



**Universidad  
Europea MADRID**

**Diferencias entre los resultados de espectrometría de masas de muestras de encéfalo al someterlas a diferentes métodos de preparación.**

**Autor: Lino Arturo Rojas Cruz.**

**Tutor: Verónica Moral Dardé.**

**Facultad de Ciencias Biomédicas y de la Salud.**

**Titulación: Máster Universitario en Bioinformática.**

**Curso 2021-2022**

**INDICE.**

<b>Indice de gráficos.....</b>	<b>3</b>
<b>Indice de tablas.....</b>	<b>6</b>
<b>Glosario de abreviaturas. ....</b>	<b>7</b>
<b>Resumen.....</b>	<b>8</b>
<b>Palabras clave.....</b>	<b>8</b>
<b>Planteamiento de hipótesis. ....</b>	<b>12</b>
<b>Objetivos del trabajo.....</b>	<b>13</b>
<b>Materiales y métodos. ....</b>	<b>14</b>
<b>Resultados. ....</b>	<b>19</b>
<b>Discusión. ....</b>	<b>55</b>
<b>Conclusiones. ....</b>	<b>58</b>
<b>Agradecimiento. ....</b>	<b>59</b>
<b>Anexos. ....</b>	<b>68</b>

## Índice de gráficos.

Gráfico 1. Esquematación de un espectrómetro de masas. Elaboración propia. ....	10
Gráfico 2. Ilustración explicativa sobre las muestras obtenidas de cada individuo de estudio..	14
Gráfico 3. División de las zonas para el análisis de las cromatografías. ....	19
Gráfico 4. Resultados de la cromatografía, realizada con rangos de tiempo entre 2500 y 3500 segundos, del individuo número 1.....	20
Gráfico 5. Resultados de la cromatografía, realizada con rangos de tiempo entre 2500 y 3500 segundos, del individuo número 2.....	21
Gráfico 6. Resultados de la cromatografía, realizada con rangos de tiempo entre 2500 y 3500 segundos, del individuo número 3.....	22
Gráfico 7. Detección de picos cromatográficos usando el algoritmo centWave del paquete de R “xcms: LC-MS and GC-MS Data Analysis” utilizando un tiempo de retención entre 2500 y 3500 segundos y utilizando una brecha de M/Z entre 434.9 y 435.1. Individuo número 1. ....	23
Gráfico 8. Detección de picos cromatográficos usando el algoritmo centWave del paquete de R “xcms: LC-MS and GC-MS Data Analysis” utilizando un tiempo de retención entre 2500 y 3500 segundos y utilizando una brecha de M/Z entre 434.9 y 435.1. Individuo número 1. ....	24
Gráfico 9. Detección de picos cromatográficos usando el algoritmo centWave del paquete de R “xcms: LC-MS and GC-MS Data Analysis” utilizando un tiempo de retención entre 2500 y 3500 segundos y utilizando una brecha de M/Z entre 434.9 y 435.1. Individuo número 3. ....	25
Gráfico 10. Visualización cromatográfica de la relación que existe entre intensidad, M/Z y tiempo de retención, utilizando un tiempo de retención entre 2500 y 3500 segundos y utilizando una brecha de M/Z entre 434.9 y 435.1. Individuo número 1. ....	27
Gráfico 11. Visualización cromatográfica de la relación que existe entre intensidad, M/Z y tiempo de retención, utilizando un tiempo de retención entre 2500 y 3500 segundos y utilizando una brecha de M/Z entre 434.9 y 435.1. Individuo número 2. ....	28
Gráfico 12. Visualización cromatográfica de la relación que existe entre intensidad, M/Z y tiempo de retención, utilizando un tiempo de retención entre 2500 y 3500 segundos y utilizando una brecha de M/Z entre 434.9 y 435.1. Individuo número 3. ....	29
Gráfico 13. Cromatograma realizado utilizando, exclusivamente, iones extraídos de las muestras, lo cual nos ayudará a diferenciar las distintas configuraciones de detección máxima. utilizando un tiempo de retención entre 2500 y 3500 segundos, usando el algoritmo de centWave del paquete de R “xcms: LC-MS and GC-MS Data Analysis”. Individuo 1.....	30
Gráfico 14. Cromatograma realizado utilizando, exclusivamente, iones extraídos de las muestras, lo cual nos ayudará a diferenciar las distintas configuraciones de detección máxima. utilizando un tiempo de retención entre 2500 y 3500 segundos, usando el algoritmo de centWave del paquete de R “xcms: LC-MS and GC-MS Data Analysis”. Individuo 2.....	31

Gráfico 15. Cromatograma realizado utilizando, exclusivamente, iones extraídos de las muestras, lo cual nos ayudará a diferenciar las distintas configuraciones de detección máxima, utilizando un tiempo de retención entre 2500 y 3500 segundos, usando el algoritmo de centWave del paquete de R “xcms: LC-MS and GC-MS Data Analysis”. Individuo 3.....	32
Gráfico 16. Visualización de picos cromatográficos en el conjunto de datos completo, considerando las señales con un valor de intensidad de los iones entre 6 a 5000 cps en el paso de detección de picos, cubriendo todos los posibles picos, utilizando un tiempo de retención entre 2500 y 3500 segundos. Individuo 1. ....	33
Gráfico 17. Visualización de picos cromatográficos en el conjunto de datos completo, considerando las señales con un valor de intensidad de los iones entre 6 a 5000 cps en el paso de detección de picos, cubriendo todos los posibles picos, utilizando un tiempo de retención entre 2500 y 3500 segundos. Individuo 2. ....	34
Gráfico 18. Visualización de picos cromatográficos en el conjunto de datos completo, considerando las señales con un valor de intensidad de los iones entre 6 a 5000 cps en el paso de detección de picos, cubriendo todos los posibles picos, utilizando un tiempo de retención entre 2500 y 3500 segundos. Individuo 3. ....	35
Gráfico 19. El siguiente gráfico presenta el conteo total de picos por muestra e individuo, con un conteo del número de picos que presento un tiempo de retención relativo (0 %,25%, 50%, 75% y 100% de tiempo de retención), visualizando así, el número de picos identificados por muestra y ancho/duración de los picos cromatográficos.....	36
Gráfico 20. Visualización de la frecuencia de los picos cromatográficos identificados a lo largo del eje del tiempo de retención, identificando así, la frecuencia de los picos detectados a lo largo del eje de tiempo de retención. Lo que nos permite identificar períodos de tiempo a lo largo de la ejecución de la espectrometría de masa con una mayor cantidad de picos cromatográficos, visualizando de manera más comprensible las diferencias entre cada una de las muestras en cuento este aspecto. Individuo 1. ....	37
Gráfico 21. Visualización de la frecuencia de los picos cromatográficos identificados a lo largo del eje del tiempo de retención, identificando así, la frecuencia de los picos detectados a lo largo del eje de tiempo de retención. Lo que nos permite identificar períodos de tiempo a lo largo de la ejecución de la espectrometría de masa con una mayor cantidad de picos cromatográficos, visualizando de manera más comprensible las diferencias entre cada una de las muestras en cuento este aspecto. Individuo 2. ....	38
Gráfico 22. Visualización de la frecuencia de los picos cromatográficos identificados a lo largo del eje del tiempo de retención, identificando así, la frecuencia de los picos detectados a lo largo del eje de tiempo de retención. Lo que nos permite identificar períodos de tiempo a lo largo de la ejecución de la espectrometría de masa con una mayor cantidad de picos cromatográficos,	

