (NICHOLSON et al., 1995). Sin embargo, esta secuencia proporciona una línea base y fases distorsionadas que podrían afectar el análisis estadístico posterior

Por otra parte, la secuencia CPMG provee una mejor línea base y fase, si bien la pre-saturación del solvente no es tan efectiva como con la secuencia NOESY, es mejor con respecto a la secuencia zgpr. Por otra parte, se evidenció una mejor resolución de las señales, con bandas mejor definidas en la región entre 0,5 y 4 ppm (**Figura 2**). Los resultados obtenidos son similares a los observados por Gabbani y colaboradores (GABBANI et al., 2017), donde desarrollaron un estudio metabolómico basado en RMN y compararon las secuencias NOESY y CPMG para discriminación de pacientes con hepatitis. Los modelos obtenidos con suero sanguíneo utilizando la secuencia CPMG mostraron una mayor precisión en comparación con los obtenidos con la secuencia NOESY. Estos resultados son atribuidos a la pérdida de información de las pequeñas moléculas debido a la superposición de señales de las macromoléculas presentes en el suero cuando se utiliza la secuencia NOESY.

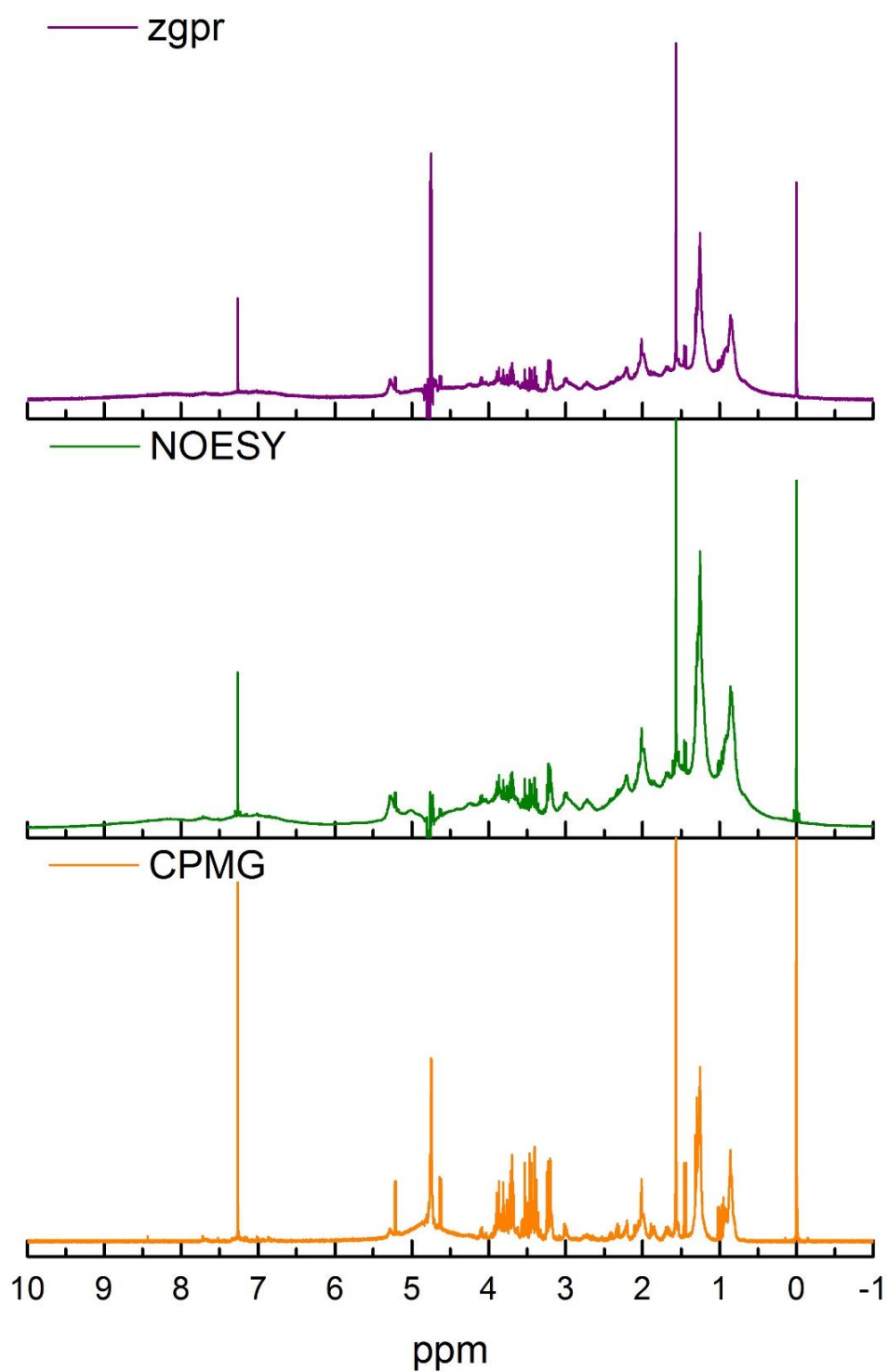


Figura 2. Espectros obtenidos con las secuencias zgpr, CPMG, NOESY. Elaboración propia.

En contraste, Li y colaboradores (LI et al., 2014) realizaron un análisis metabolómico en plasma sanguíneo ovino y obtuvieron resultados consistentes cuando emplearon las secuencias CPMG y NOESY. Por otra parte, Bharti y colaboradores (BHARTI et al., 2008) compararon la secuencia CPMG y la secuencia *one pulse* para una muestra de suero, lograron la eliminación de las señales ensanchadas de proteína con una mejor línea de base. Esto es debido a que los metabolitos tienen un tiempo de relajación más largo con respecto a las proteínas. La Ecuación 7 describe que tan rápido la magnetización transversal, M_{xy} decae a cero.

$$M_{xy} = M_o(e^{-\frac{t}{T_2}})$$

Ecuación 7

El tiempo de relajación T_2 es inversamente proporcional al tiempo de correlación de las moléculas, τ_c , (Ecuación 8) que a su vez está asociado con el tamaño molecular (Ecuación 9). Por tanto, entre mayor sea el tamaño molecular el tiempo de relajación T_2 será más corto, tal y como se ilustra en la **Figura 3**.

$$\frac{1}{T_2} = (\frac{H_d}{h})^2 \tau_c$$

Ecuación 8

$$\tau_c = \frac{4\pi a^3 \eta}{3kT} = \frac{V\eta}{kT}$$

Ecuación 9

Donde H_d es la energía de interacción dipolar, a es el radio molecular, η la viscosidad del medio, V volumen molecular, k la constante de Boltzman, T temperatura.

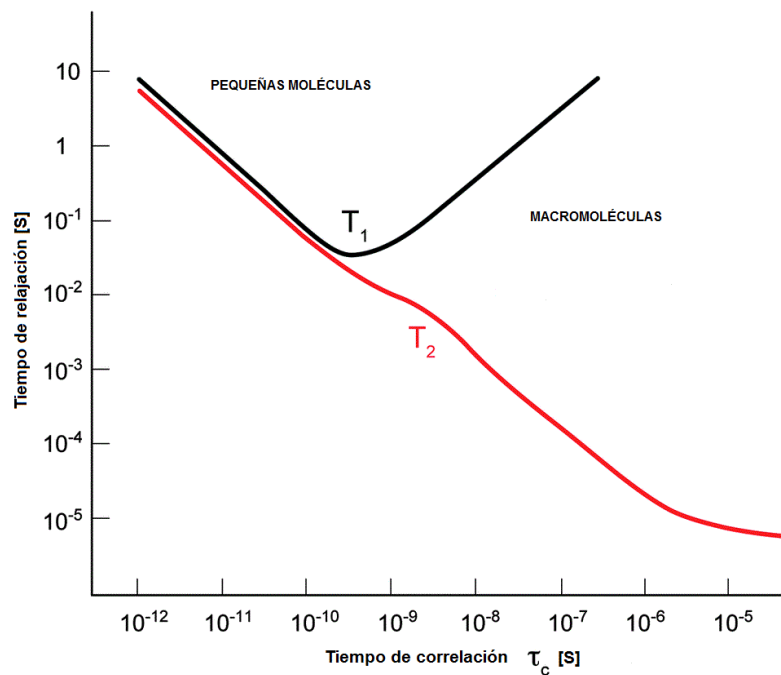


Figura 3. Diagrama esquemático de la relación entre la relajación longitudinal T_1 , la relajación transversal T_2 y el tiempo de correlación molecular τ_c . Adaptado de: (KESHARI; WILSON, 2014)

Este diferencial de T_2 puede ser utilizado en la secuencia CPMG para diseñar el experimento que permita la detección selectiva de las moléculas pequeñas tal como los metabolitos y la eliminación de las señales correspondientes a las macromoléculas (YAN, 2004). Además, la relajación transversal, T_2 , está relacionada con el ancho de la amplitud de la señal, $\Delta\nu_{1/2} = 1/\pi T_2$, y $\Delta\nu_{1/2}$ se define como el ancho a la altura media de la señal. Por tanto, al filtrar las señales de las macromoléculas la resolución espectral mejora ya que se eliminan las señales anchas y se evita el solapamiento con las señales correspondientes a los metabolitos.

En lugar de adquirir la FID después del pulso de 90° como en el experimento RMN unidimensional zgpr [presat- 90° - Adquisición], en la secuencia CPMG después del pulso de 90° se inserta un tren de pulsos de 180° para lograr el filtrado T_2

$[90^\circ_x - (\tau - 180^\circ_y - \tau)_n - \text{Adquisición}]$ (**Figura 1**, cpmgpr1d).

Inicialmente, la magnetización se encuentra orientada en la dirección z (**Figura 4a**), después del pulso inicial de 90° , la magnetización se encuentra en el plano xy (**Figura 4b**), la magnetización decaerá debido a la relajación T_2 durante un *delay* τ . El pulso de 180° rota la magnetización (**Figura 4c**) y esta se reenfoca después de otro *delay* τ (**Figura 4e**) (YAN, 2004). Para lograr el filtrado de las señales de interés es importante optimizar τ y n para lograr una longitud total del tren de pulsos ($2n\tau$) que sea lo suficientemente larga para suprimir las señales de las proteínas y lo suficientemente corta para preservar las señales de los metabolitos (YAN, 2004).

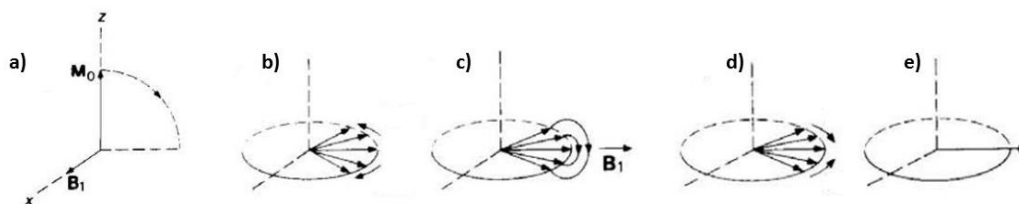


Figura 4. Diagrama de la secuencia de pulsos CPMG. Adaptado de (MEIBOOM; GILL, 1958)

3.3.2 Optimización de la secuencia de pulsos CPMG Previo a la optimización de la secuencia y para garantizar que los cambios en el espectro se dan sólo por los parámetros ajustados y no por errores en la preparación de muestra, se realizó un ensayo de la reproducibilidad de la preparación de la muestra. Para ello, se realizó la preparación de la muestra cuatro veces utilizando el mismo suero, y se realizaron las cuatro mediciones. Todas las mediciones se realizaron utilizando la secuencia CPMG y con los mismos parámetros.

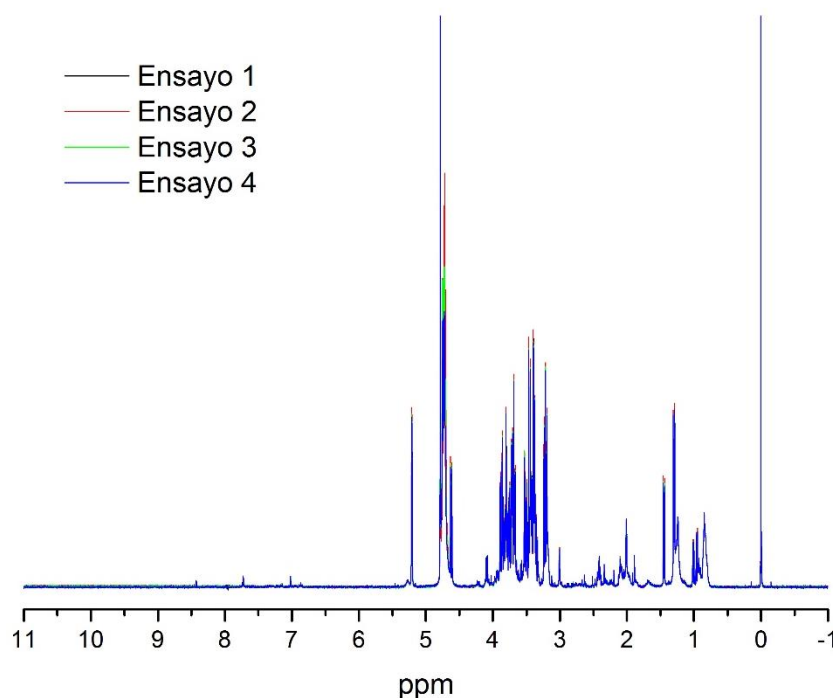


Figura 5. Ensayos de reproducibilidad utilizando la secuencia CPMG, con único suero.

De acuerdo con los resultados obtenidos **Figura 5** se puede observar que la preparación de muestras es reproducible, se observan pequeñas diferencias en las señales de la región de 4.5-5 ppm, es debido a que esa región se está pre-saturando la señal del agua y por tanto no es reproducible. Posteriormente, se evaluaron diferentes valores de τ y n con el objetivo de fijar los valores que dan una mejor resolución espectral. Inicialmente, para un número de ciclos $n=100$, se evaluaron valores de τ entre 0,0005 y 0,100 segundos (**Figura 6**) Por otra parte, para un $\tau=0,0005$, se evaluó el número de ciclos (n) entre 50 y 500 (**Figura 7**).

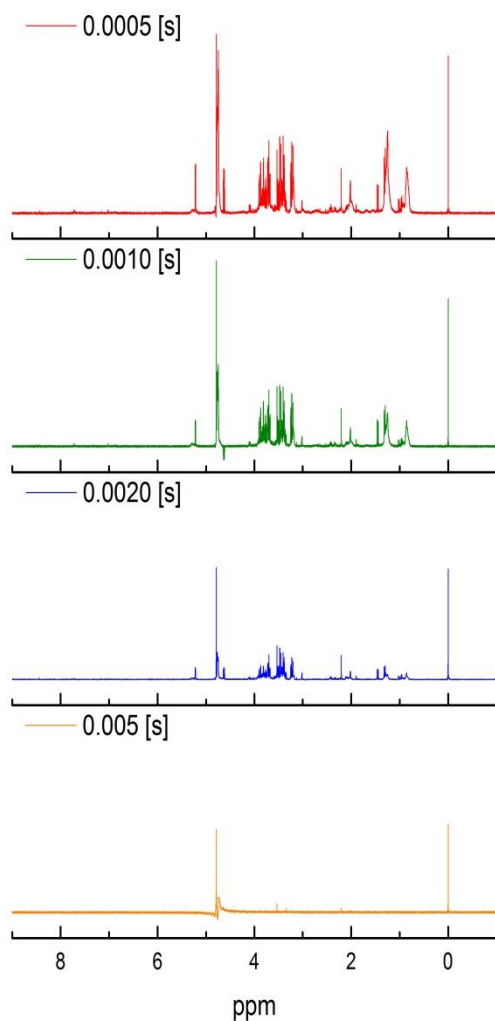


Figura 6. Espectros obtenidos con la secuencia CPMG con $n=100$ y para diferentes valores de τ .

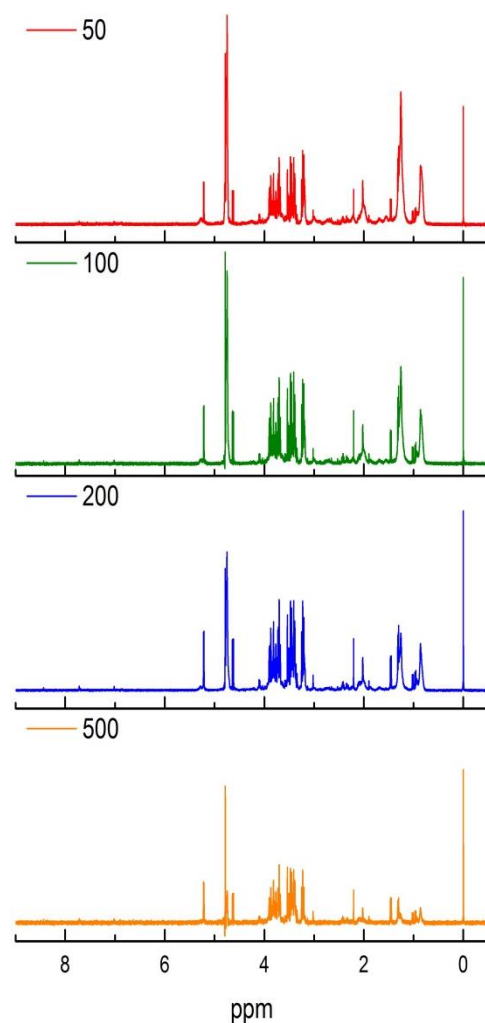


Figura 7. Espectros obtenidos con la secuencia CPMG con $\tau= 0.0005$ y diferentes valores de n .

La señal de las proteínas fue atenuada principalmente en la región entre 0,5 y 2 ppm. El tiempo T_2 total para cada experimento se lista en la **Tabla 2**. De acuerdo con estos resultados la señal de las proteínas fue atenuada en los experimentos con $n=100$ y $\tau=0,0020$, para un T_2 de 0,4 [s] y en el experimento $n=500$ y $\tau=0,005$ para un $T_2=0,5$. De acuerdo con estos resultados se decidió elegir el T_2 más corto

con el que se logra la atenuación, es decir, 0,4 [s]. Para ello se definieron los parámetros en $n=500$ y $\tau=0,005$.

Tabla 2. Cálculos de T_2 para los experimentos realizados. Elaboración propia.

τ	n	$2n\tau$
0,0005	100	0,1
0,001	100	0,2
0,002	100	0,4
0,005	100	1
0,0005	50	0,05
0,0005	100	0,1
0,0005	200	0,2
0,0005	500	0,5
0,0004	500	0,4

3.4 CONCLUSIONES

La edición espectral realizada a través de las secuencias de pulsos CPMG y NOESY permitió atenuar la señal del agua, componente mayoritario del suero sanguíneo, y observar la señal de los metabolitos, obteniéndose la eliminación total de la señal de agua en la secuencia NOESY.

Se seleccionó la secuencia CPMG para realizar las mediciones de los sueros sanguíneos, debido a que esta generó un espectro con mejor línea base y permitió la atenuación de las señales de las proteínas con un tiempo de eco de 400 μ s y con un número de ecos igual a 500, para un T_2 de 400 ms.

4 CONSTRUCCIÓN DE UN MODELO ESTADÍSTICO PARA DETERMINAR LA DIFERENCIACIÓN METABÓLICA

4.1 INTRODUCCIÓN

La espectroscopia de RMN aplicada a estudios metabolómicos genera un gran conjunto de datos complejos que requieren de métodos estadísticos multivariados para extraer la información. Una metodología de evaluación integrada de los datos del metaboloma obtenidos por RMN debe facilitar la interpretación e identificación de los metabolitos implicados en la diferenciación metabólica que tienen un alto potencial como biomarcadores de enfermedades.

El primer paso del análisis multivariado es encontrar la estructura y tendencia intrínseca de los datos, para ello se debe realizar un análisis estadístico basado en métodos estadísticos no supervisados, tal como, el análisis de componentes principales (PCA). Posteriormente, se realiza una análisis discriminante por mínimos cuadrados parciales (PLS-DA) para clasificar las muestras en categorías predefinidas o clases (BRINDLE et al., 2003; ZHANG et al., 2013). Finalmente, se identifican las regiones espectrales que conducen a la discriminación entre grupos y se optimiza utilizando criterios de selección de variables predictoras y variables latentes.

En este trabajo, se realizó un análisis quimiométrico al perfil metabólico obtenido a partir del suero sanguíneo utilizando RMN como herramienta analítica, a partir de los datos obtenidos se construyó un modelo PLS-DA, el cual fue optimizado realizando una selección de variables predictoras basada en el estadístico VIP y la selección de variables latentes basada en el mínimo del error de clasificación que permitió diferenciar pacientes con la EC de pacientes considerados como sanos.

4.2 METODOLOGÍA

4.2.1 Población de estudio La población de estudio está compuesta de 134 sujetos (Infectados, $n = 134$) con EC en fase crónica y 105 sujetos comparables que se incluyeron como control sano (sanos=105). La clasificación de los sujetos se llevó a cabo con dos pruebas serológicas, prueba ELISA con antígeno crudo, más prueba confirmatoria con antígenos recombinantes. Se utilizó como criterio de inclusión personas de áreas rurales donde el nivel de endemidad es alto.

4.2.2 Pre-procesado de los espectros El pre-procesado de los espectros se llevó a cabo utilizando el software topspin 3.5pl7 (bruker), en el cual se realizó la transformada de Fourier, el ajuste de la fase y línea base. El procesamiento de los espectros se realizó con el software NMRprocFlow (JACOB et al., 2017), el alineamiento se llevó a cabo utilizando un algoritmo basado en los mínimos cuadrados tomando como referencia el promedio de espectros con máximo de desplazamiento relativo de 0,05 ppm, a manera de ejemplo se muestra la región comprendida entre 0.5-1.6 ppm antes y después de realizar el alineamiento, **Figura 8** y **Figura 9**, respectivamente.

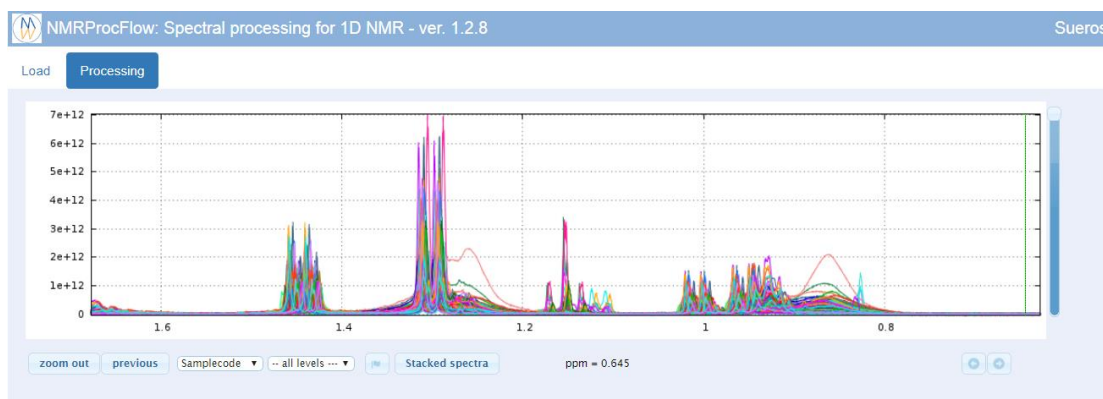


Figura 8. Región espectral de 0.5 -1.6 ppm antes del alineamiento. Elaboración propia.

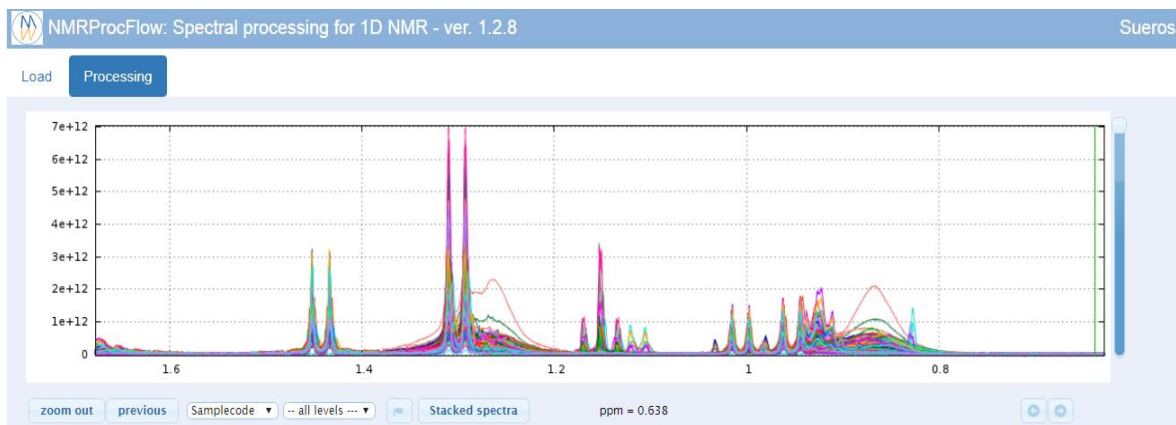


Figura 9. Región espectral de 0.5 -1.6 ppm después del alineamiento. Elaboración propia.

Posteriormente, se llevó a cabo la exclusión de la señal residual del agua en la región de 4,65-5,16 ppm, y de la señal del TSP en la región de -0,2-0,2 ppm. Los *buckets* fueron obtenidos en la región de 0,5-9 ppm utilizando un *bucketing* inteligente con factor de resolución de 0,2 y un umbral de la relación señal/ruido igual a 2, se obtuvo un total de 196 *buckets* (**Figura 10**). Finalmente, la normalización de los espectros se llevó a cabo utilizando una normalización integral.

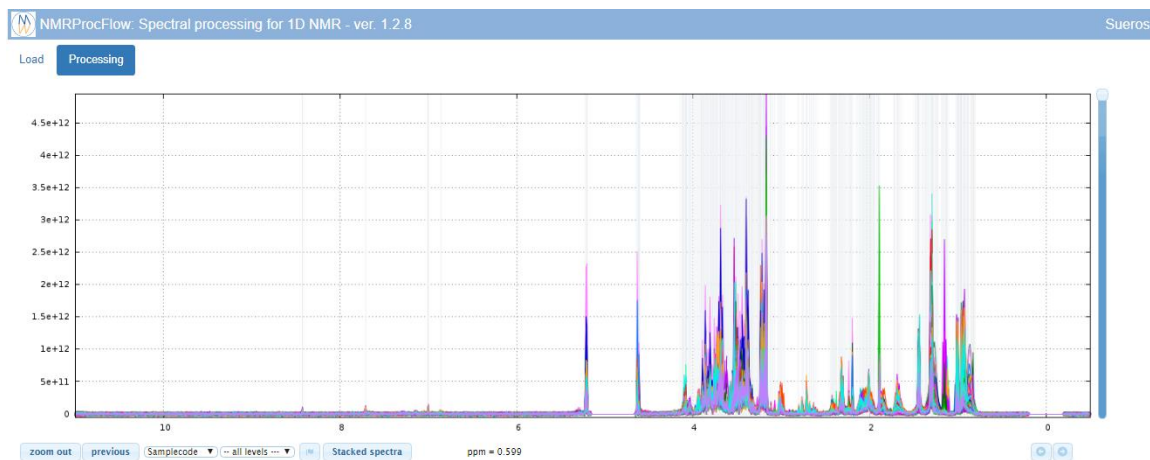


Figura 10 Segmentación de los espectros utilizando *Bucketing* inteligente entre la región de 0.5-9 ppm. Elaboración propia.

4.2.3 Análisis estadístico multivariado El análisis estadístico multivariado se llevó a cabo utilizando el software Metaboanalyst (XIA; WISHART, 2016) y el software R (R core team, 2017) empleado las librerías y FactorMineR *mdatools*.

El escalado de las variables se llevó a cabo aplicando un auto-escalado. Una vez los datos fueron escalados (**Figura 11**), se realizó el análisis multivariado, se llevó a cabo el análisis de componentes principales, donde se realizó un filtrado de datos, eliminando los datos que estaban fuera del intervalo de confianza. Posteriormente, se realizó un análisis supervisado, para ello se aplicó un análisis PLS-DA.

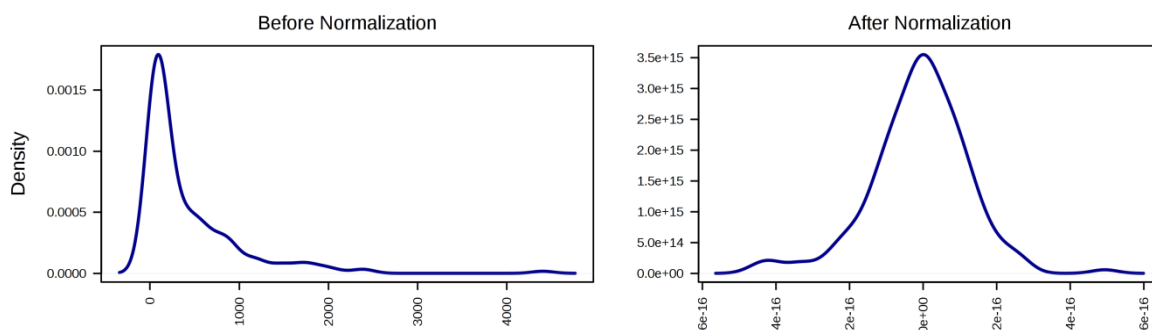


Figura 11. Distribución normal. Izquierda: Antes del auto-escalado. Derecha: Después del escalado. Elaboración propia

4.3 RESULTADOS Y DISCUSIÓN

4.3.1 Selección de muestras Se realizó la medición de un total de 239 muestras, de las cuales se seleccionaron 93 muestras de pacientes seronegativos y 126 de pacientes seropositivos. Se eliminaron un total de 20 muestras debido a que presentan señales atípicas comparadas con el grupo. En 15 de estos espectros se observó una posible contaminación (**Figura 12**), debido a la aparición con alta intensidad relativa de un triplete en las señales comprendidas entre las regiones 1-

1,2 ppm y un cuarte en la región 3,5-3,7 ppm. Probablemente esta contaminación corresponda al etanol, que presenta estas señales características. El etanol es empleado durante la manipulación de las muestras para descontaminar los equipos y/o áreas de trabajo.

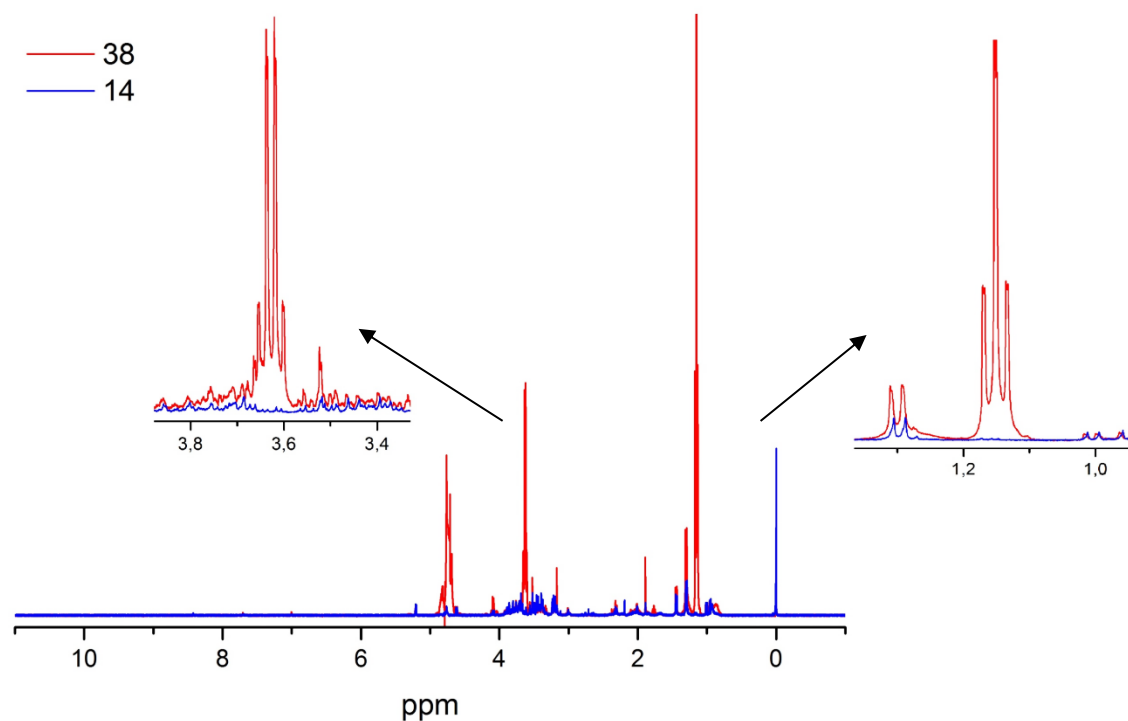


Figura 12. Comparación espectros de ^1H -NMR de suero sanguíneo muestra 38 con posible contaminación y muestra 14 con señales características del suero sanguíneo. Elaboración propia.

Por otra parte, las demás muestras eliminadas se observan unas señales atípicas ensanchadas en la región de 0,6-2 ppm (**Figura 13**). Estas muestras fueron eliminadas debido a que dificultan el pre-procesado de los espectros, específicamente la etapa de alineamiento espectral.

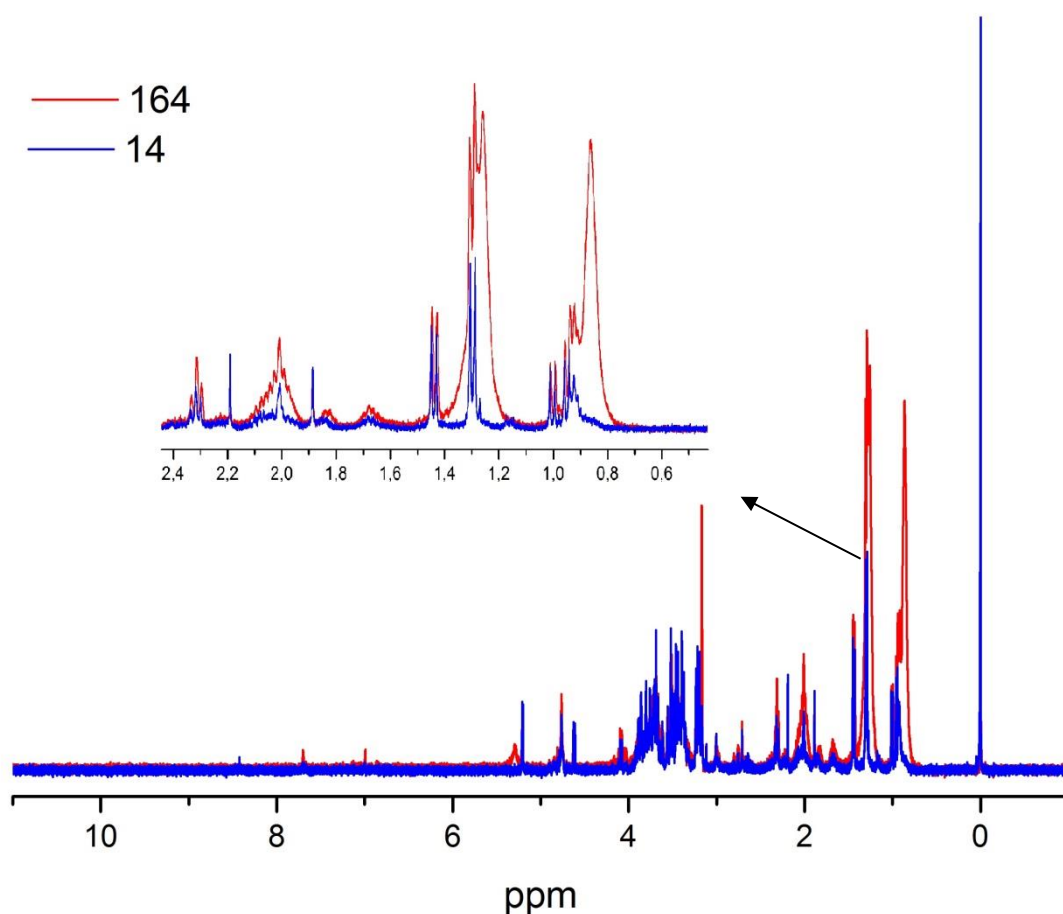


Figura 13. Espectro de ^1H -NMR de suero sanguíneo de la muestra 164 con señales atípicas comparada con la muestra 14 que presenta las señales características. Elaboración propia.