visualizando de manera más comprensible las diferencias entre cada una de las muestras en cuento este aspecto. Individuo 3. ....	39
Gráfico 23. Visualización cromatográfica individualizando los picos para un valor específico de M/Z entre 465 y 499, con un tiempo de retención entre 2500 y 3500 segundos. Individuo 1. ....	40
Gráfico 24. Visualización cromatográfica individualizando los picos para un valor específico de M/Z entre 465 y 499, con un tiempo de retención entre 2500 y 3500 segundos. Individuo 2. ....	41
Gráfico 25. Visualización cromatográfica individualizando los picos para un valor específico de M/Z entre 465 y 499, con un tiempo de retención entre 2500 y 3500 segundos. Individuo 3. ....	42
Gráfico 26. Huella de Datos del análisis estadístico de M/Z, tiempo de retención, tiempo de retención máximo y mínimo e intensidad por cada muestra, con un M/Z entre 750 y 1500. Muestra la relación espacial de los diferentes datos. Individuo 1. ....	45
Gráfico 27. Huella de Datos del análisis estadístico de M/Z, tiempo de retención, tiempo de retención máximo y mínimo e intensidad por cada muestra, con un M/Z entre 750 y 1500. Muestra la relación espacial de los diferentes datos. Individuo 2. ....	46
Gráfico 28. Huella de Datos del análisis estadístico de M/Z, tiempo de retención, tiempo de retención máximo y mínimo e intensidad por cada muestra, con un M/Z entre 750 y 1500. Muestra la relación espacial de los diferentes datos. Individuo 3. ....	47
Gráfico 29. Representación gráfica de la relación entre tiempo de retención con intensidad, con su media, desviación estándar y la densidad del conteo en los dos ejes: X y Y. Individuo 1. ....	49
Gráfico 30. Representación gráfica de la relación entre tiempo de retención con M/Z, con su media, desviación estándar y la densidad del conteo en los dos ejes: X y Y. Individuo 1. ....	50
Gráfico 31. Representación gráfica de la relación entre tiempo de retención con intensidad, con su media, desviación estándar y la densidad del conteo en los dos ejes: X y Y. Individuo 2. ....	51
Gráfico 32. Representación gráfica de la relación entre tiempo de retención con M/Z, con su media, desviación estándar y la densidad del conteo en los dos ejes: X y Y. Individuo 2. ....	52
Gráfico 33. Representación gráfica de la relación entre tiempo de retención con intensidad, con su media, desviación estándar y la densidad del conteo en los dos ejes: X y Y. Individuo 3. ....	53
Gráfico 34. Representación gráfica de la relación entre tiempo de retención con M/Z, con su media, desviación estándar y la densidad del conteo en los dos ejes: X y Y. Individuo 3. ....	54

**Índice de tablas.**

Tabla 1. Distribución de procesos de preparación de las muestras. ....	16
Tabla 2. Tabla resumen del conteo total de picos por muestra e individuo, con un conteo del número de picos que presento un tiempo de retención relativo (0 %,25%, 50%, 75% y 100% de tiempo de retención), visualizando así, el número de picos identificados por muestra y ancho/duración de los picos cromatográficos . ....	68
Tabla 3. Resumen de las medidas de dispersión, desviación estándar, valores máximo y mínimo de los valores más importantes de cada muestra de cada réplica biológica. ....	71

**Glosario de abreviaturas.**

MS: espectrometría de masas.

M/Z: masa/carga.

ug: microgramos.

Tu/U: tampón de lisis de tiourea/urea.

M: moles/litro.

CHAPS: 3-[(3 - Dholamidopropil ) Dimetilamonio ] -1 -Propano Sulfonato.

DTT: Ditiol.

mg: miligramos.

uL: microlitros.

RIPA: tampón para análisis de radioinmunoprecipitación.

SP3: digestión de la muestra mediante la preparación de muestras mejorada en fase sólida de un solo recipiente.

kV: kilovoltios.

cps: conteos por segundo.

3D: tres dimensiones.

Rt: tiempo de retención.

Rtmin: tiempo de retención mínimo.

Rtmax: tiempo de retención máximo.

Into/10000: intensidad en conteos por segundo, dividido para 10000.

## **Resumen.**

El estudio de la neuroproteómica es uno de los campos que más ha avanzado en los últimos años, el estudio de las particularidades del cerebro, su composición y estructura, son de especial relevancia en la actualidad por su complejidad y su relación con el resto de los sistemas de cuerpo. Así, la espectrometría de masas es una herramienta potente que brinda gran información cuantitativa sobre el mismo. La hipótesis planteada ha sido que, los resultados de la espectrometría de masas de muestras de cerebro, que tengan protocolos de preparación diferentes, presentarán resultados diversos sin alterar la estructura básica de los parámetros de interés. El objetivo principal del estudio es analizar los datos de espectrometría de masas de las muestras de cada individuo usando un laboratorio virtual, creado con herramientas de fuente libre. Se ha usado como materiales, datos de muestras de 3 réplicas biológicas de cerebro, con 4 protocolos diferentes de preparación de muestra, analizadas mediante espectrometría de masas, realizando un total de 12 experimentos, obteniendo un total de 144 análisis diferentes como resultado. La metodología del análisis de cada experimento se creó en lenguaje de programación R, usando varias librerías de fuente abierta para analizar datos masivos, espectrometría de masas y datos bioestadísticos. Dentro de los resultados cabe señalar que, se ha encontrado patrones comunes entre muestras en los diferentes experimentos, pero a pesar de ello, se han presentado resultados aberrantes en varios resultados, rompiendo los patrones comunes, sin embargo, no se ha logrado encontrar un factor común que sea el causante de esta perturbación en los resultados. Se ha llegado a la conclusión de que se tiene que estudiar el tema con mayor profundidad, existe poca información con este enfoque en investigaciones sobre tejido de cerebro, sobre todo porque se conoce que las células cerebrales no se comportan de la misma en comparación con las demás.

## **Palabras clave.**

Español: espectrometría de masas; cromatografía de gases y espectrometría de masas; cerebro; lenguajes de programación; “R”; software de fuente abierta.

Inglés: Mass spectrometry; gas chromatography-mass spectrometry; brain, programming languages; “R”, open source software.



## **Introducción.**

La neuroproteómica se define como el campo científico que estudia la estructura, conformación, expresión, función e interacciones de las proteínas del sistema nervioso central de individuos que presentan, o no, una patología, así como las variaciones fisiológicas entre individuos y su repercusión en la relación que estos tienen con el medio externo e interno. (5,7–9) Las células que conforman el sistema nervioso son únicas, no se ha logrado entender completamente como realizan las funciones correspondientes o conocer el total de sus funciones; la relación entre sí y con otros sistemas, es un campo de estudio que está en constante avance, muchas veces incluso con resultados contraintuitivos. (1,10,11)

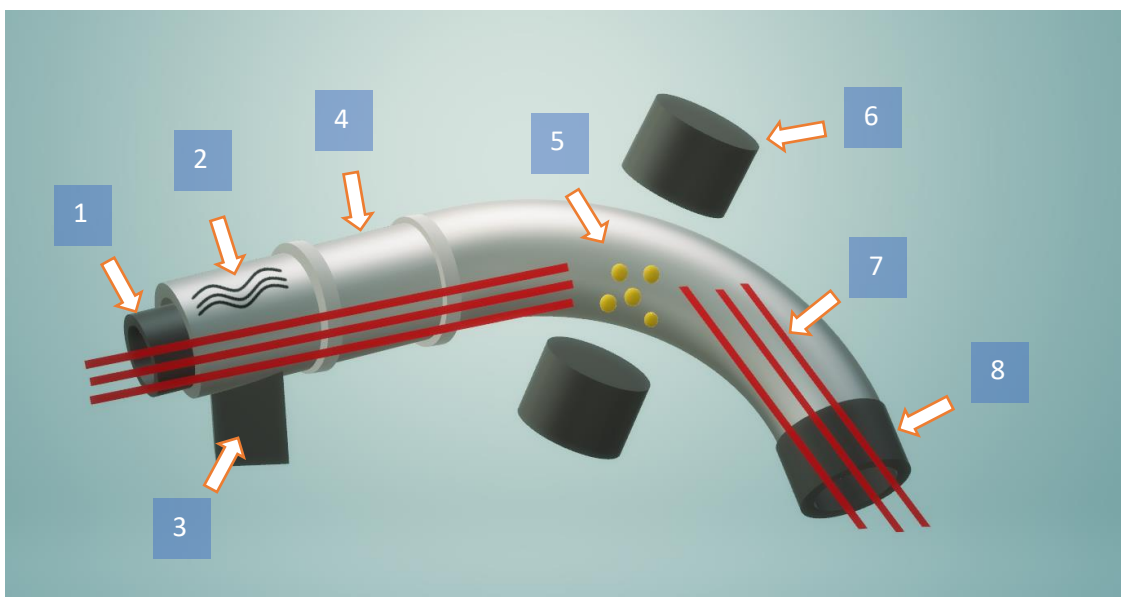
El estudio del sistema nervioso a través de la espectrometría de masas (MS, del inglés mass spectrometry) es uno de los campos que avanza con mayor rapidez actualmente. La capacidad de analizar, a un nivel más avanzado y completo, el complejo entramado de la conformación, interacción y consecuencias de las variaciones encontradas en las células nerviosas ha impulsado un avance en el conocimiento del órgano irremplazable. Para que se lleve a cabo una correcta investigación del sistema nervioso, a través de la MS, es importante entender que cada parte del proceso tiene un rol fundamental; desde la selección adecuada del individuo de estudio, la correcta toma de muestra, la preparación de esta, hasta el correcto procesamiento de los resultados, la interpretación de la información y la adecuada divulgación de los nuevos conocimientos. Es así, que el presente trabajo se basará en tres zonas fundamentales de estos campos de la investigación: la preparación de las muestras, las variaciones de los resultados de la MS y la transmisión de conocimientos.

La MS es una técnica analítica avanzada que identifica y detecta moléculas específicas en muestras complejas mediante la clasificación de iones en campos eléctricos y magnéticos de acuerdo con la relación de masa-carga ( $M/Z$ ) que presentan, permitiendo de esta manera, un análisis completo de los elementos que conforman la muestra. (1–3) Actualmente, la MS es uno de los mejores métodos para analizar, de forma más completa, el perfil cuantitativo de proteínas, las interacciones entre elementos y las modificaciones que presentan las proteínas.(4)

A pesar de que el primer espectrómetro de masas se creó en 1912, los avances que esta tecnología ha presentado en estas últimas décadas nos han permitido analizar los datos de forma masiva gracias al resurgimiento de la investigación bioinformática, la ciencia de datos, las nuevas y avanzadas capacidades de análisis masivo de información y el avance de los estudios ómicos.(4–6) Esto en realidad abre nuevas oportunidades de investigación.

En el siguiente gráfico se puede observar un esquema del funcionamiento principal de un espectrómetro de masas. Así, con este esquema, entender mejor el complejo proceso de la MS. En el siguiente gráfico podemos ver enumerado:

- 1: entrada de la muestra.
- 2: fuente de calor para vaporizar la muestra.
- 3: fuente de corriente de electrones.
- 4: aceleración de iones.
- 5: realineación de elementos por campos magnéticos.
- 6: imán.
- 7: salida de iones.
- 8: detector.



*Gráfico 1. Esquemización de un espectrómetro de masas. Elaboración propia.*

Los datos obtenidos a través de la MS son un conjunto de datos masivos con varios parámetros. En una muestra como la que se ha usado, se pueden obtener cientos de miles de parámetros analizados, lo que hace que el análisis de estos sea todo un reto desde el punto de vista técnico. De tal forma que encontrar los patrones en los resultados se vuelve crítico para dar sentido a la ingente cantidad de datos.