4.3.2 Análisis de componentes principales PCA fue realizado para analizar los datos espectrales, con el fin de identificar y estudiar cualquier muestra que presente un comportamiento atípico en relación con su grupo (*outliers*) y decidir la inclusión o exclusión de muestras en el estudio. Para cada grupo de muestras se analizaron los *score plots* y aquellos datos que se ubican por fuera de la elipse que representa un intervalo de confianza del 95% utilizando el estadístico T2 de Hotelling, fueron considerados *outliers* (Figura 14). Este proceso se realizó iterativamente hasta que se eliminaron todas las muestras fuera de intervalo de confianza. Los resultados

obtenidos en el software Metaboanalyst fueron comparados con los resultados obtenidos usando el software R con la librería FactorMineR (Anexo 3).

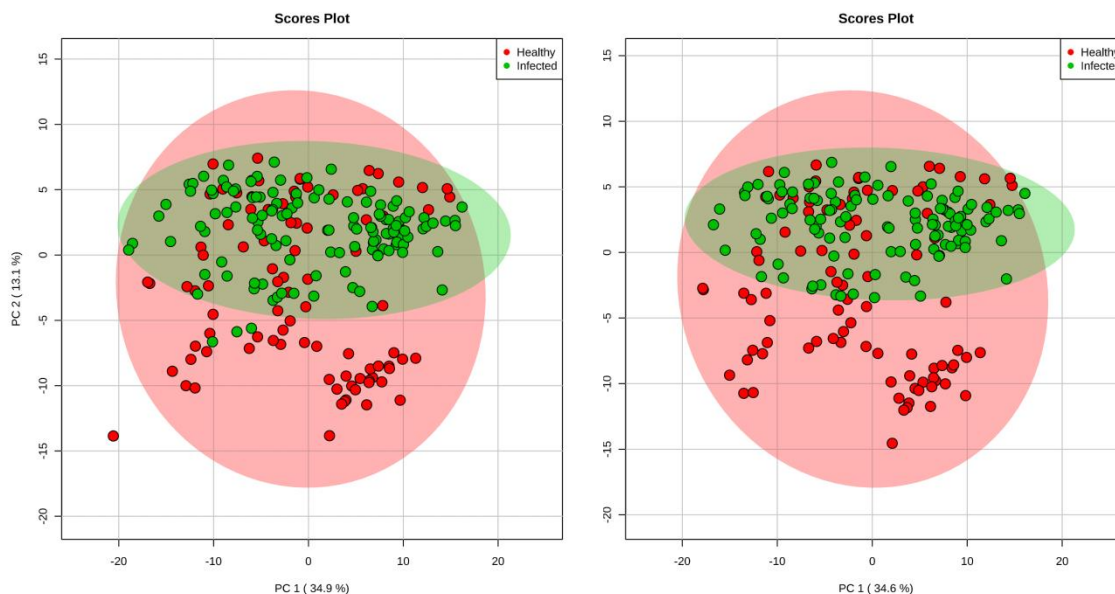


Figura 14. Score plots. Izquierda: antes de realizar el filtrado de datos. Derecha: después de retirar los outliers. Elaboración propia.

Se identificaron como *outliers* nueve muestras, una seronegativa etiquetada con el número 80 y ocho muestras seropositivas con las siguientes etiquetas: 6, 722, 188, 2251, 393, 2882, 2808, 61. Con el objeto de comprender el comportamiento atípico de los datos espectrales que fueron identificados como *outliers*, se realizó una revisión del espectro de RMN-¹H. El espectro de la muestra 80 fue comparado con dos espectros del mismo grupo que se encontraron dentro del intervalo de confianza, muestras 2247 y 14, en la **Figura 15** se muestra la comparación realizada en la región de 0,5-4,3 ppm, se puede observar claramente diferencias en las señales donde algunas presentan una mayor intensidad y en otros casos aparecen nuevas señales.

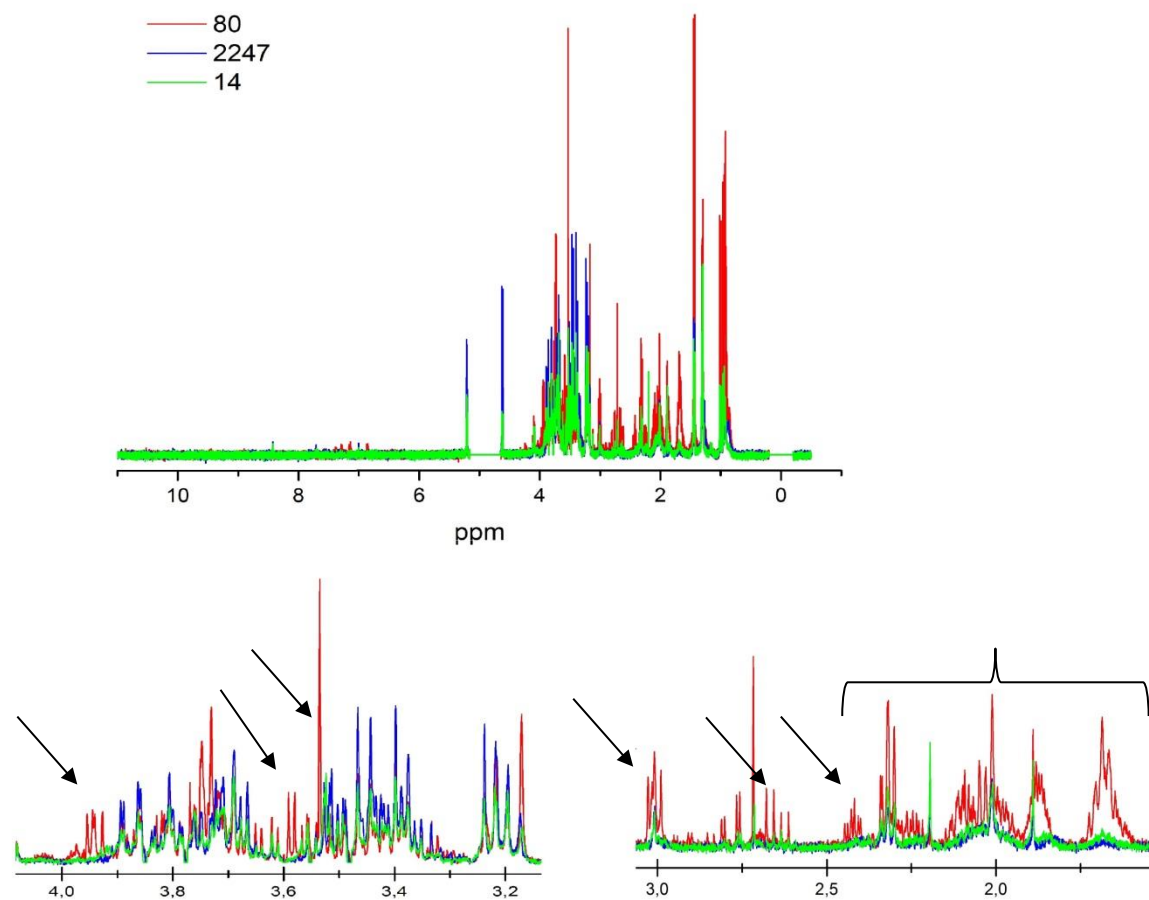


Figura 15. Comparación espectros RMN-H de la muestra 80 identificada como outlier con dos muestras (14 y 2247) elegidas al azar que se encuentran dentro del intervalo de confianza. Elaboración propia.

Por otra parte, también se analizaron los *outliers* obtenidos para las muestras seropositivas, en la **Figura 16** se muestra la comparación de la muestra 2808 identificada como *outlier* con las muestras 59 y 154 que se encuentran dentro del intervalo de confianza. Como puede observarse la principal diferencia se encuentra en la región de 1-1,2 ppm donde aparece una señal inexistente en las otras muestras, además de un aumento en la intensidad de las señales que aparecen en la región de 1,2-1,4 ppm. Todas las muestras consideradas como *outliers* fueron eliminadas para el posterior análisis supervisado.

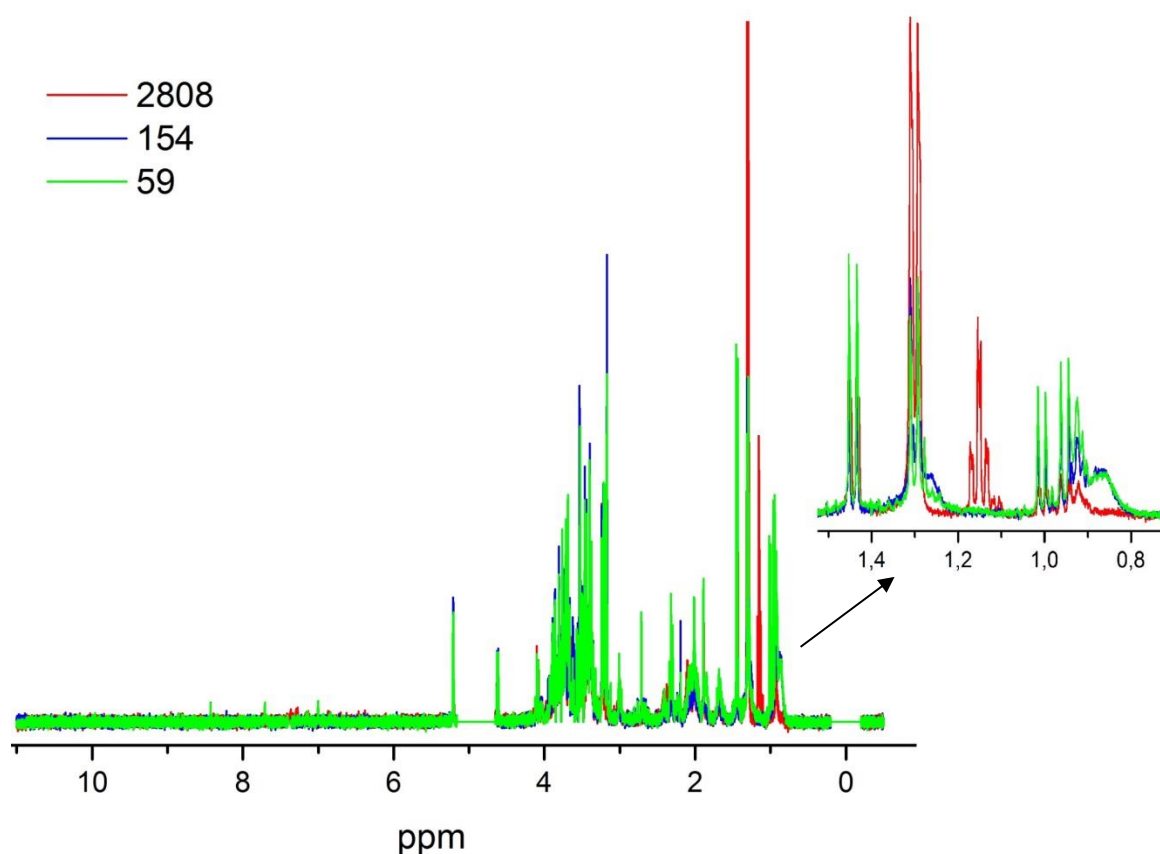


Figura 16. Comparación espectros RMN- ^1H de la muestra 2808 identificada como outlier con dos muestras 154 y 59 elegidas al azar que se encuentran dentro del intervalo de confianza. Elaboración propia.

Además, para observar si hay alguna agrupación intrínseca dada por el género o la edad, es decir, si estas características actúan como un factor de confusión se realizó el análisis de componentes principales etiquetando la muestra por género Femenino, F y Masculino, M. Por otro lado, para generar los grupos acordes con la edad se tuvo en cuenta que el promedio de edad es de 57 años se etiquetaron como “Mayor” aquellos sujetos cuya edad es mayor o igual a 57 años y como “Menor” los pacientes cuya edad es menor a 57 años. La **Figura 17** muestra los agrupamientos obtenidos cuando las muestras fueron etiquetadas de acuerdo al género y la **Figura 18** muestra el agrupamiento obtenido cuando las muestras fueron separadas de acuerdo con la edad, como puede observarse los grupos se encuentran altamente

solapados de manera que estas características no están formando interferencias para el análisis entre los grupo seropositivos y seronegativos.

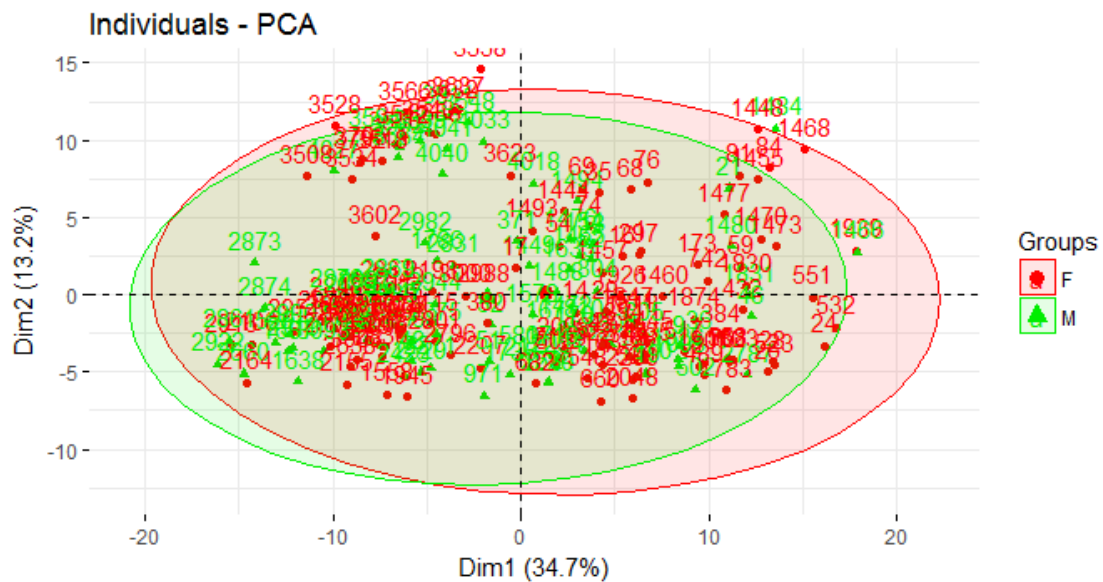


Figura 17. Score plot del análisis PCA utilizando las etiquetas de acuerdo con el género. Elaboración propia.

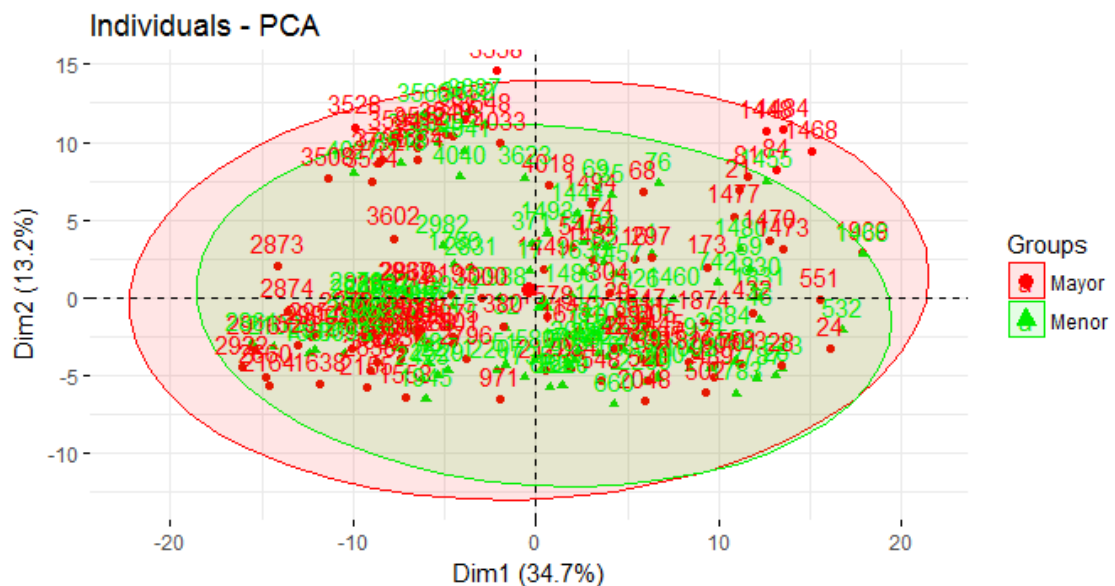


Figura 18. Score plot del análisis PCA utilizando las etiquetas de acuerdo con la edad. Elaboración propia.

4.3.3. Análisis Discriminante por mínimos cuadrados parciales Una vez seleccionadas las muestras a través del PCA, se procedió a realizar el análisis supervisado PLS-DA, para el cual se emplearon 210 muestras con 196 variables definidas como variables predictoras y una variable clasificatoria que permite distinguir entre los dos grupos pacientes seronegativos (n=92) y pacientes seropositivos (n=118), etiquetados como “*Healthy*” e “*Infected*”, respectivamente. La **Tabla 3** muestra los datos del grupo de estudio del análisis discriminante.

Tabla 3. Muestras empleadas para realizar el análisis supervisado PLS-DA. Elaboración propia.

Condición	Masculino		Femenino	
	Número de casos	Edad promedio	Número de casos	Edad promedio
Seronegativos “Healthy”	39	57±6,4	53	58±6
Seropositivos “Infected”	49	54,6±6,5	69	59±6

Para construir el modelo las muestras fueron divididas en dos conjuntos, el conjunto de calibración o “*Training set*” que contiene el 80% de las muestras (n=168) y el conjunto de prueba “*test set*” que contiene el 20% de las muestras (n=42), las muestras fueron divididas utilizando una función de azar ‘sample’ de R. El modelo fue desarrollado y validado internamente utilizando el conjunto de entrenamiento realizando una validación cruzada de tipo “*Leave one out (LOO-CV)*”. Se obtuvo un primer modelo el cual se desarrolló con 8 variables latentes o componentes y las 196 variables predictoras. Este modelo fue etiquetado como “Modelo sin selección de variables, Mssv. La **Tabla 4** muestra la matriz de confusión obtenida para el modelo. Los parámetros de calidad del modelo Mssv se enlistan en la **Tabla 5**.

Tabla 4. Matriz de confusión obtenida para el modelo Mssv. Elaboración propia.

		Clase real	
		Seropositivo	Seronegativo
Clase estimada	Seropositivo	87	10
	Seronegativo	5	66

Tabla 5. Parámetros de calidad obtenidos para el modelo Mssv. Elaboración propia.

Parámetro	Valor
R^2	0,84
Q^2	0,62
Especificidad (cv)	0,868
Sensibilidad (cv)	0,946
Precisión (cv)	0,911
Error de clasificación (cv)	0,089

cv: parámetro calculado a partir de la validación cruzada

Con el objeto de mejorar desempeño del modelo Mssv se realizó una selección de variables predictoras basado en el estadístico $VIP > 1$ (**Figura 19**)(FARRÉS et al., 2015; GUPTA; GUPTA; MAHDI, 2015) y se construyó un modelo el cual se etiquetó como MVIP. Por otra parte, se realizó la selección de variables latentes basado en el mínimo del error de clasificación para la validación cruzada (**Figura 20**) (FERNÁNDEZ PIERNA et al., 2009), a partir de esta selección se construyó un modelo etiquetado como M_{SME} .

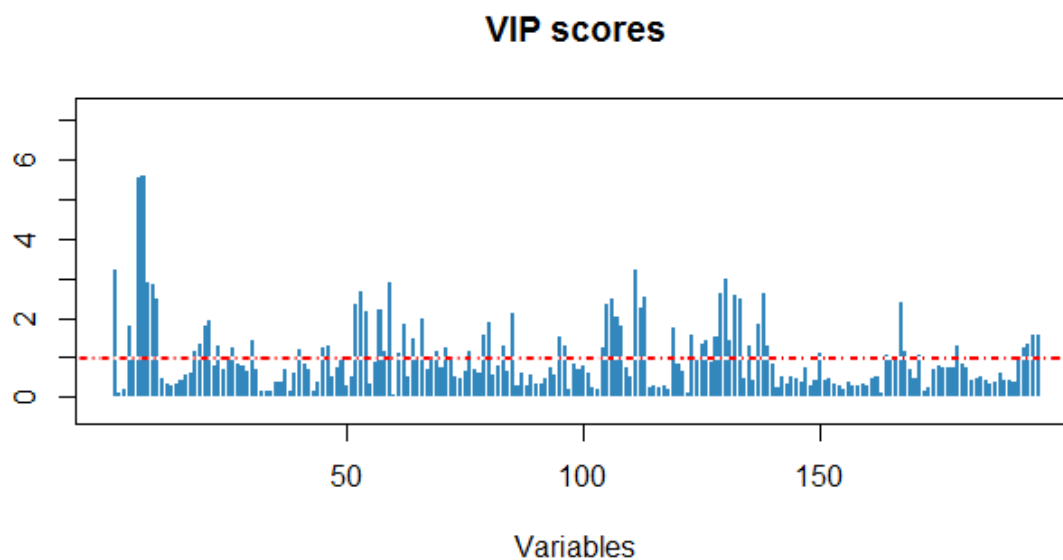


Figura 19. Estadístico VIP obtenido para las variables predictor a partir del modelo Mssv. Elaboración propia.

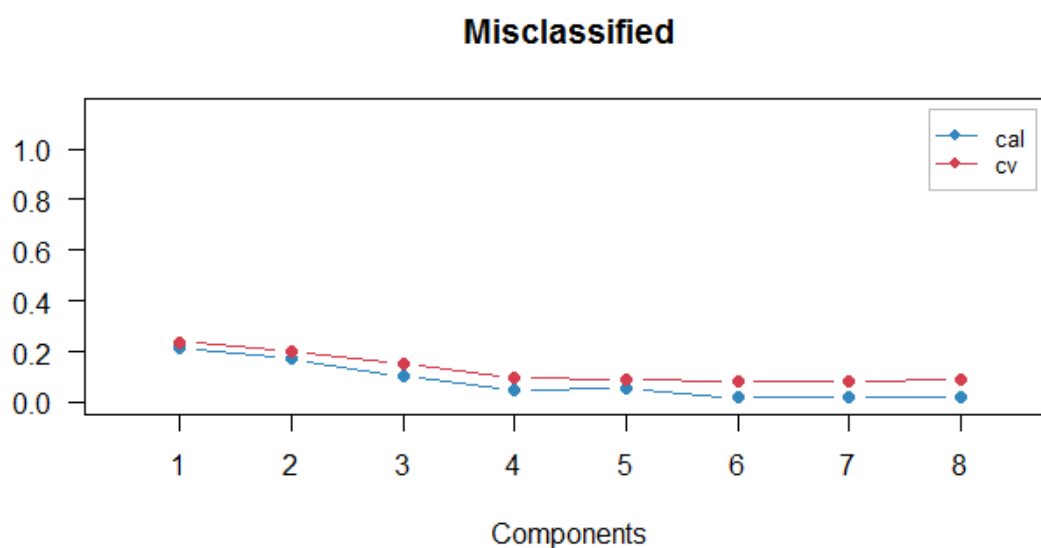


Figura 20 Error de clasificación para el modelo Mssv obtenido a través de las 8 variables latentes. Elaboración propia.

Finalmente, se desarrolló un modelo donde se realizó la selección de variables predictoras de acuerdo con el estadístico $VIP > 1$ obtenidos en el modelo M_{SME} y la selección de variables latentes de acuerdo con el mínimo del error de clasificación, el modelo fue etiquetado como $M_{SME-VIP}$. Los parámetros de calidad para todos los modelos evaluados se resumen en la **Tabla 6**.

Tabla 6. Parámetros de calidad para cada uno de los modelos evaluados. Elaboración propia.

Modelo	#VL	Criterio SV	R^2	Q^2	Esp.	Sen.	Prec.	ErrorClas
MSSV	8	N.A	0,84	0,62	0,868	0,946	0,911	0,095
M_{VIP}	8	$VIP > 1$ nc=8	0,83	0,67	0,895	0,946	0,923	0,077
M_{SME}	6	N.A	0,81	0,63	0,895	0,935	0,917	0,083
$M_{SME-VIP}$	6	$VIP > 1$ nc=6	0,79	0,68	0,908	0,967	0,940	0,059
Predicción Test set	6	$VIP > 1$ nc=6	0,71	N.A	0,938	0,962	0,952	0,048

#LV: Número de variables latentes. Criterio SV: Criterio de selección de variables. nc: número de componentes. Esp: Especificidad. Sen: Sensibilidad. Prec: Precisión. ErrorClas: Error de Clasificación.

De acuerdo con los parámetros de calidad obtenidos el mejor modelo se obtuvo con seis variables latentes, y realizando una selección de variables $VIP > 1$ con seis componentes, $M_{SME-VIP}$, el cual arroja una precisión del 94%, la especificidad del 91% y sensibilidad del 97%. El *score plot* del modelo $M_{SME-VIP}$ (**Figura 21**) donde se puede observar el agrupamiento por clases, la separación no es completamente clara ya que el modelo fue elaborado con seis variables latentes y *score plot*

contiene las dos primeras. Las regiones responsables de la discriminación con valor $VIP > 1$ son resumidas en la **Tabla 7**. Además, se realizó el test de permutación con 1000 modelos y se obtuvo un $p\text{ value} < 0,001$ demostrando que el modelo es significativamente válido y la curva ROC la cual presenta un área bajo la curva de 0,96 evidenciando la buena capacidad predictiva del modelo (¡Error! No se encuentra el origen de la referencia.).

Teniendo en cuenta que este fue el mejor modelo, se realizó la evaluación del mismo a partir del conjunto de prueba “*Test set*” el cual nunca hizo parte del desarrollo del modelo (GUPTA; GUPTA; MAHDI, 2015), se obtuvieron los siguientes valores de predicción: R^2 de 0,71, con una precisión del 95%, especificidad del 94% y una sensibilidad del 96% (**Tabla 6, Figura 22**).

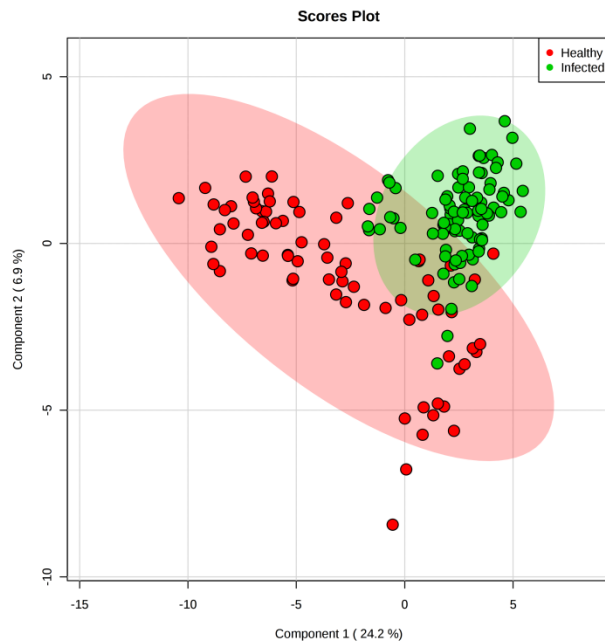


Figura 21. Score plot obtenido para el modelo $M_{SME-VIP}$. Elaboración propia.

**Tabla 7. Regiones responsables de la discriminación en el modelo M_{SME-VIP}.
Elaboración propia.**

Bucket	Valor VIP	Región espectral RMN ¹ H δ ppm		
		Centro	Mín	Máy
B4_6119	2,62	4,612	4,602	4,622
B4_6328	2,55	4,633	4,622	4,643
B4_0803	2,39	4,080	4,071	4,089
B4_1000	2,32	4,100	4,089	4,111
B4_0636	2,12	4,064	4,061	4,066
B2_3864	1,92	2,386	2,384	2,389
B2_1055	1,80	2,106	2,103	2,108
B3_5347	1,70	3,535	3,531	3,538
B2_0597	1,56	2,060	2,057	2,063
B2_0936	1,54	2,094	2,092	2,095
B2_1305	1,49	2,130	2,128	2,133
B2_0890	1,47	2,089	2,086	2,092
B2_4359	1,44	2,436	2,433	2,438
B3_5849	1,42	3,585	3,583	3,587
B8_4260	1,42	8,426	8,421	8,432
B2_4474	1,42	2,447	2,445	2,449
B3_5895	1,40	3,589	3,587	3,592
B2_3523	1,38	2,352	2,350	2,355
B3_4547	1,38	3,455	3,451	3,458
B3_5769	1,32	3,577	3,571	3,583
B2_4277	1,30	2,428	2,422	2,433
B3_9476	1,25	3,948	3,943	3,952
B2_0666	1,24	2,067	2,063	2,070
B2_4171	1,21	2,417	2,412	2,422
B3_2994	1,10	3,299	3,296	3,302
B3_9549	1,08	3,955	3,952	3,958
B3_5480	1,06	3,548	3,546	3,550
B1_3830	1,04	1,383	1,380	1,386
B2_1358	1,02	2,136	2,133	2,139

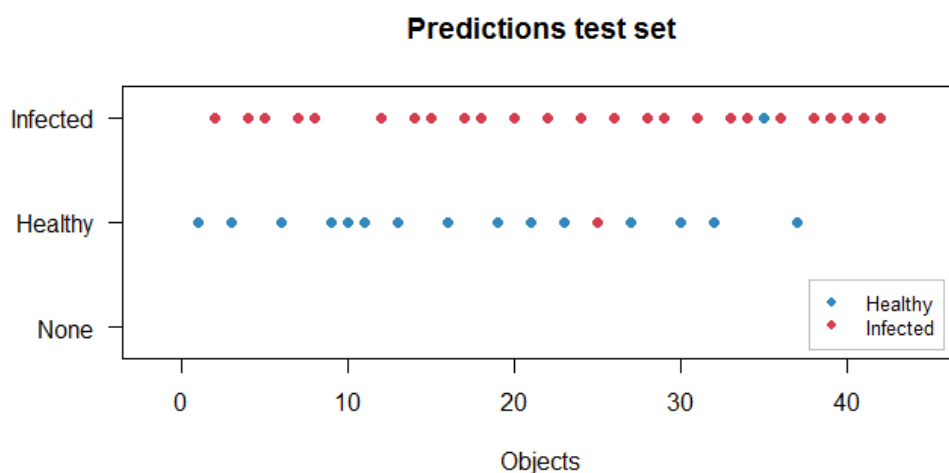


Figura 22. Evaluación del conjunto *test set* en el modelo $M_{SME-VIP}$. Elaboración propia.

Actualmente, no se encuentran estudios publicados de metabolómica para la EC utilizando RMN como herramienta analítica. Los resultados obtenidos muestran la buena clasificación del modelo, donde se consideran como valores aceptables R^2 y $Q^2 > 0,5$ (GULERIA et al., 2016). El modelo obtenido en este estudio es comparable con los modelos PLS-DA recientemente obtenidos para otras enfermedades para diferenciar casos vs control, empleando suero sanguíneo y RMN como herramienta analítica **Tabla 8**.

Tabla 8. Estudios de enfermedades a través de la diferenciación metabólica en suero sanguíneo utilizando RMN y modelado PLS-DA. Elaboración propia.

Autor	Enfermedad	Casos	Parámetros estadísticos
(MACINTYRE et al., 2010)	Leucemia linfocítica crónica	Control: 38 Enfermos: 29	$R^2=0,8$ $Q^2=0,5$
(DAWISKIBA et al., 2014)	Enfermedad Inflamatoria intestinal	Control: 17 Enfermos: 43	$Q^2=0,79$
(EMBADE et al., 2016)	Hepatitis C crónica	Control: 30 Enfermos: 27	$R^2=0,8$ $Q^2=0,6$
(ZABEK et al., 2016)	Artritis Reumatoide	Control: 30 Enfermos: 20	$R^2=0,635$ $Q^2=0,845$
(GULERIA et al., 2016)	Lupus eritematoso sistémico	Control: 30 Enfermos: 22	$R^2=0,75$ $Q^2=0,87$
	Nefritis lúpica	Control:30 Enfermos: 40	$R^2=0,75$ $Q^2=0,86$
(PALISI et al., 2017)	Síndrome antifosfolípido	Control: 27- Sanos: 27	$R^2=0,74$ $Q^2=0,50$
(ZHENG et al., 2017)	Hepatitis B	Control: 25 Enfermos: 20	$R^2=0,61$ $Q^2=0,50$
En esta investigación	Enfermedad de Chagas	Control:92 Enfermos:118	$R^2=0,79$ $Q^2=0,68$

Los modelos se compararon en términos de los parámetros R^2 y Q^2 . Para estudios relacionados con otras enfermedades el parámetro R^2 se encontró en un rango de 0,61-0,8, en el modelo desarrollado $R^2_{MSME-VIP}$ tiene valor de 0,79. El parámetro Q^2 que se encuentra en un rango de 0,5-0,86, y para el modelo $MSME-VIP$ tiene un valor 0,68. Por otra parte, es necesario comparar la capacidad de discriminación del modelo estadístico con otros métodos de diagnóstico clínico empleados convencionalmente para la detección de la EC. La **Tabla 9** muestra el resumen de los diferentes métodos de diagnóstico empleados para la detección de la EC en fase

crónica, la sensibilidad de los métodos se encuentra entre el 92-100% y la especificidad entre el 82-100%. El modelo obtenido y evaluado $M_{SME-VIP}$ se encuentra dentro de este rango de valores evidenciando la buena capacidad predictiva del modelo. Se debe tener en cuenta que este es un estudio preliminar basado en el perfil metabólico, que requiere de más etapas de evaluación pueda ser aplicado en el campo clínico, de acuerdo con los resultados obtenidos se evidencia el potencial de RMN para ser empleada como herramienta de diagnóstico clínico. Además, en comparación con los métodos de diagnóstico de la Tabla 9 cabe resaltar que el modelo $M_{SME-VIP}$ es el único que basa su clasificación en la fluctuación de metabolitos, en este sentido RMN podría ser útil en la identificación de metabolitos que puedan utilizarse como biomarcadores para diseñar un método de diagnóstico clínico sencillo y confiable, que tenga la suficiente robustez analítica.

Tabla 9. Comparación con otros métodos de diagnóstico de la EC. Elaboración propia.