La preparación de una muestra es uno de los procesos más importantes dentro del estudio del proteoma mediante la MS. La MS, al ser una técnica cuantitativa y completa, como lo hemos explicado anteriormente, puede presentar variaciones dependiendo de la forma de preparar una muestra. “La preparación de una muestra proteómica es realmente un desafío y puede considerarse tanto un arte como una ciencia”(4), por lo que en el siguiente trabajo se presentarán los resultados obtenidos mediante MS, de muestras de encéfalo de tres individuos diferentes, con 4 preparaciones diferentes cada una, obteniendo en total 12 resultados de MS.

El sistema nervioso, sobre todo el cerebro, es el sistema más complejo, importante y el menos conocido hoy en día. El órgano que nos ha permitido construir y derribar imperios, crear cohetes y llevarnos a la luna, crear otros cohetes y con ellos destruir ciudades enteras; el órgano que nos hace capaces de tener conciencia de nosotros mismos y de entender la existencia de los demás, que nos hace capaces de la compasión y del odio, del amor y la ira; el órgano que nos permite percibir el medio externo y concebir la realidad, solo tanto y cuanto, según la forma de activarse y transmitir información a partir de pequeñas células dentro de aquella bóveda ósea. Tan importante es, que al no poder mostrar su actividad se da por declarada la muerte del individuo a pesar de que su corazón siga latiendo, la muerte encefálica. Al cosmos y al cerebro se los estudiarán por siempre, son inabarcables.

### **Planteamiento de hipótesis.**

Como lo hemos explicado anteriormente, conocemos que la MS es una técnica cuantitativa, produce un análisis de la muestra de una forma muy completa(3), esto es una ventaja ya que nos permite estudiar muestras complejas. Al mostrar los resultados de una forma tan precisa, es muy probable que existan variaciones en los resultados al preparar la muestra siguiendo diferentes protocolos.(5,7) En principio esto no debería ser un problema ya que el análisis de los datos obtenidos y contrastados con las bases de datos debería mostrarnos un resultado coherente con la muestra estudiada.(7) Sabemos también que el sistema nervioso central presenta características únicas a diferencia de otros tejidos(7,9,10), lo que hace que también su interacción con los diferentes elementos de preparación de la muestra para diferentes protocolos de preparación, probablemente, muestre interacciones diversas. La pregunta que cabe aquí es, ¿qué tanto pueden cambiar los resultados de una MS al tener diferentes preparaciones en muestras biológicas similares de cerebro?

Los resultados de MS de muestras de cerebro de un mismo individuo con protocolos de preparación diferentes darán resultados diferentes que, a pesar de sus diferencias, logren conservar sus estructuras básicas en cuanto a los parámetros de interés, a saber: picos de intensidad, tiempos de retención y su aparición en una ventana temporal específica, sin distorsionar sus estructuras básicas.

### **Objetivos del trabajo.**

Investigar las particularidades del órgano más complejo que el ser humano posee no es, en absoluto, poca cosa. Entre más luces tenemos sobre el cerebro más dudas se plantean. Sin dejar de lado la importancia de estudiar el resto de los órganos y tejidos, es preciso que todas las nuevas tecnologías brinden su granito de claridad en cuanto al órgano irremplazable. Es por esto, por lo que el objetivo principal de este trabajo es:

- Analizar los resultados de la MS de las diferentes muestras biológicas de cerebro de cada individuo con cada protocolo de preparación, y los cambios que producen en los mismos, utilizando un laboratorio virtual con varios tipos de experimentos, creado con herramientas de fuente libre.

Al ser un trabajo con tantos puntos de interés y sumergirse en un tema esencial para el presente y el futuro, como es el análisis de los datos masivos, en inglés conocido como “big data”, se presentará además como objetivos secundarios del presente trabajo:

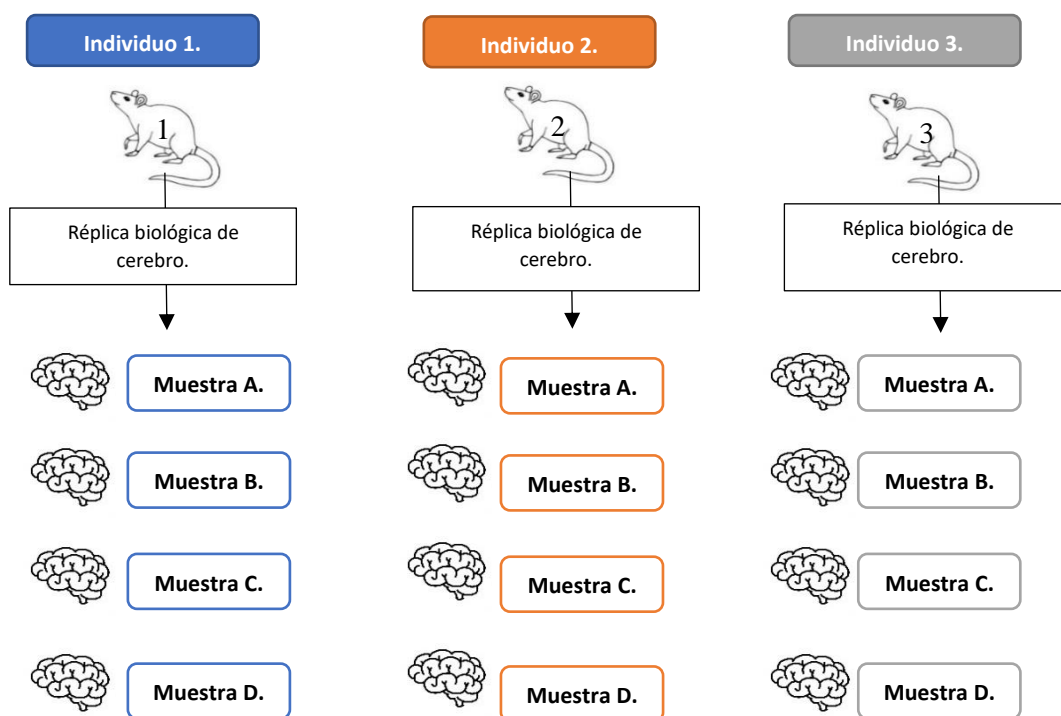
- Explorar el uso de varias herramientas de visualización “clásica” de datos masivos para representar los resultados de la MS.
- Elaborar un laboratorio virtual con herramientas de fuente abierta para generar varios.
- Plantear el uso de nuevas herramientas de visualización de datos masivos para representar los resultados de la MS.
- Experimentos que permitan analizar los datos de la MS.
- Encontrar patrones en los resultados de la MS.