Autor	Método referencia	Número de casos	Método evaluado	%Sen.	%Esp.
(CABALLERO et al., 2007)	TESA	E. Chagas: 53 NoCh: 134	ELISA-IMT	100	85,07
			BIOCHAGAS III (BIOSChile)	100	90,30
			ELISAcruzi (bioMérieux Brasil SA)	100	82,84
			Chagaek (bioMérieux Brasil SA)	100	87,31
			Chagatest Rec v3.0 (Wiener)	100	100
			Pathozyme Chagas (Omega)	100	100
(FERRERA et al., 2013)	Diagnóstico por cardiólogo	E. Chagas: 42 NoCh: 30	ELISA	95,2	93,3
			IFI	95,2	93,3
			HAI	88,1	100
			PCR-ADNk	76,9	100
			PCR-ADNsat	26,2	100
(CERVANTES-LANDÍN et al., 2014)	ELISA y Western Blot	E. Chagas: 96 NoCh: 224	ELISA-Dot	97	89
(FLORES-CHÁVEZ et al., 2010)	Clasificación previa por los grupo de trabajo: Qpanel y CNM	E.Chagas: 66 NoCh: 157	IFI-CNM	98,5	85,4
			ELISA-CNM	100	88,5
			Certest	100	84,1
			Ortho	100	88,5
			BKL	97,6	90,3
			BioMérieux	98,5	94,3
			Biokit	100	94,9
			ID-PaGIA2	92,4	98,7
			ID-PaGIA3	98,5	98,1
			ICT Operon	92,4	92,4
(NEVES SANTOS et al., 2016)	Dos pruebas serológicas positivas + valoración médica	E. Chagas: 246 NoCh: 1247	ICT CTK	95,5	93
			ELISA Chagas III	97,3	100
			Imuno-ELISA Chagas	99,5	99,2
			Gold ELISA Chagas	100	100
(ABRAS et al., 2016)	Dos pruebas serológicas	E.Chagas: 108 NoCha: 205	Pathozyme Chagas	99,5	97,0
			Architect Chagas	100	97,56
En esta investigación	Dos pruebas serológicas	E. Chagas: 118 NoCh: 92	RMN	96,7	90,8

E.Chagas: Enfermedad de Chagas. NoCh: No chagásicos. HAI: Hemaglutinación indirecta. IFI: inmunofluorescencia indirecta. ELISA: ensayos inmunoenzimáticos. PCR-ADNsat: reacción en cadena de la polimerasa ADN satélite. PCR-ADNk: reacción en cadena de la polimerasa ADN cinetoplasto

4.4 CONCLUSIONES

El análisis de componentes principales evidenció el agrupamiento intrínseco que presentan las muestras debido a su condición fisiopatológica, descartando como factores de confusión el género y la edad. Además, permitió la selección de muestras eliminando aquellas que contienen señales atípicas debido a posibles contaminaciones.

El perfil metabólico del suero sanguíneo procedente de pacientes con enfermedad de Chagas obtenido por RMN permitió construir un modelo de clasificación PLS-DA, el cual permite discriminar pacientes enfermos de pacientes sanos, con 90,8% de especificidad, 96,7% de sensibilidad, 94 % de precisión y un error de clasificación de 6%.

5 IDENTIFICACIÓN DE LOS POSIBLES METABOLITOS CON POTENCIAL COMO BIOMARCADORES PARA MÉTODOS DE DIAGNÓSTICO CLÍNICO DE INFECCIÓN POR *T. cruzi*.

5.1 INTRODUCCIÓN

El descubrimiento de biomarcadores ha jugado un rol importante en la medicina, ayudando a mejorar el diagnóstico y pronóstico de diferentes patologías. El descubrimiento de nuevos biomarcadores implica una serie de etapas rigurosas de ensayos y estudios hasta lograr su aplicación. La primera etapa consiste en la búsqueda de los metabolitos diferenciadores en particular para comparaciones bidireccionales, sanos versus enfermos ó antes de la dosificación versus después de la dosificación. Una vez se hallan estos metabolitos diferenciadores se procede a realizar suficientes pruebas de precisión, eliminación de interferencias químicas o biológicas, para promover el/los metabolitos candidatos a biomarcadores, los cuales deben probarse y demostrar utilidad para continuar en evaluación o descartarse (KOULMAN et al., 2009; STRIMBU; TAVEL, 2011).

Los estudios metabolómicos han sido útiles en el descubrimiento de nuevos biomarcadores encontrando metabolitos que permiten diferenciar entre dos o más estados fisiopatológicos (ZHANG et al., 2013). A pesar del gran potencial de la metabolómica para hallar metabolitos diferenciadores, la identificación de los metabolitos sigue siendo el cuello de botella. Actualmente, se han desarrollado un gran número de bases de datos de RMN, entre las que se encuentran: Base de Datos del Metaboloma Humano (*Human Metabolome Data Base, HMDB*), Banco de datos biológico de Resonancia Magnética Nuclear (*Biological Magnetic Resonance Data Bank, BioMagResBank*), Bases de datos de RMN de Linkoping (*NMR database of Linkoping*), Base de Datos Metabolómica de Resonancia Magnética Nuclear (*Magnetic Resonance Metabolomics Database*), entre otras. Estas bases de datos proporcionan la información necesaria, espectros de RMN de referencia,

para la identificación de metabolitos y facilitan la elucidación de las vías biológicas pertinentes (ZHANG et al., 2013). Una vez se identifica el conjunto de metabolitos es necesario conferir un sentido biológico a los datos adquiridos y determinar las funciones de los metabolitos en el sistema biológico. Para ello se debe comparar los datos con estudios previos y además correlacionar la información del estudio metabolómico con la obtenida por otras ciencias 'ómicas' como genómica y proteómica. Los niveles de metabolitos que se encuentran en el huésped durante la infección por *T. cruzi* podrían proporcionar información valiosa sobre la patogénesis identificando la ruta metabólica desregulada que podría ser blanco potencial para el tratamiento o servir como biomarcadores pronósticos del desarrollo y severidad de la enfermedad. Con este propósito, se asignaron los posibles metabolitos relacionados con las variables de mayor poder de discriminación en el modelo PLS-DA.

5.2 METODOLOGÍA

Para una interpretación bioquímica de las variaciones espectrales observadas, se intentó asignar los *buckets* a sus respectivos metabolitos. Con ese fin, cada *bucket* se representó por su rango de desplazamiento químico que luego se relacionó con un metabolito basado en la literatura (EMBADE et al., 2016; JOBARD et al., 2014; MACINTYRE et al., 2010; MAITRE et al., 2014; NICHOLSON et al., 1984; RANKIN et al., 2014) y las bases de datos HMDB y *BioMagResBank*. Por otra parte, se relacionó la fluctuación de los metabolitos identificados como relevantes en la discriminación del modelo $M_{SME-VIP}$ con estudios previos relacionados con la EC.

5.3 RESULTADOS Y DISCUSIÓN

De acuerdo con los resultados obtenidos se asignaron los metabolitos presentes en las muestras de suero sanguíneo tal como lo ilustran la **Figura 23** y **Figura 24**. La

Tabla 10 resume los metabolitos asignados con la respectiva multiplicidad de la señal, y la región espectral que se utilizó en el análisis multivariado, así mismo con el estadístico VIP obtenido en el modelo $M_{SME-VIP}$ y el incremento o disminución relativa del metabolito.

.

Tabla 10. Metabolitos identificados en los espectros de RMN-H de sueros sanguíneos. Elaboración propia.

Metabolito	$\delta^1\text{H}$ ppm	Asignación	Región bucket VIP	Variación		VIP
				Healthy	Infected	
Valina	0,95(d), 1,00(d)	CH ₃		-----	-----	<1
Lactato	1,30 (d), 4,09 (q)	CH ₃ , CH	4,071-4,089 4,089-4,111	Aumentó	Disminuyó	2,39 2,32
Alanina	1,45 (d)	CH ₃		-----	-----	<1
Acetato	1,89 (s)	CH ₃		-----	-----	<1
Glutamato	2,04(m), 2,32 (m)	CH ₃	2,057-2,063 2,063-2,070	Disminuyó	Aumentó	1,56 1,24
+ Grupos N- acetil de glicoproteínas	2,01 (m)		2,086-2,092 2,092-2,095 2,103-2,108 2,128-2,133	Aumentó	Disminuyó	1,47 1,54 1,80 1,49
Acetoacetato	2,19(s)	CH ₃		-----	-----	<1
Sarcosina	2,71 (s)	CH ₃		-----	-----	<1
Creatina	3,01(s), 3,90(s)	CH ₃ , CH ₂		-----	-----	<1
Colina	3,17 (s)	N(CH ₃) ₃		-----	-----	<1
Malonato	3,12 (s)	CH ₂		-----	-----	<1
Glicina	3,52 (s)	CH ₂		-----	-----	<1
Glucosa	3,19-3,90 (m)	Varios H	3,531-3,538 3,451-3,458	Disminuyó Aumentó	Aumentó Disminuyó	1,70 1,38
Treonina	3,57 (d)	CH	3,587-3,592 3,583-3,587 3,571-3,583	Aumentó	Disminuyó	1,40 1,42 1,32
α Glucosa	5,21 (d)	H2		-----	-----	<1
β Glucosa	4,62 (d)	H2	4,602-4,622 4,622-4,643	Aumentó	Disminuyó	2,62 2,55
Histidina	7,70(s)			-----	-----	<1
Formiato	8,43 (s)	CH	8,421-8,432	Disminuyó	Aumentó	1,42

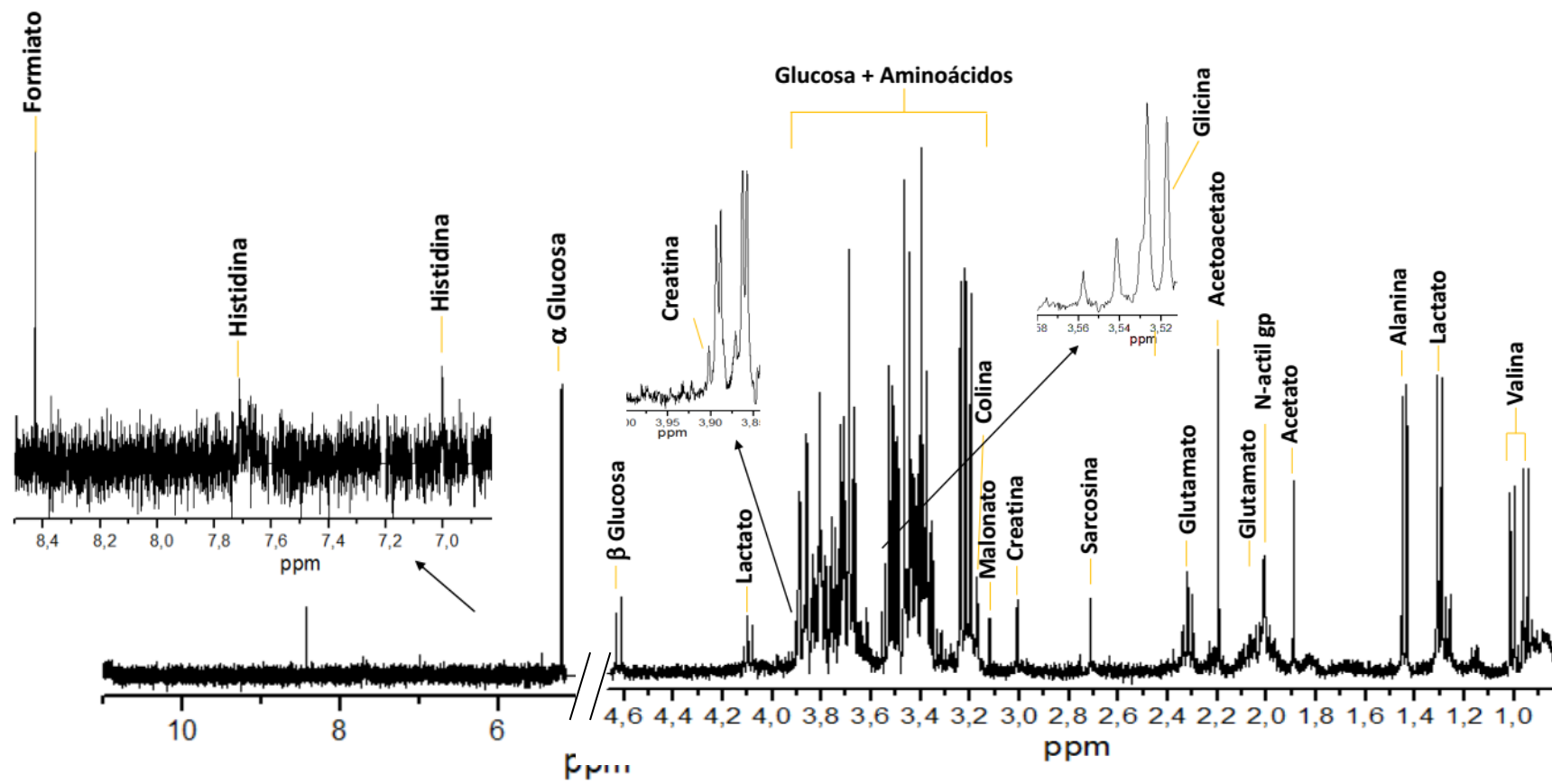


Figura 23. Asignación de metabolitos para un espectro de RMN- ^1H obtenido para una muestra de suero sanguíneo con diagnóstico seropositivo. *Izquierda*) Región aromática. *Derecha*) Región alifática. Elaboración propia.

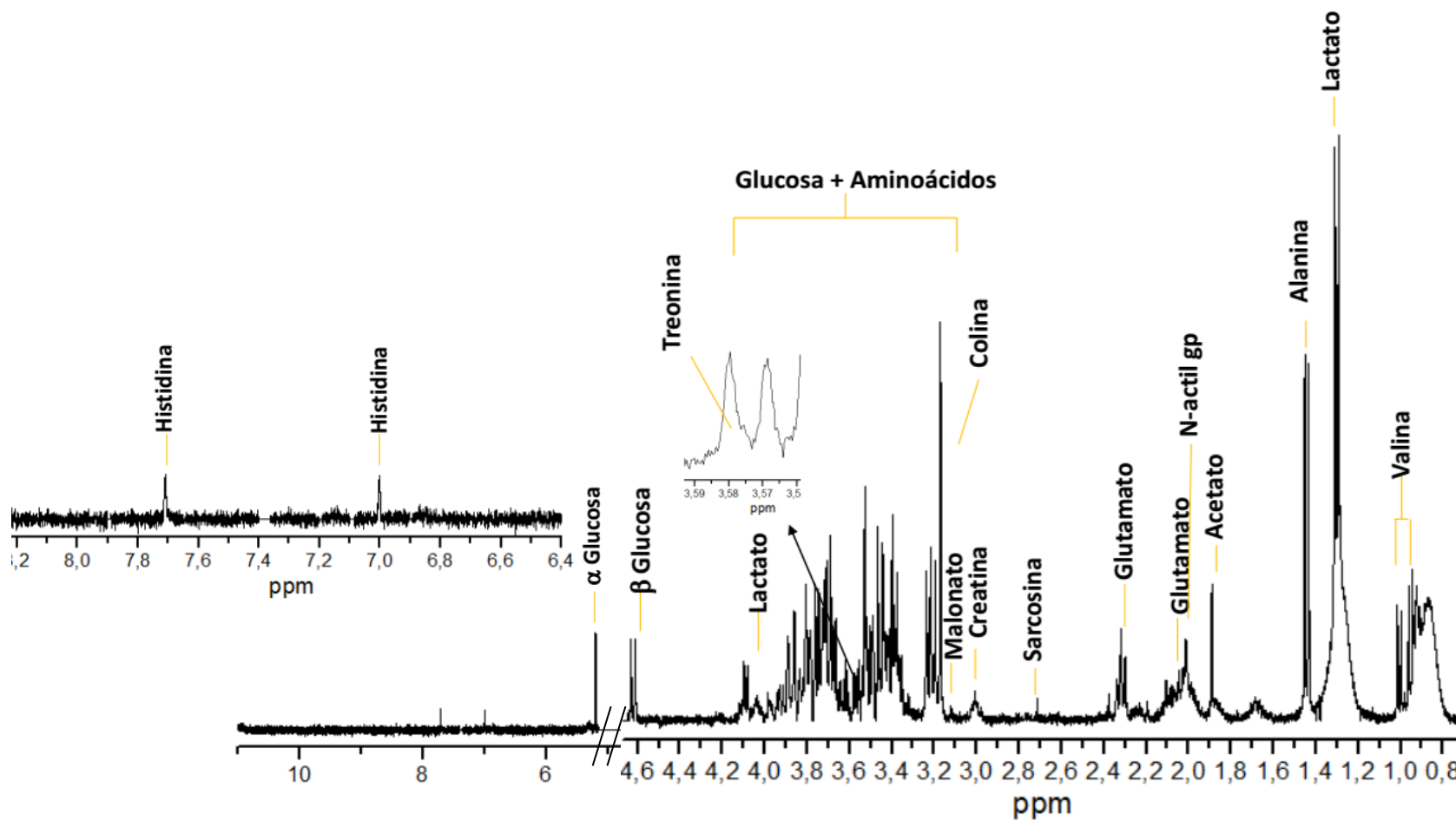


Figura 24 Asignación de metabolitos para un espectro de RMN- ^1H obtenido para una muestra de suero sanguíneo con diagnóstico seronegativo. *Izquierda*) Región aromática. *Derecha*) Región alifática

Los espectros están dominados por las señales de glucosa, especialmente en la región comprendida entre 3,19-3,90 ppm, este hecho dificulta la identificación de metabolitos cuyas señales características se encuentren en estas regiones. De las 29 regiones que tiene un valor VIP>1 para el modelo M_{SME-VIP} (**Tabla 7**) se lograron asignar 18 regiones a las señales correspondientes a metabolitos, es decir, el 62% de las regiones. De acuerdo con el estadístico VIP se encontraron los posibles metabolitos que contribuyen principalmente a la separación de los dos grupos en el modelo PLS-DA:

Glucosa

Las regiones identificadas con un valor VIP>1 para la glucosa fueron, principalmente los dos *buckets* del doblete comprendido entre las regiones 4,602-4,643 ppm donde se observó una disminución en los pacientes infectados respecto a los sanos. Además, la región 3,531-3,538 ppm donde se observó una disminución para los pacientes infectados y la región 3,451-3,458 ppm donde se observó un incremento en los pacientes infectados. Aunque estas dos últimas regiones fueron asignadas a la glucosa, se debe considerar que existe una alta superposición de señales de glucosa con señales correspondientes a otros metabolitos.