## Materiales y métodos.

Uno de los avances más importantes de la civilización occidental es la democratización de la investigación, es decir, la apertura de barreras que permiten que, el estatus socioeconómico, el nivel de educación o el país de origen, dejen de ser fronteras infranqueables para las mentes inquietas. Permitiendo, de esta forma, un desarrollo más rápido y justo de la raza humana. Se ha utilizado herramientas de fuente abierta, conocidas en inglés como “open source tools”, para realizar esta investigación. (12)

Los materiales que se ha utilizado en el presente trabajo los podemos dividir en 2 grupos. El primero, los materiales preprocesamiento, estos son: los datos obtenidos. Estos, han sido obtenidos a través del apoyo del grupo de investigación de neuroproteómica de la “Unidad de Proteómica del Hospital Nacional de Paraplégicos de Toledo (España)”.

Para entender mejor el trabajo de investigación, es importante entender el origen y características de las muestras que han producido los resultados que se han analizado. En primer lugar cabe destacar que, se han usado 3 individuos diferentes, estos han sido ratones para investigación científica, sin alteraciones ni modificaciones genéticas. De cada individuo se han tomado 4 muestras de sistema nervioso central, específicamente de cerebro. Para cada individuo, cada una de las 4 muestras, se la ha procesado con un protocolo de preparación diferente. Obteniendo, en resumen, 12 muestras, de tres individuos diferentes, preparadas con 4 protocolos diferentes, como lo indica el gráfico 1.



*Gráfico 2. Ilustración explicativa sobre las muestras obtenidas de cada individuo de estudio.*

Cada una de las muestras, de cada uno de los individuos, se representarán de ahora en adelante con la nomenclatura correspondiente, primero, al individuo de estudio, y segundo, al tipo de preparación específica de la muestra. De la siguiente manera, al referirnos a la muestra del individuo 1 cuya preparación corresponde a la preparación de todas las muestras A, se lo llamará: “individuo 1, muestra A”; así, con cada uno de los individuos y tipo de procesamiento de la muestra. La explicación de como se ha procesado cada tipo de muestra se realizará en los párrafos siguientes, de acuerdo con lo indicado por el grupo de investigación en neuroproteómica que obtuvo las muestras y realizó la MS.

Es fundamental conocer que cada individuo de estudio aportó con una réplica biológica, por lo que se puede indicar que para realizar la investigación se cuenta con al menos 3 réplicas biológicas, lo cual entrega robustez al estudio (13–17). Cada una de las réplicas biológicas es una muestra de cerebro de 520 microgramos (ug) de proteína total.

Para esquematizar de una forma clara el proceso de preparación de cada una de las muestras, se va a realizar en primer lugar la explicación de los procesos a los que estuvieron expuestas cada una de las 4 partes de cada réplica biológica. Posteriormente se indicará la pertenencia de cada proceso con su correspondiente muestra. Así, tenemos que se han realizado 4 procedimientos diferentes a acuerdo a los protocolos establecidos y normados en cada uno, por lo que mencionaremos 2 momentos en la preparación de cada muestra, el primero, el tipo de tampón de extracción, es decir, la forma de realizar el lisado de la muestra. El segundo, el tipo de digestión de la muestra.

El lisado, se ha realizado mediante 2 procesos diferentes, el primero, utilizando el tampón de lisis de tiourea/urea (Tu/U), utilizado como gel de electroforesis de 2 dimensiones(18,19), que utiliza para cada tampón 7 moles/litro (M) de Urea, 2 M de Tiourea, 4% de 3-[(3 - Dholamidopropil ) Dimetilamonio ] -1 -Propano Sulfonato (CHAPS), 2% de Pharmalyte®, 40 milimoles/litro (mM) de Ditiol (DTT) (20), por lo que el cálculo del compuesto para 1 ml de tampón sería: 152.2 miligramos (mg) de Tiourea, 420,4 mg de Urea, 40 mg de CHAPS, 10 mg de DTT, y 20 microlitros (uL). (18) El segundo, utilizando el tampón para análisis de radioinmunoprecipitación (RIPA), el cual se utiliza para la lisis celular rápida al mismo tiempo que evita la degradación de las proteínas en las muestras biológicas. (21)

En cuanto al proceso de digestión, también podemos observar el procesamiento de la muestra mediante dos protocolos diferentes. El primero, la digestión de la muestra “en gel” conocido en inglés como “in-gel-digestion”, la que se realizó de acuerdo con los protocolos de preparación de la muestra. (22–24) El segundo, la digestión de la muestra mediante la preparación de muestras mejorada en fase sólida de un solo recipiente (SP3), la cual se realizó según los protocolos de preparación. (25–27)

De esta forma vamos a tener una distribución de procesamiento de muestras que puede interpretarse de la siguiente manera, cada una de las réplicas biológicas se ha dividido en 4 muestras, las muestra A: tiene un lisado en RIPA con una digestión “en gel”, las muestras B: tienen un lisado en RIPA con una digestión SP3, la muestra C: tiene un lisado en Tu/U con una digestión “en gel” y la muestra D: tiene un lisado en Tu/U con una digestión SP3, como lo muestra la tabla 1.

Las muestras se analizaron en un espectrómetro de masas de tipo nanoESI qQTOF (TripleTOF® 6600plus, Sciex). Las muestras fueron ionizadas a través de una fuente de tipo Optiflow 1-50 uL aplicando 4.5 kilovóltios (kV) al emisor de spray a 100°C. El análisis de las muestras en el espectrómetro de masas se llevarón a cabo en modo dependiente de los datos.

Tipo de muestra	Réplica Biológica	Tipo de tapón de Extracción	Tipo de Digestión	Nombre de la muestra.
Cerebro	Individuo 1	RIPA	“en gel”	Individuo 1, muestra A.
			SP3	Individuo 1, muestra B.
		Tu/U	“en gel”	Individuo 1, muestra C.
			SP3	Individuo 1, muestra D.
	Individuo 2	RIPA	“en gel”	Individuo 2, muestra A.
			SP3	Individuo 2, muestra B.
		Tu/U	“en gel”	Individuo 2, muestra C.
			SP3	Individuo 2, muestra D.
	Individuo 3	RIPA	“en gel”	Individuo 3, muestra A.
			SP3	Individuo 3, muestra B.
		Tu/U	“en gel”	Individuo 3, muestra C.
			SP3	Individuo 3, muestra D.