Dentro de las rutas biológicas donde posiblemente esté involucrada la glucosa se encuentra la ruta de la glucólisis/gluconeogénesis. Estos resultados podrían ser atribuidos a que *T. cruzi* no tiene una reserva intracelular de glucosa como glucógeno o almidón, sino que debe tomarla del medio, como la sangre del mamífero, a través de un sistema de transporte de glucosa. De acuerdo con un estudio realizado por Silber (SILBER et al., 2009) sobre la captación de glucosa de *T. cruzi* en el ciclo de vida del huésped mamífero, se encontró que los tripomastigotes presentan una actividad elevada de captación de glucosa con respecto a la etapa intracelular (amastigote), evidenciando que los tripomastigotes dependen de la glucosa del hospedero para sus requerimientos de energía. Recientemente, Gironès y colaboradores (GIRONÈS et al., 2014) realizaron análisis

metabolómico del corazón y plasma sanguíneo de ratones infectados con *T. cruzi*, entre sus hallazgos se logró identificar un aumento de glucosa en el tejido cardíaco y una disminución de la misma en el plasma de los ratones infectados. Resultados acordes con los obtenidos durante esta investigación, donde los niveles de glucosa en suero sanguíneo disminuyen en el grupo infectado con respecto al grupo control. En otros estudios se han encontrado lesiones a nivel de páncreas en pacientes con EC crónica, de manera que podría verse afectada la regulación de glucosa (GUARIENTO et al., 1993; ROCHA et al., 1998).

Glutamato-Glicoproteínas

El glutamato presenta un multiplete en la región ~1,98-2,15 ppm. Sin embargo, en esa región también se presentan señales correspondientes a las N-acetil glicoproteínas. Se identificaron dos variables con valor VIP>1, comprendidas entre 2,057-2,063 y 2,063-2,070 ppm en los cuales se presenta un aumento de este metabolito con respecto al grupo de pacientes sanos. Por otra parte, en esta región se encuentran cuatro regiones más en las cuales se observó una disminución de los metabolitos con respecto al grupo control, estas variables se encuentran 2,086-2,092, 2,092-2,095, 2,103-2,108 y 2,128-2,133 ppm. De acuerdo con lo revisado en la literatura, los aminoácidos tienen un rol importante durante el ciclo de vida de *T. cruzi* como fuente de energía y participan en diversas rutas biológicas que ayudan al parásito a adaptarse a los cambios de ambiente (SILBER et al., 2005). El glutamato tiene gran importancia en el ciclo de vida del *T. cruzi*, se ha demostrado que el glutamato estimula la metacicloogénesis (HOMSY; GRANGER; KRASSNER, 1989). La absorción de glutamato se produce a través de un único sistema de transporte cinético, que es impulsado por el gradiente de concentración de H⁺ de la membrana citoplásmica (SILBER et al., 2006). Una vez en el interior de la célula el glutamato puede ser convertido en α -cetoglutarato por glutamato deshidrogenasa y ser procesado vía ciclo de Krebs para la obtención de energía (SILBER et al., 2005). Por otra parte, la respuesta del hospedero a la infección persistente de *T. cruzi* la generación sostenida de especies reactivas de oxígeno y nitrógeno, el *T. cruzi* utiliza

el tripanotión como vía de defensa antioxidante (ARIYANAYAGAM; FAIRLAMB, 2001) . En este sentido, el glutamato también puede servir como sustrato para la biosíntesis de glutatión y tripanotión vía γ -glutamylcisteína sintasa y glutatión sintasa (TELLERIA; TIBAYRENC, 2010). Por esta razón, se ha propuesto dirigir fármacos contra enzimas del metabolismo de glutamato como una estrategia tripanocida novedosa (DAMASCENO et al., 2014; MAGDALENO et al., 2011). Sin embargo, no se puede establecer la fluctuación del metabolito asociada a la EC debido a que las señales se encuentran superpuestas con las señales de la N-glicoproteínas, tampoco es claro si hay un efecto sinérgico debido a que en dos de las cuatro regiones la concentración relativa aumenta, mientras en las cuatro restantes la concentración disminuye.

Lactato

El lactato presenta un cuádruplete en la región ~4,06-4,12 ppm. En esta región se hallaron tres variables con un valor VIP>1, comprendidas entre 4,061-4,066, 4,071-4,089 y 4,089-4,111 ppm

Formiato

Presenta un singlete en 8,45, en esta región se encontró una variable con valor VIP>1, entre 8,421-8,432 ppm, la cual evidencia un incremento de este metabolito con respecto al grupo control.

Treonina

Presenta un doblete en ~3,57 ppm, donde se hallaron tres regiones con un valor VIP>1, 3,571-3,583, 3,583-3,587, y 3,587-3,592 ppm. Este metabolito presentó una disminución con respecto al grupo control.

De últimos tres metabolitos no se encontró información relacionada con la EC.

Adicionalmente, se realizó una revisión de los biomarcadores para la EC en suero sanguíneo. Ndao y colaboradores (NDAO et al., 2010) estudiaron espectrometría de masas para buscar biomarcadores en sueros sanguíneos de pacientes asintomáticos y encontraron como posibles biomarcadores la proteína C3a cuya función principal es la moderación y activación de vías inflamatorias y la proteína ApoA1 que es el principal componente proteico estructural y funcional de alta densidad, estimula el transporte inverso del colesterol e interviene en la estructura. Santamaría y colaboradores (SANTAMARIA et al., 2014) estudiaron biomarcadores séricos predictivos de curación en pacientes con EC después del tratamiento con nifurtimox y encontraron como biomarcadores fragmentos de la proteína ApoA1 y la proteína FN1 para discriminar pacientes enfermos de pacientes sanos y basados en estos biomarcadores poder evaluar la acción del medicamento. No se logró establecer una relación entre las proteínas con potencial como biomarcadores en los estudios y los metabolitos encontrados durante el desarrollo de esta investigación.

En el estudio de Gironès y colaboradores (GIRONÈS et al., 2014) realizado en plasma y tejido cardíaco de ratones infectados con *T. cruzi* utilizando cromatografía líquida acoplada a masas, además de cambios en la utilización de glucosa como ya se mencionó, se encontraron alteraciones en el metabolismo de aminoácidos, en el ciclo de Krebs y membrana lipídica. En el tejido cardíaco se encontró un aumento de isoleucina, leucina y valina. Mientras que en el tejido y plasma sanguíneo se halló un aumento de 2-butil-butil carnitina e isovalerilcarnitina. Los ácidos grasos de cadena larga como, palmato, estereato y oleato aumentaron en el tejido cardíaco pero disminuyeron en plasma. En nuestra investigación no se observaron cambios en estos metabolitos, probablemente porque se analizaron muestras en fase crónica. Con respecto a los demás metabolitos no se logró establecer una relación ya que estas fluctuaciones no se observaron en la presente investigación. Esto puede deberse a la sensibilidad de las técnicas empleadas y las diferencias en la preparación de la muestra. Con RMN sólo se pueden observar los metabolitos

mayoritarios presentes en el sueros debido a la baja sensibilidad (WISHART, 2008). Por otro lado, durante la presente investigación se trabajó con el suero completo, sin realizar ningún tratamiento, mientras que en el estudio de Gironés el plasma recibió un tratamiento térmico y extracción con solventes para el posterior análisis. Además, son hospederos diferentes y en algunos casos no se pueden extrapolar los resultados de modelos animales al humano.

5.4 CONCLUSIÓN

Se identificaron quince metabolitos en el espectro RMN-¹H, de los cuales la glucosa, lactato, formiato, treonina y glutamato presentan fluctuaciones en los pacientes con EC con respecto al grupo control, los cuales podrían tener un potencial como biomarcadores para desarrollar un método de diagnóstico clínico.

6 CONCLUSIÓN GENERAL

Los resultados demostraron el potencial de RMN-¹H como herramienta analítica práctica para la discriminación de pacientes infectados con *T. cruzi* de pacientes sanos utilizando el suero sanguíneo sin ningún tratamiento previo. Además, cabe destacar que se logró la clasificación de los pacientes utilizando información del metaboloma, en contraste con los métodos actuales que emplean la interacción antígeno anticuerpo para clasificar el paciente de acuerdo con su estado fisiopatológico.

7 RECOMENDACIONES

- Realizar experimentos adicionales de RMN en dos dimensiones, tal como, 2D ^1H JRES, COSY y TOCSY para confirmar la identidad de los metabolitos identificados por RMN- ^1H .
- Validar los resultados obtenidos en esta investigación evaluando un nuevo grupo de estudio, con el fin de verificar que los resultados obtenidos están asociados a la EC y no a factores de confusión sobre los que no se tiene control, como la dieta alimenticia de los individuos y/o hábitos de vida.
- Comparar los resultados obtenidos por RMN con otras técnicas analíticas como espectrometría de masas.
- Utilizar esta herramienta para la búsqueda de biomarcadores asociados con la patogénesis y que permitan determinar susceptibilidad al desarrollo de la fase sintomática y severa de la enfermedad.

8 ASPECTOS ÉTICOS

De acuerdo con la norma internacional establecida en el informe BELMONT (febrero, 1976) donde se estipulan los principios y guías éticas para la protección de los sujetos humanos de investigación y teniendo en cuenta la Resolución 008430 de Octubre 4 de 1993 del Ministerio de Salud de Colombia, en la cual se establecen las normas científicas, técnicas y administrativas para la investigación en salud y en cumplimiento con los aspectos mencionados de la presente Resolución se tendrán en cuenta los siguientes criterios:

Este estudio se considera de **riesgo mínimo** para los participantes de acuerdo al artículo 11, numeral “b”, porque se trata de un estudio observacional que emplea el registro de datos a través de procedimientos comunes como una encuesta, exámenes físicos y la extracción de sangre por punción venosa.

El principio de *respeto y autonomía*, teniendo en cuenta que cada persona es autónoma, es decir opta por lo que considere que es mejor para él en concordancia con sus intereses, deseos y creencias y dado que la libertad personal es inviolable los investigadores tienen la obligación de respetar los valores y opciones personales de cada individuo en aquellas decisiones básicas que le atañen. En este estudio se garantizarán estos principios a los participantes cuando se les ofrece la libertad de participar o no en el estudio y de retirarse en el momento que lo deseen. Así mismo, se les brindará toda la información (el procedimiento de la investigación, sus propósitos, riesgos y beneficios anticipados, procedimientos alternos) se responderán todas las inquietudes y se contará con el Consentimiento Informado y por escrito del participante en la investigación o su representante legal con las excepciones dispuestas en la presente resolución.

El principio de *beneficencia y no maleficencia*, se refiere a la obligación de prevenir o aliviar el daño, hacer el bien u otorgar beneficios, deber de ayudar al

prójimo por encima de los intereses particulares, en otras palabras, obrar en función del mayor beneficio posible para el paciente procurando el bienestar de la persona enferma. En este trabajo se garantizará el principio de beneficencia dado que está dirigido hacia la búsqueda de una solución para las personas afectadas por la enfermedad de Chagas, que permita optimizar el sistema de prevención y diagnóstico temprano de enfermedad de Chagas en las regiones endémicas de Santander, lo que impactaría a mediano plazo en la salud de los participantes, así como de la población en general del departamento.

El principio de *justicia* se tendrá en cuenta dado que ninguno de los posibles participantes será objeto de discriminación por cualquier circunstancia.

También, se hará énfasis, para este estudio, **en la *confidencialidad y privacidad* de la información**, evitando utilizar el nombre de la persona tanto en las encuestas como en las bases de datos. Se asignará un código de identificación para cada paciente. Además, los registros serán custodiados por los investigadores, lo que impide que sean de dominio público. Los resultados no estarán disponibles para terceras personas a menos que el participante autorice o la Ley lo permita. En todo caso se cumplirá con lo reglamentado en la Ley estatutaria 1581 de 2012 sobre el tratamiento de datos personales.

Aspectos relacionados con la bioseguridad de los investigadores, en este proyecto se cumplió con los protocolos institucionales de bioseguridad y protección de los investigadores los cuales están consignados en el manual de procedimientos para los laboratorios de microbiología (UIS). El personal involucrado en la investigación es idóneo y competente en el manejo manipulación, transporte, utilización de muestras sanguíneas y descontaminación y eliminación de desechos.

Disposición adecuada de los residuos químicos y/o biológicos teniendo en cuenta el sistema de gestión interno:

La metodología descrita para el desarrollo de esta investigación permite que se generen algunos residuos o químicos y biológicos, que deben ser eliminados de una manera ambientalmente responsable. Por tanto, todo residuo o desecho generado en este proyecto fue debidamente tratado de acuerdo con la reglamentación institucional de etiquetado y embalaje para que pueda ser dispuesto al PGIR (Plan de Gestión Integral de Residuos en la Universidad), órgano institucional que se encarga de velar por la aplicación de estrategias preventivas y cumpliendo con las normas legales ambientales para el manejo adecuado de los residuos peligrosos.

REFERENCIAS BIBLIOGRÁFICAS

ABRAS, Alba. et al. Serological Diagnosis of Chronic Chagas Disease : Is It Time for a Change? *Journal of clinical microbiology*. 2016. v. 54, n. 6, p. 1566–1572.

ANDERSEN, C. M.; BRO, R. Variable selection in regression-a tutorial. *Journal of Chemometrics*. 2010. v. 24, n. 11–12, p. 728–737.

ARIYANAYAGAM, Mark; FAIRLAMB, Alan. Ovothiol and trypanothione as antioxidants in trypanosomatids. *Molecular and Biochemical Parasitology*. 2001. v. 115, n. 2, p. 189–198.

BECKONERT, Olaf. et al. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nature protocols*. 2007. v. 2, n. 11, p. 2692–2703.

BERN, Caryn. Chagas' Disease. *New England Journal of Medicine*. 2015 v. 373, n. 5, p. 456–466.

BHARTI, Santosh Kumar. et al. Improved quantification from ¹H-NMR spectra using reduced repetition times. *Metabolomics*. 2008. v. 4, n. 4, p. 367–376.

BIJLSMA, Sabina. et al. Large-scale human metabolomics studies: A strategy for data (pre-) processing and validation. *Analytical Chemistry*. 2006. v. 78, n. 2, p. 567–574.

BLASCO, Helene et al. ¹H-NMR-Based metabolomic profiling of CSF in early amyotrophic lateral sclerosis. *PLoS ONE*. 2010 v. 5, n. 10.

BRASIL, Pedro. et al. ELISA versus PCR for diagnosis of chronic Chagas disease:

systematic review and meta-analysis. BMC infectious diseases. 2010. v. 10, p. 337.

BRINDLE, Joanne. et al. Application of chemometrics to ¹H NMR spectroscopic data to investigate a relationship between human serum metabolic profiles and hypertension. The Analyst. 2003. v. 128, n. 1, p. 32–36.

CABALLERO, Zuleima. et al. Evaluation of serological tests to identify Trypanosoma cruzi infection in humans and determine cross-reactivity with Trypanosoma rangeli and Leishmania spp. Clinical and Vaccine Immunology. 2007. v. 14, n. 8, p. 1045–1049.

CASTELLANOS, Yenny Zulay. et al. Reproducibility of serological tests for the diagnosis of Trypanosoma cruzi infection in pregnant women in an endemic area of Santander, Colombia. Biomédica : revista del Instituto Nacional de Salud. 2014. v. 34, n. 2, p. 198–206.

CERVANTES-LANDÍN, Alejandra et al. Estandarización de la técnica Dot-ELISA para la detección de anticuerpos anti-Trypanosoma cruzi y su comparación con ELISA y Western blot. Enfermedades Infecciosas y Microbiología Clínica. 2014 v. 32, n. 6, p. 363–368.

DAMASCENO, Flavia Silvia. et al. Memantine, an Antagonist of the NMDA Glutamate Receptor, Affects Cell Proliferation, Differentiation and the Intracellular Cycle and Induces Apoptosis in Trypanosoma cruzi. PLoS Neglected Tropical Diseases. 2014. v. 8, n. 2.

DAWISKIBA, Tomas. et al. Serum and urine metabolomic fingerprinting in diagnostics of inflammatory bowel diseases. World Journal of Gastroenterology. 2014. v. 20, n. 1, p. 163–174.

DIETERLE, Frank. et al. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in ^1H NMR metabonomics. *Analytical Chemistry*. 2006. v. 78, n. 13, p. 4281–4290.

DUNN, Warwick; ELLIS, David. Metabolomics: Current analytical platforms and methodologies. *TrAC - Trends in Analytical Chemistry*. 2005 v. 24, n. 4, p. 285–294.

EMBADE, Nieves. et al. Metabolic characterization of advanced liver fibrosis in HCV patients as studied by serum ^1H -NMR spectroscopy. *PLoS ONE*. 2016. v. 11, n. 5, p. 1–19.

ENEA, C. et al. ^1H NMR-based metabolomics approach for exploring urinary metabolome modifications after acute and chronic physical exercise. *Analytical and Bioanalytical Chemistry*. 2010. v. 396, n. 3, p. 1167–1176.

FARRÉS, Mireia. et al. Comparison of the variable importance in projection (VIP) and of the selectivity ratio (SR) methods for variable selection and interpretation. *Journal of Chemometrics*. 2015. v. 29, n. 10, p. 528–536.

FAWCETT, Tom. An introduction to ROC analysis. *Pattern Recognition Letters*. 2006. v. 27, n. 8, p. 861–874.

FERNÁNDEZ, Juan Antonio. et al. A Backward Variable Selection method for PLS regression (BVSPLS). *Analytica Chimica Acta*. 2009. v. 642, n. 1–2, p. 89–93.

FERRER, Elizabeth. et al. Comparación entre técnicas inmunológicas y moleculares para el diagnóstico de la enfermedad de Chagas. *Enfermedades Infecciosas y Microbiología Clínica*. 2012. v. 31, n. 5, p. 277–282.

FERRERA, Elizabeth. et al. Comparación entre técnicas inmunológicas y moleculares para el diagnóstico de la enfermedad de Chagas. *Enfermedades Infecciosas y Microbiología Clínica*. 2013. v. 31, n. 5, p. 277–282.

FLORES-CHÁVEZ, María. et al. Comparación de técnicas serológicas convencionales y no convencionales para el diagnóstico de la enfermedad de Chagas importada en España. *Enfermedades Infecciosas y Microbiología Clínica*, 2010 v. 28, n. 5, p. 284–293.

GABBANI, Tommaso. et al. Metabolomic analysis with ¹H-NMR for non-invasive diagnosis of hepatic fibrosis degree in patients with chronic hepatitis C. *Digestive and Liver Disease*. 2017. v. 49, n. 12, p. 1338–1344.