Tabla 1. Distribución de procesos de preparación de las muestras.

*Nota de la tabla 1: Lisado mediante tampón para análisis de radioinmunoprecipitación (RIPA), Lisado mediante tampón de lisis de tiourea/urea (Tu/U), digestión de la muestra con el protocolo “en gel” (“en gel”), digestión de la muestra mediante la preparación de muestras mejorada en fase sólida de un solo recipiente (SP3).*

El segundo grupo de materiales del presente trabajo son, los materiales del procesamiento, estos son: el lenguaje de programación: R, el entorno de desarrollo integrado para el lenguaje de programación usado: RStudio, lo cual permite crear “a medida” un entorno de desarrollo y



experimentación ideal, utilizando software de código abierto. (28–31) A este grupo de materiales debemos añadir el software de fuente abierta Jamovi. (32,33)

Siguiendo el concepto bioinformático del término, un laboratorio virtual es una herramienta que nos ayuda a procesar con un orden específico un tipo concreto de muestras para generar uno o varios resultados de los experimentos realizados a través del código creado,(34–36) existen muchas formas de crear un laboratorio virtual, en el presente trabajo se ha creado un algoritmo de procesamiento de la información de las muestras. Así, analizar a través de varios experimentos los datos ingresados. En cuanto al ingreso de la muestra, y como se podrá apreciar en los documentos de resultados del procesamiento del código para cada muestra, se ha realizado de tal forma que sea necesario solamente cambiar el archivo de ingreso y correr el código.

En cuanto a la estructura del laboratorio virtual, se ha tratado de replicar el “espíritu” de los laboratorios físicos, donde una muestra diferente a través de un proceso concreto proporciona resultados específicos. La metodología que se ha seguido y el código creado para elaborar el laboratorio virtual, utilizando el lenguaje de programación R y el entorno de desarrollo RStudio, se puede revisar en el siguiente enlace [https://1drv.ms/f/s!An\\_DYEB6yVClxASVOF29VIYEuvix](https://1drv.ms/f/s!An_DYEB6yVClxASVOF29VIYEuvix) , en el mismo se podrá apreciar el código usado para cada resultado de análisis de la MS para cada muestra biológica.

Es importante recalcar que se ha utilizado varios paquetes de R de Bioconductor para la creación del laboratorio virtual. Bioconductor se concibe como un proyecto de software de código y desarrollo abiertos que utiliza R como su lenguaje de programación, que tiene por finalidad analizar y comprender los datos de alto rendimiento en las ciencias de la genómica y la biología molecular.(37–39) Los paquetes usados son: “RforMassSpectrometry”(40,41), “msdata”(42), “mzR”(43–46), “Rcpp”(47–49), “lattice”(50), “gridExtra”(51), “MSstatsTMT”(52), “MSnBase”(40,41), “Spectra”(53), “rpx”(54), “xcms”(55–57), “Rcolorbrewer”(58), “pander”(59), “madritr”(60), “pheatmap”(61), “jamovi”(62) y “SummarizedExperiment”(63).

La metodología de investigación utilizada, en el presente trabajo, se ha basado principalmente en el análisis de los datos masivos obtenidos como resultado de la MS. Así, se ha llevado a cabo un total de 12 experimentos en cada una de las 12 muestras pertenecientes a las 3 diferentes réplicas biológicas, dando como resultado un total de 144 análisis, los cuales han sido organizados en el apartado de resultados, más adelante, por experimento y por individuo. De esta manera, poder tener una mejor comprensión de la información obtenida como resultado. A continuación se enunciarán cada uno de los experimentos realizados.

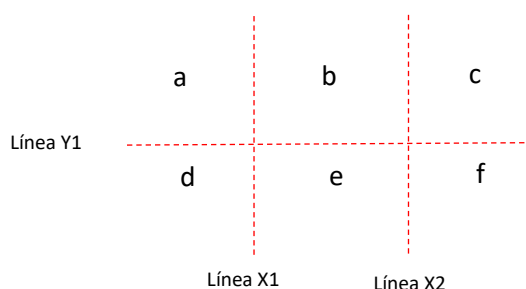
- 1- Resultados de la cromatografía, realizada con rangos de tiempo entre 2500 y 3500 segundos.

- 2- Detección de picos cromatográficos usando el algoritmo centWave del paquete de R “xcms: LC-MS and GC-MS Data Analysis” utilizando un tiempo de retención entre 2500 y 3500 segundos y utilizando una brecha de M/Z entre 434.9 y 435.1.
- 3- Visualización cromatográfica de la relación que existe entre intensidad, M/Z y tiempo de retención, utilizando un tiempo de retención entre 2500 y 3500 segundos y utilizando una brecha de M/Z entre 434.9 y 435.1.
- 4- Cromatograma realizado utilizando, exclusivamente, iones extraídos de las muestras, lo cual nos ayudará a diferenciar las distintas configuraciones de detección máxima, utilizando un tiempo de retención entre 2500 y 3500 segundos.
- 5- Visualización de picos cromatográficos en el conjunto de datos completo, considerando las señales con un valor de intensidad de los iones entre 6 a 5000 en el paso de detección de picos, cubriendo todos los posibles picos, utilizando un tiempo de retención entre 2500 y 3500 segundos.
- 6- Conteo total de picos por muestra e individuo, con un conteo del número de picos que presento un tiempo de retención relativo (0 %,25%, 50%, 75% y 100% de tiempo de retención), visualizando así, el número de picos identificados por muestra y ancho/duración de los picos cromatográficos.
- 7- Visualización de la frecuencia de los picos cromatográficos identificados a lo largo del eje del tiempo de retención, identificando así, la frecuencia de los picos detectados a lo largo del eje de tiempo de retención. Lo que nos permite identificar períodos de tiempo a lo largo de la ejecución de la espectrometría de masa con una mayor cantidad de picos cromatográficos.
- 8- Visualización cromatográfica individualizando los picos para un valor específico de M/Z entre 465 y 499, con un tiempo de retención entre 2500 y 3500 segundos.
- 9- Análisis estadístico de los datos de M/Z, tiempo de retención, tiempo de retención máximo y mínimo e intensidad por cada muestra, con un M/Z entre 750 y 1500.
- 10- Huella de datos del análisis estadístico de los datos de M/Z, tiempo de retención, tiempo de retención máximo y mínimo e intensidad por cada muestra, con un M/Z entre 750 y 1500. Muestra la relación espacial de los diferentes datos.
- 11- Representación gráfica de la relación entre tiempo de retención e intensidad, con su media, desviación estándar y la densidad del conteo en los dos ejes: X y Y.
- 12- Representación gráfica de la relación entre tiempo de retención y M/Z, con su media, desviación estándar y la densidad del conteo en los dos ejes: X y Y.

## Resultados.

Como se explicó en el apartado previo, de materiales y métodos, se ha realizado un total de 12 experimentos virtuales para cada una de las 4 muestras biológicas de cada una de las 3 réplicas biológicas de cerebro. Resultando así, un total de 144 análisis de la información de obtenida mediante espectrometría de masas de cada muestra. Se ha agrupado los resultados de una forma lógica, por experimento, por muestras y por individuo, así poder apreciar de una manera más clara los resultados.

Para el análisis de cada experimento, vamos a dividirlos en 6 zonas, cada zona tendrá un nombre específico asignándole una letra del alfabeto entre la a y f, en orden descendente y de derecha a izquierda, como lo indica el siguiente gráfico.



*Gráfico 3. División de las zonas para el análisis de las cromatografías.*

Una de las fortalezas del análisis cromatográfico de una MS es la posibilidad de poder estudiar la conformación de elementos a través de patrones concretos, esto se realiza mediante análisis de cotejamiento con bibliotecas de fuente libre o con softwares dedicados a este fin. En el siguiente trabajo, no se realizará una búsqueda de elementos, en primer lugar, porque ya se ha realizado el análisis de las proteínas que conforman estos datos, a través de softwares dedicados, como lo indica el equipo de investigación en neuroproteómica que proporcionó los datos. Así, la visión del presente trabajo será la búsqueda de patrones en determinadas secciones de la MS. Es importante recordar que, la MS se la debe comprender como un elemento en 3 dimensiones pertenecientes a los 3 principales datos que presentan sus resultados, estos son: M/Z, tiempo de retención e intensidad.

Uno de los retos que ha presentado el presente trabajo ha sido, tanto, la realización del análisis de la información, cuanto, la eficacia en el uso de los recursos computacionales del análisis. Con esta premisa, el autor del presente trabajo ha pensado que la mejor solución de este particular es realizar un análisis sectorizado de los resultados, fijando: determinados tiempos de retención y M/Z, para analizar los picos cromatográficos importantes por zona de interés, medidos en conteos por segundo (cps), tratando de buscar patrones, usando múltiples paquetes de R con cualidades específicas que nos permitan análisis distintos.

- **Resultados de la cromatografía, realizada con rangos de tiempo entre 2500 y 3500 segundos.**

En el siguiente gráfico podemos observar los resultados del individuo 1 del primer experimento. Lo que más llama la atención es que existe una variación en cada una de las muestras, pero que a pesar de ello presenta características comunes. Se puede observar un pico cromatográfico, el más alto de todos en las 3 primeras muestras, alrededor de los 3200 segundos, que en la muestra A, B y C se ubica en la zona b, en la muestra D se encuentra en la zona c, no siendo sino el segundo pico más intenso. Otra de las características que podemos observar en las 4 cromatografías, es la constancia de un pico cromatográfico que se encuentra entre los 3400 y 3500 segundos, que no llega a sobrepasar la línea divisoria entre las zonas c y f en las 2 primeras muestras, pero que sí lo hace en las muestras C y D. Las cromatografías más similares del individuo 1 son las de las muestras A y B. La muestra con mayores diferencias entre todas en la muestra C, que es la única que se presenta con 3 picos cromatográficos en tiempos de retención más apartados del común del resto. Como lo muestra el siguiente gráfico.