GIRONÈS, Nuria. et al. Global Metabolomic Profiling of Acute Myocarditis Caused by *Trypanosoma cruzi* Infection. *PLoS Neglected Tropical Diseases*. 2014. v. 8, n. 11, p. 1–11.

GUARIENTO, M. E. et al. Heterogenous insulin response to an oral glucose load by patients with the indeterminate clinical form of Chagas' disease. *Brazilian journal of medical and biological research* 1993. v. 26, n. 5, p. 491—495.

GULERIA, Anupam. et al. NMR based serum metabolomics reveals a distinctive signature in patients with Lupus Nephritis. *Scientific Reports*. 2016. v. 6, n. October, p. 1–11.

GUPTA, Ashis.; GUPTA, Shalini.; MAHDI, Abbas. ¹H NMR-derived serum metabolomics of leukoplakia and squamous cell carcinoma. *Clinica Chimica Acta*. 2015. v. 441, p. 47–55.

HOMSY, James.; GRANGER, Barbara.; KRASSNER, Stuart. Some Factors

Inducing Formation of Metacyclic Stages of *Trypanosoma cruzi*. The Journal of Protozoology. 1989 v. 36, n. 2, p. 150–153.

HORE, P. J. Solvent suppression in fourier transform nuclear magnetic resonance. Journal of Magnetic Resonance. 1983. v. 55, n. 2, p. 283–300.

JACOB, Daniels. et al. NMRProcFlow: a graphical and interactive tool dedicated to 1D spectra processing for NMR-based metabolomics. Metabolomics. 2017. v. 13, n. 4, p. 1–5.

JOBARD, Elodie. et al. A serum nuclear magnetic resonance-based metabolomic signature of advanced metastatic human breast cancer. Cancer Letters. 2014. v. 343, n. 1, p. 33–41.

KAY, Lewis. Field gradient techniques in NMR spectroscopy. Current Opinion in Structural Biology. 1995. v. 5, n. 5, p. 674–681.

KESHARI, Kayvan; WILSON, David. Chemistry and biochemistry of ^{13}C hyperpolarized magnetic resonance using dynamic nuclear polarization. v. 43, p. 1627-1659.

KOULMAN, Albetr. et al. From differentiating metabolites to biomarkers. Analytical and Bioanalytical Chemistry. 2009. v. 394, n. 3, p. 663–670.

LI, Juan. et al. Impacts of the callipyge mutation on ovine plasma metabolites and muscle fibre type. PLoS ONE. 2014. v. 9, n. 6, p. 1–15.

LIU, Maili.; NICHOLSON, Jeremy.; LINDON, John. High-Resolution Diffusion and Relaxation Edited One- and Two-Dimensional ^1H NMR Spectroscopy of Biological Fluids. Analytical chemistry. 1996. v. 68, n. 19, p. 3370–3376.

MACINTYRE, D. A. et al. Serum metabolome analysis by ^1H -NMR reveals differences between chronic lymphocytic leukaemia molecular subgroups. *Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund U.K.* 2010. v. 24, n. 4, p. 788–797.

MADSEN, Rasmus.; LUNDSTEDT, Torbjorn.; TRYGG, Johan. Chemometrics in metabolomics-A review in human disease diagnosis. *Analytica Chimica Acta.* 2010. v. 659, n. 1–2, p. 23–33.

MAGDALENO, Anahí. et al. The Involvement of Glutamate Metabolism in the Resistance to Thermal, Nutritional, and Oxidative Stress in *Trypanosoma cruzi*. *Enzyme Research.* 2011. v. 2011, p. 1–13.

MAITRE, Lea. et al. Urinary metabolic profiles in early pregnancy are associated with preterm birth and fetal growth restriction in the Rhea mother-child cohort study. *BMC Medicine.* 2014. v. 12, n. 1, p. 1–14.

MANRIQUE-ABRIL, Fred. et al. Diagnóstico de enfermedad de Chagas en maternas y recién nacidos de Moniquirá y Miraflores, Boyacá, Colombia. *Infectio.* 2013. v. 17, n. 1, p. 28–34.

MANTILLA, Julio Cesar. Enfermedad de Chagas : correlación clínico-patológica . Serie de casos del Hospital Universitario de Santander - Departamento de Patología , Universidad Industrial de Santander. *Revista Colombiana de Cardiología.* 2011. v. 18, n. 5, p. 249–261.

MEIBOOM, S.; GILL, D. Modified spin-echo method for measuring nuclear relaxation times. *Review of Scientific Instruments.* 1958. v. 29, n. 8, p. 688–691.

MOLINA, Israel.; SALVADOR, Fernando; SÁNCHEZ-MONTALVÁ, Adrian.

Actualización en enfermedad de Chagas. *Enfermedades Infecciosas y Microbiología Clínica*. 2016. v. 34, n. 2, p. 132–138.

MONCAYO, Álvaro.´; SILVEIRA, Antonio. Current epidemiological trends for Chagas disease in Latin America and future challenges in epidemiology, surveillance and health policy. *Memorias do Instituto Oswaldo Cruz*. 2009. v. 104, n. SUPPL. 1, p. 17–30.

MOURE, Z. et al. Serodiscordance in chronic Chagas disease diagnosis: a real problem in non-endemic countries. *Clinical Microbiology and Infection*. 2016. v. 22, n. 9, p. 788–792.

NAGANA, Gowda; RAFTERY, Daniel. Can NMR solve some significant challenges in metabolomics? *Journal of magnetic resonanc-* 2015. v. 260, p. 144–60.

NDAO, Momar. et al. Identification of novel diagnostic serum biomarkers for chagas' disease in asymptomatic subjects by mass spectrometric profiling. *Journal of Clinical Microbiology*. 2010. v. 48, n. 4, p. 1139–1149.

NEVES, Fred. et al. Chronic Chagas Disease Diagnosis: A Comparative Performance of Commercial Enzyme Immunoassay Tests. *The American Journal of Tropical Medicine and Hygiene*. 2016. v. 94, n. 5, p. 1034–1039.

NICHOLSON, Jeremy. et al. Proton-Nuclear-Magnetic-Resonance Studies of Serum, Plasma and Urine From Fasting Normal and Diabetic Patients. *Biochemical Journal*. 1984. v. 217, p. 365–375.

NICHOLSON, Jeremt. K. et al. 750-Mhz H-1 and H-1-C-13 Nmr-Spectroscopy of Human Blood-Plasma. *Analytical Chemistry*. 1995. v. 67, n. 2, p. 793–811.

OPS. ESTIMACIÓN CUANTITATIVA DE LA ENFERMEDAD DE CHAGAS EN LAS AMÉRICAS. Organización Mundial de la Salud. 2006. p. 1–28.

PALISI, Angélica. et al. A serum nuclear magnetic resonance-based metabolomic signature of antiphospholipid syndrome. *Journal of Pharmaceutical and Biomedical Analysis*. 2017. v. 133, p. 90–95.

PEREIRA, Maria. et al. Chagas disease: an overview of clinical and epidemiological aspects. *Journal of the American College of Cardiology*. 2013. v. 62, n. 9, p. 767–76.

PONCE, Carlos. et al. Validation of a Rapid and Reliable Test for Diagnosis of Chagas Disease by Detection of *Trypanosoma cruzi* -Specific Antibodies in Blood of Donors and Patients in Central America Validation of a Rapid and Reliable Test for Diagnosis of Chagas ' Disease. *Journal of clinical microbiology*. 2005. v. 43, n. 10, p. 5065–5068.

RANKIN, Naomi. et al. The emergence of proton nuclear magnetic resonance metabolomics in the cardiovascular arena as viewed from a clinical perspective. *Atherosclerosis*. 2014. v. 237, n. 1, p. 287–300.

RASSI, Anis Jr.; RASSI, Anis.; MARIN-NETO, Jose Antonio. Chagas disease. *The Lancet*. 2010. v. 375, n. 9723, p. 1388–1402.

RASTRELLI, Federico; JHA, Satadru; MANCIN, Fabrizio. Seeing through macromolecules: T2-filtered NMR for the purity assay of functionalized nanosystems and the screening of biofluids. *Journal of the American Chemical Society*. 2009. v. 131, n. 40, p. 14222–14224.

REYES, Marlene. et al. Riesgo de transmisión de la enfermedad de Chagas por

intrusión de triatominos y mamíferos silvestres en la ciudad de Bucaramanga (Santander, Colombia). *Biomédica Revista del Instituto Nacional de Salud*. 2017. v. 33, n. 4, p. 24.

ROCA, Carme. et al. Documento de consenso sobre el abordaje de la enfermedad de Chagas en atención primaria de salud de áreas no endémicas. *Atención Primaria*. 2015. v. 47, n. 5, p. 308–317.

ROCHA, Ademir. et al. Despopulação neuronal pancreática em chagásicos crônicos
Neuronal pancreatic depopulation in chronic Chagas ' disease patients. *Revista da Sociedade Brasileira de Medicina Tropical* . 1998. v. 31, n. 1, p. 43–49.

ROSAS, Fernando. Enfermedad de Chagas. *Revista Colombiana de Cardiología*. v. 18, p. 186–317, 2011.

SANTAMARIA, Cynthia. et al. Serum biomarkers predictive of cure in Chagas disease patients after nifurtimox treatment. *BMC Infectious Diseases*. 2014. v. 14, n. 1, p. 302.

SERKOVA, Natalie. J.; NIEMANN, Claus. Pattern recognition and biomarker H-NMR-based metabolomics. *Expert Review of Molecular Diagnostics*. 2006. p. 717–731.

SILBER, Ariel. et al. Amino Acid Metabolic Routes in *Trypanosoma cruzi*: Possible Therapeutic Targets Against Chagas; Disease. *Current Drug Target -Infectious Disorders*. 2005. v. 5, n. 1, p. 53–64.

SILBER, Ariel. et al. Biochemical characterization of the glutamate transport in *Trypanosoma cruzi*. *International Journal for Parasitology*. 2006. v. 36, n. 2, p. 157–163.

SILBER, Ariel. M. et al. Glucose uptake in the mammalian stages of *Trypanosoma cruzi*. *Molecular and Biochemical Parasitology*. 2009. v. 168, n. 1, p. 102–108.

SIMMLER, Charlotte. et al. Universal quantitative NMR analysis of complex natural samples. *Current Opinion in Biotechnology*. 2014. v. 25, p. 51–59.

SINCLAIR, Alexandra. et al. NMR-based metabolomic analysis of cerebrospinal fluid and serum in neurological diseases - A diagnostic tool? *NMR in Biomedicine*. 2010. v. 23, n. 2, p. 123–132.

SMOLINSKA, Agnieszka. et al. NMR and pattern recognition methods in metabolomics: From data acquisition to biomarker discovery: A review. *Analytica Chimica Acta*. 2012 v. 750, p. 82–97.

STRIMBU, Kyle.; TAVEL, Jorge. What are Biomarkers? *Curr Opin HIV AIDS*. 2011. v. 5, n. 6, p. 463–466.

SZYMAŃSKA, Ewa. et al. Double-check: Validation of diagnostic statistics for PLS-DA models in metabolomics studies. *Metabolomics*. 2012. v. 8, p. 3–16.

TANG, Huiru. et al. Use of relaxation-edited one-dimensional and two dimensional nuclear magnetic resonance spectroscopy to improve detection of small metabolites in blood plasma. *Analytical Biochemistry*. 2004. v. 325, n. 2, p. 260–272.

TANG, Huiru.; WANG, Yulan. High-Resolution NMR Spectroscopy in Human Metabolism and Metabonomics. *Modern Magnetic Resonance Part III*. 2006. p. 1623–1630, 2006.

TELLERIA, J.; TIBAYRENC, M. American trypanosomiasis Chagas disease. *One*

Hundred Years of Research. ELSEVIER INSIGHTS. 2010. p. 150-154

TIZIANI, Stefano.; LOPES, Victor.; GÜNTHER, Ulrich. Early stage diagnosis of oral cancer using ^1H NMR-based metabolomics. *Neoplasia*. 2009. v. 11, n. 3, p. 269–276.

TRIBA, Mohamed. et al. PLS/OPLS models in metabolomics: the impact of permutation of dataset rows on the K-fold cross-validation quality parameters. *Mol. BioSyst*. 2015. v. 11, n. 1, p. 13–19.

UNGER, Kristian. Integrative radiation systems biology. *Radiation oncology*. 2014., v. 9, n. 1, p. 21, 2014.

VIANT, Mark. Improved methods for the acquisition and interpretation of NMR metabolomic data. *Biochemical and Biophysical Research Communications*. 2003. v. 310, n. 3, p. 943–948.

WEN, He. et al. A new NMR-based metabolomics approach for the diagnosis of biliary tract cancer. *Journal of Hepatology*. 2010. v. 52, n. 2, p. 228–233.

WESTERHUIS, Johan. et al. Assessment of PLS-DA cross validation. *Metabolomics*. 2008. v. 4, n. 1, p. 81–89.

WISHART, David. Quantitative metabolomics using NMR. *TrAC - Trends in Analytical Chemistry*. 2008. v. 27, n. 3, p. 228–237.

WOLD, Svante.; SJÖSTRÖM, Michael.; ERIKSSON, Lennart. PLS-regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*. 2001. v. 58, n. 2, p. 109–130.

WORLEY, Bradley.; POWERS, Robert. Multivariate Analysis in Metabolomics. Current Metabolomics. 2013. v. 1, n. 1, p. 92–107.

XIA, Jianguo.; WISHART, David. Using MetaboAnalyst 3.0 for Comprehensive Metabolomics Data Analysis. Current Protocols in Bioinformatics. 201614.10.1-14.10.91

YAN, Bing. Analysis for Feasibility and Optimization of Library Synthesis With Nmr Spectroscopy. 2004. p. 230-250.

ZABEK, Adam. et al. Application of ¹H NMR-based serum metabolomic studies for monitoring female patients with rheumatoid arthritis. Journal of Pharmaceutical and Biomedical Analysis. 2016. v. 117, p. 544–550.

ZHANG, Aihua. et al. Recent and potential developments of biofluid analyses in metabolomics. Journal of Proteomics. 2012. v. 75, n. 4, p. 1079–1088.

ZHANG, Aihua. et al. NMR-based metabolomics coupled with pattern recognition methods in biomarker discovery and disease diagnosis. Magnetic Resonance in Chemistry. 2013. p. 549–556, 2013.

ZHENG, Hong. et al. Metabolic characterization of hepatitis B virus-related liver cirrhosis using NMR-based serum metabolomics. Metabolomics. 2017. v. 13, n. 10, p. 121.

ANEXOS

Anexo 1. Carta de autorización para el uso de las muestras por parte del investigador principal.

Bucaramanga, 17 de febrero de 2017

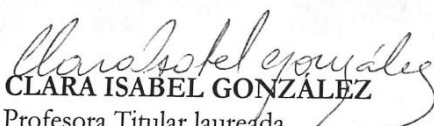
Señores
Comité de Ética en Investigación Científica, CEINCI
UIS/Presente

Asunto: Autorización del uso de las muestras e información general de los pacientes participantes en el proyecto **“Estudio multicéntrico para la evaluación de marcadores genéticos asociados con el desarrollo de la enfermedad de Chagas en Colombia”**. Código Colciencias: 38974.

Estimados señores,

Por medio de la presente comunicación me permito autorizar el uso de las muestras e información general de los pacientes participantes en el proyecto del asunto para ser utilizadas en un proyecto secundario titulado **“Estudio preliminar para el desarrollo de un método de diagnóstico clínico de Chagas a través de la diferenciación metabólica analizada por espectroscopia de resonancia magnética nuclear”**. Este proyecto es un trabajo de maestría en química que será realizado por la estudiante Keyla Maydee Ortiz Ortiz, dirigido por el profesor Juan Carlos Poveda Jaramillo y en el cual participo como co-directora.

Cordial saludo,


CLARA ISABEL GONZALEZ
Profesora Titular laureada
Escuela de Microbiología
Líder Grupo Inmunología y Epidemiología Molecular
Facultad de Salud
Universidad Industrial de Santander

Anexo 2. Aval comité de ética de la Universidad Industrial de Santander



4110

Bucaramanga,

Estudiante

KEYLA MAYDEE ORTIZ ORTIZ

Investigadora principal

Doctor

JUAN CARLOS POVEDA JARAMILLO

Director trabajo de investigación

Maestría en Química

Escuela de Química

Faculta de Ciencias

UIS Presente

Asunto: Aval Comité de Ética proyecto, "Estudio preliminar para el desarrollo de un método de diagnóstico clínico de Chagas a través de la diferencia metabólica analizada por espectroscopia de Resonancia Magnética Nuclear".

Cordial Saludo. El Comité de Ética en Investigación Científica de la Universidad Industrial de Santander (CEINCI-UIS) en reunión realizada el día 7 de abril de 2017, según consta en el acta N° 6, evaluó los ajustes realizados al proyecto del asunto y al respecto conceptúa:

En consideración a que el proyecto cumple con todos los requerimientos del CEINCI-UIS, el Comité acuerda por consenso, **APROBAR** el documento en digital y el consentimiento informado en última versión.

Se recomienda emplear las estrategias que considere necesario, para verificar que el consentimiento informado ha sido comprendido por los participantes. De otra parte, adoptar los mecanismos necesarios para garantizar la confidencialidad de la información recabada. Todo ello amparado en lo reglamentado en la Ley Estatutaria 1581 de 2012 del Congreso de la República de Colombia, por la cual se dictan disposiciones para la protección de datos personales y en la Resolución de Rectoría 1227 del 22 de agosto de 2013, sobre el tratamiento de datos personales. Igualmente realizar los trámites necesarios en las instituciones para acceder a la información y a los participantes. Así mismo, socializar los resultados generados en este proyecto en las instancias correspondientes.

Se aconseja incluir en los procedimientos lo referente al uso de las pólizas de riesgo que tiene la Universidad y que garantizan la atención en salud.

De igual manera, aplicar según corresponda a la investigación, la normatividad del Sistema de Gestión Integral de la Universidad, que está disponible en el enlace:

Proyectó César Hastamorir, revisó Francisco Espinel y aprobó Dora I. Parra.

4110

<https://www.uis.edu.co/intranet/calidad/calidad.jsp>, especialmente lo relacionado con el Manual de Gestión Integrado.

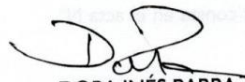
Se solicita que se remita al correo del Comité, información de las siguientes circunstancias, cuando lleguen a ocurrir:

- Reporte de mala práctica científica por parte de cualquier miembro del equipo investigador.
- Notificación previa de las modificaciones realizadas al protocolo.
- Reporte de cualquier eventualidad que usted considera deba conocer el CEINCI-UIS.
- Informe de avance sobre los aspectos éticos según guía e instructivo anexo. Este informe debe enviarse a la mitad del desarrollo de la investigación y al finalizar la misma según el cronograma establecido en el Formato FIN 65.
- El informe debe ser enviado al correo: ceinci.seguimientos@uis.edu.co

En el momento del seguimiento se verificará el cumplimiento de las consideraciones éticas.

En nombre del CEINCI-UIS le ofrecemos el apoyo que usted considere necesario, para la aplicación y salvaguarda de los asuntos éticos durante la investigación.

Atentamente,



DORA INÉS PARRA
Presidenta
CEINCI- UIS



FRANCISCO ESPINEL CORREAL
Secretario Técnico Científico
CEINCI- UIS

Copia: Doctora Clara Isabel González, co-directora del trabajo de investigación, Escuela de Microbiología.
Doctora Verónica García Rojas, Coordinadora de la Maestría en Química.
Archivo Comité de Ética en Investigación Científica.

Proyectó César Hastamorir, revisó Francisco Espinel y aprobó Dora I. Parra.

Anexo 3 Comparación de los resultados obtenidos para el análisis de componentes principales con el software Metaboanalyst y el software R.

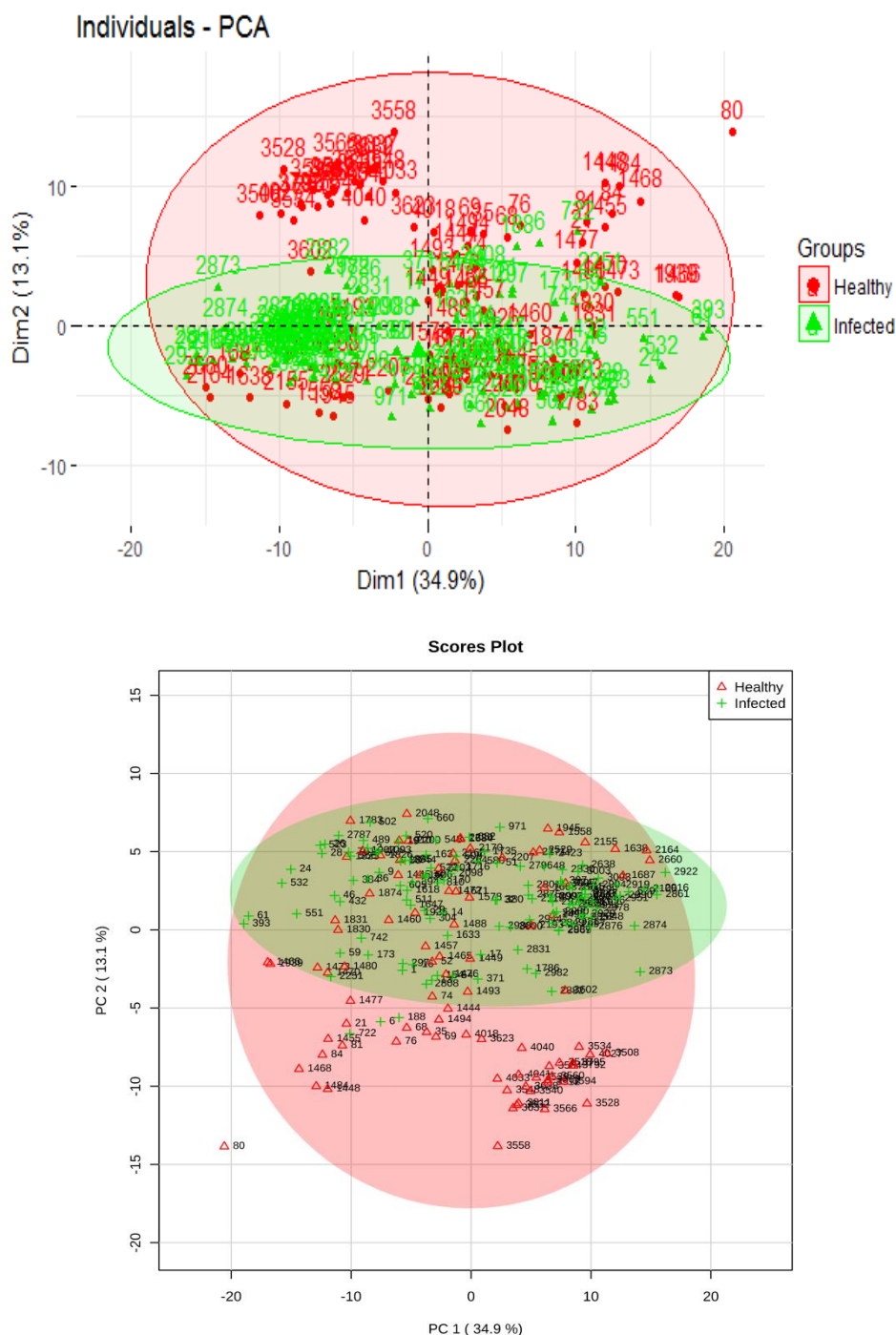


Figura 25. comparación score plots inicial. *Arriba*) R, factorRmine *abajo*) Metaboanalyst. Elaboración propia.

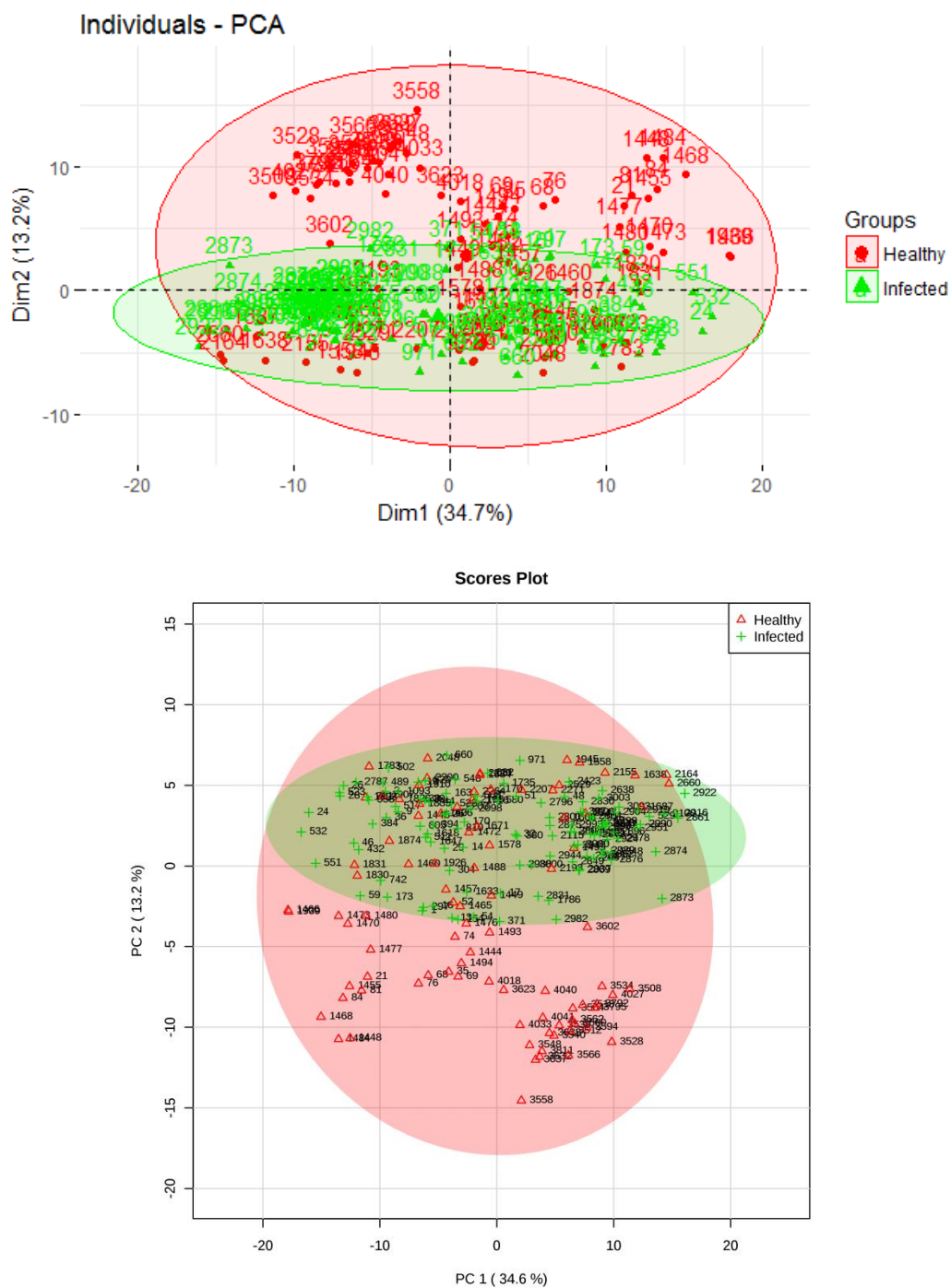


Figura 26.. comparación score plots después de eliminar outliers. arriba) R, factorRmine abajo) Metaboanalyst. Elaboración propia.

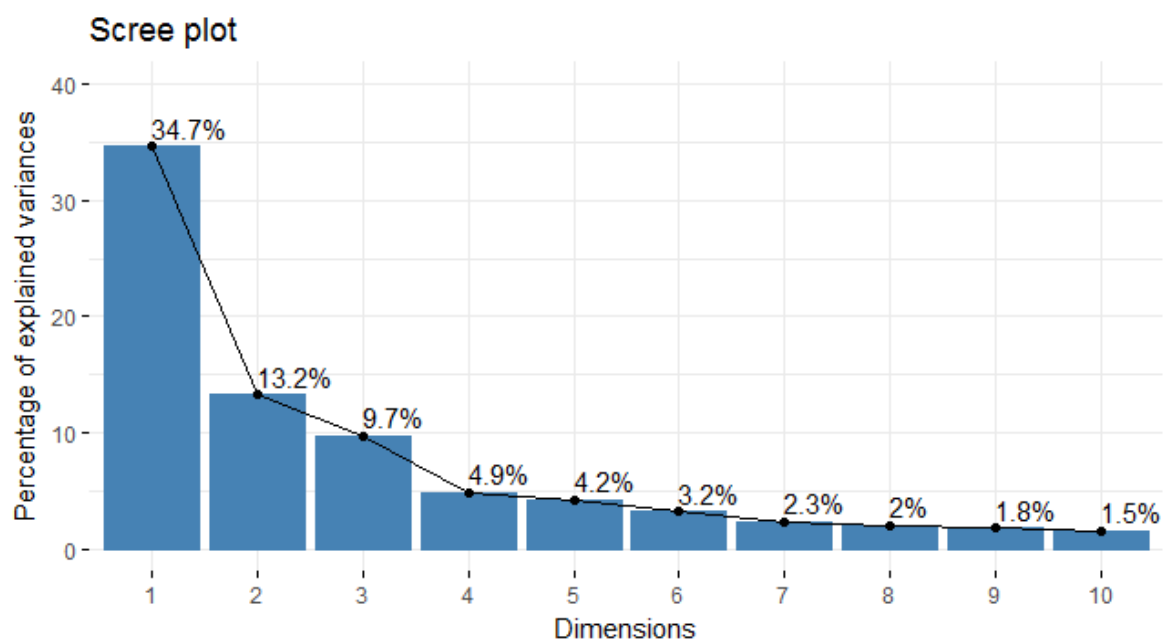
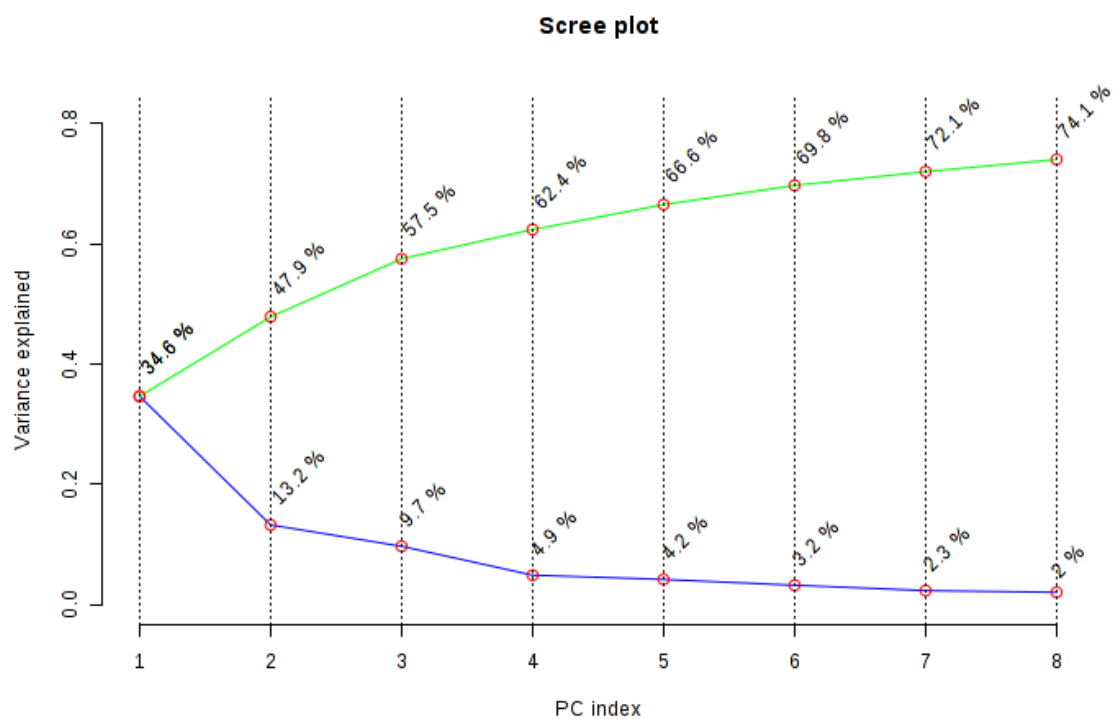


Figura 27. Comparación scree plots *Arriba)* Metaboanalyst *Abajo)* R, factorRmine *abajo*. *Elaboración propia.*

Anexo 4. Test de permutación y curva ROC

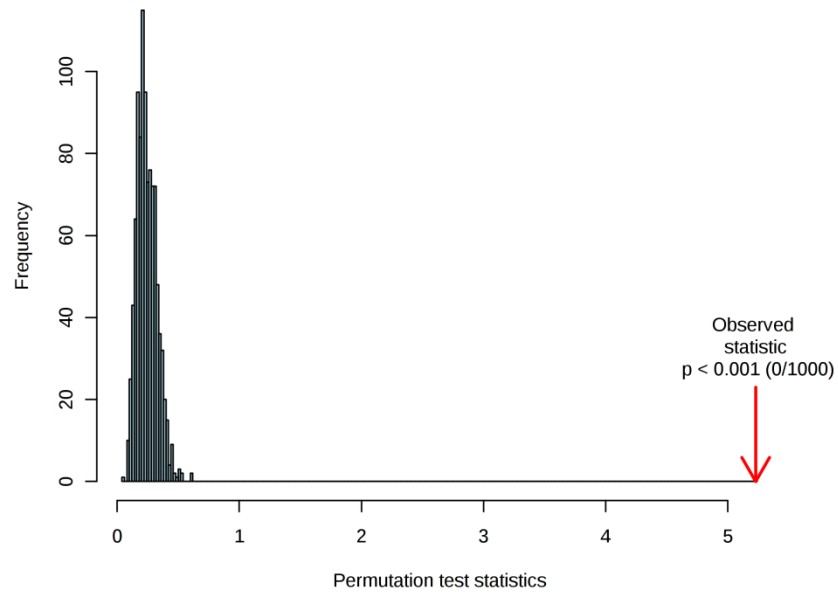


Figura 28. Test de permutación para el modelo MSME-VIP obtenido con la construcción de 1000 modelos a partir de la permutación aleatoria de la etiqueta de clase en la matriz de trabajo. Elaboración propia.

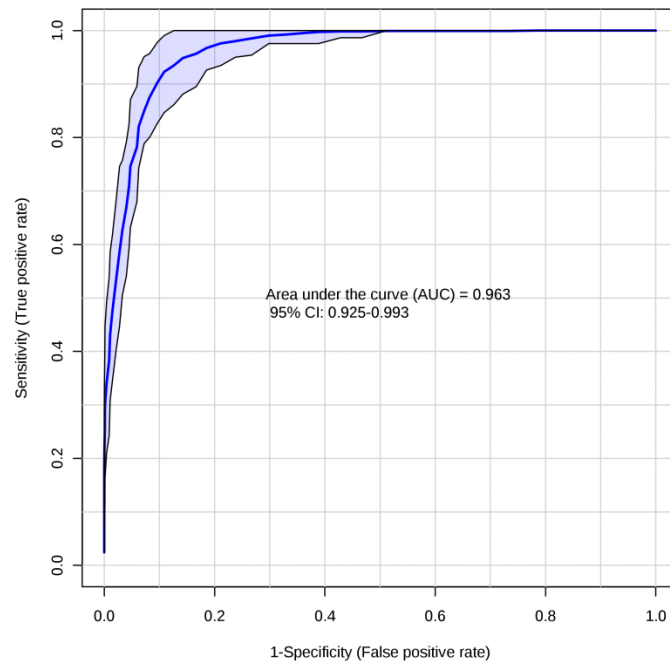


Figura 29. Curva ROC para el modelo M_{SME-VIP}. Elaboración propia