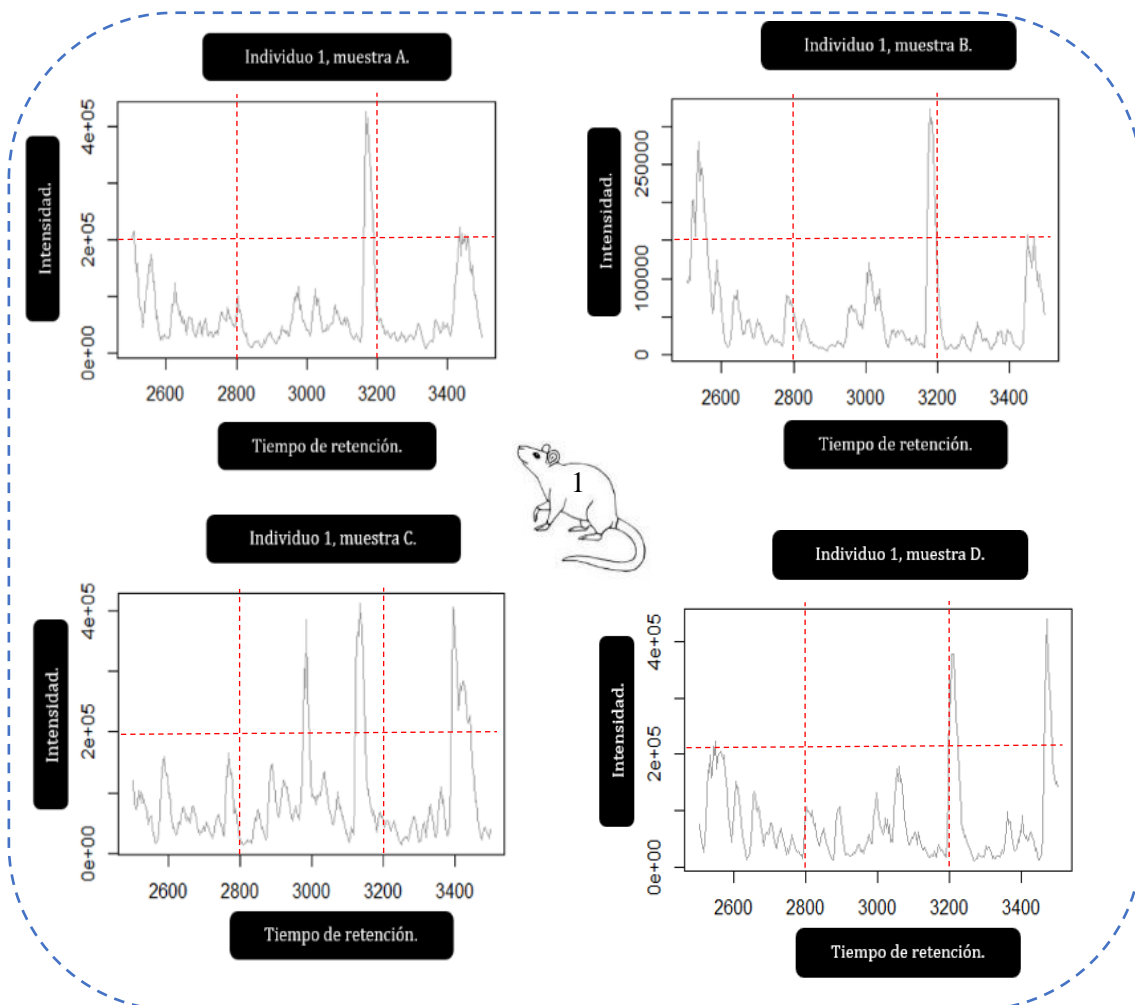


Gráfico 4. Resultados de la cromatografía, realizada con rangos de tiempo entre 2500 y 3500 segundos, del individuo número 1.

En el siguiente gráfico podemos observar que en la réplica biológica de cerebro del segundo individuo de estudio, podemos observar una similitud importante en los 2 principales picos cromatográficos que hemos enunciado en el párrafo previo. Es importante destacar la gran similitud que presentan los resultados del individuo 1, muestra A y del individuo 2, muestra A. Las cromatografías con mayor similitud en el individuo 2 son las de las muestras B y C, por la ubicación en el tiempo de retención de sus picos cromatográficos y las semejanzas en las zonas a, b, d y f. La muestra con mayores diferencias en el presente individuo es la muestra C, especialmente cuando observamos intensidad y forma irregular del segundo pico cromatográfico que se halla en la zona c, así como la aparición de 2 picos en la zona b.

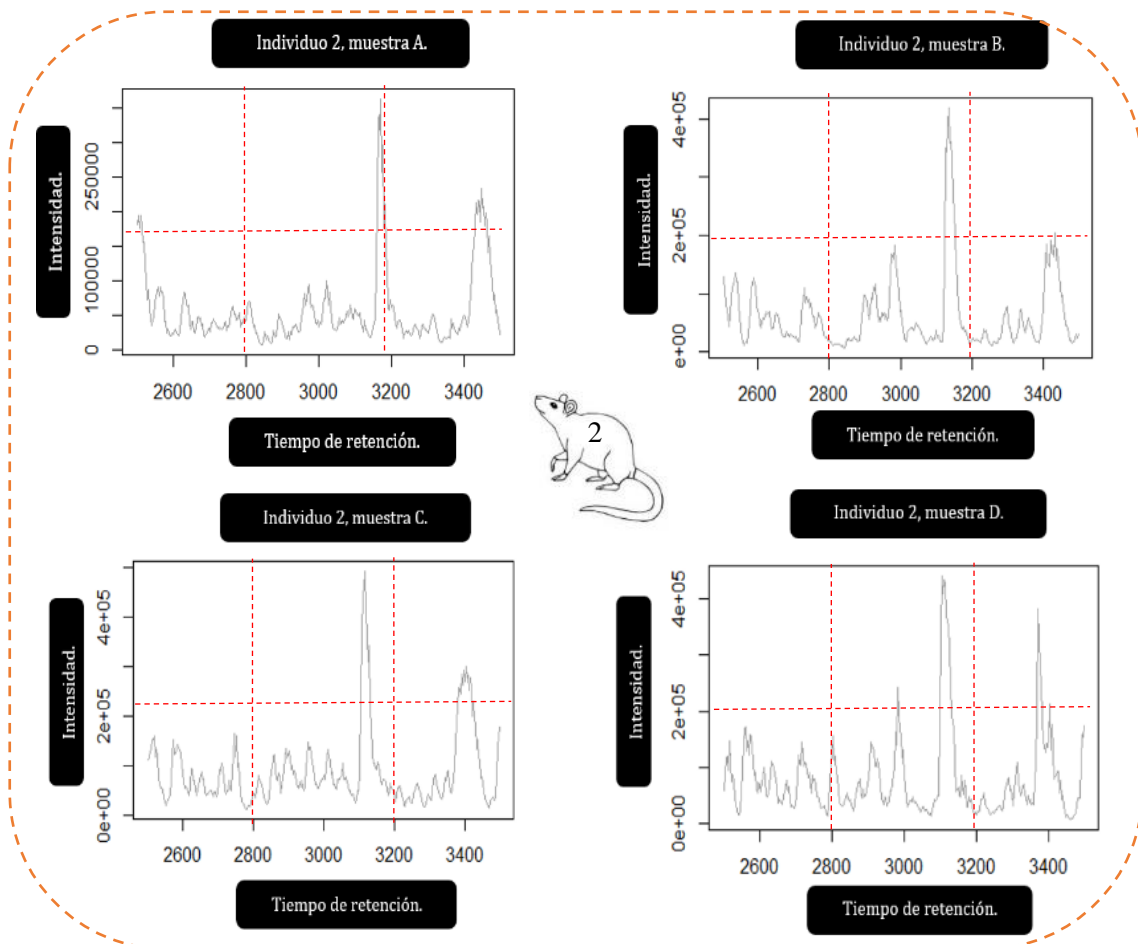


Gráfico 5. Resultados de la cromatografía, realizada con rangos de tiempo entre 2500 y 3500 segundos, del individuo número 2.

En el siguiente gráfico podemos observar la permanencia del pico cromatográfico que se evidencia alrededor de los 3200 segundos, el que hemos hallado en los 2 individuos anteriores. Llama la atención que el pico cromatográfico que se encuentra alrededor de los 3400 segundos

no es el segundo más intenso en la muestra A, existiendo un pico con una intensidad que logra superar la línea que divide las zonas superiores e inferiores. Realizando un análisis de los picos cromatográficos en cada zona de las 4 muestras de este individuo, podemos encontrar que las 2 muestras con mayores similitudes son las muestras A y C, ya que cuentan con 3 picos cromatográficos que logran superar la línea Y1, además, los tiempos de retención de los 3 picos son muy próximos al común en las 2 muestras. Es muy interesante resaltar la similitud entre la muestra B y D que, a pesar de que mediante el análisis que se ha llevado a cabo no concuerda objetivamente ya que los números de picos por zona cromatográfica son diferentes, presentan semejanzas en cuanto a la forma y ubicación de sus picos.

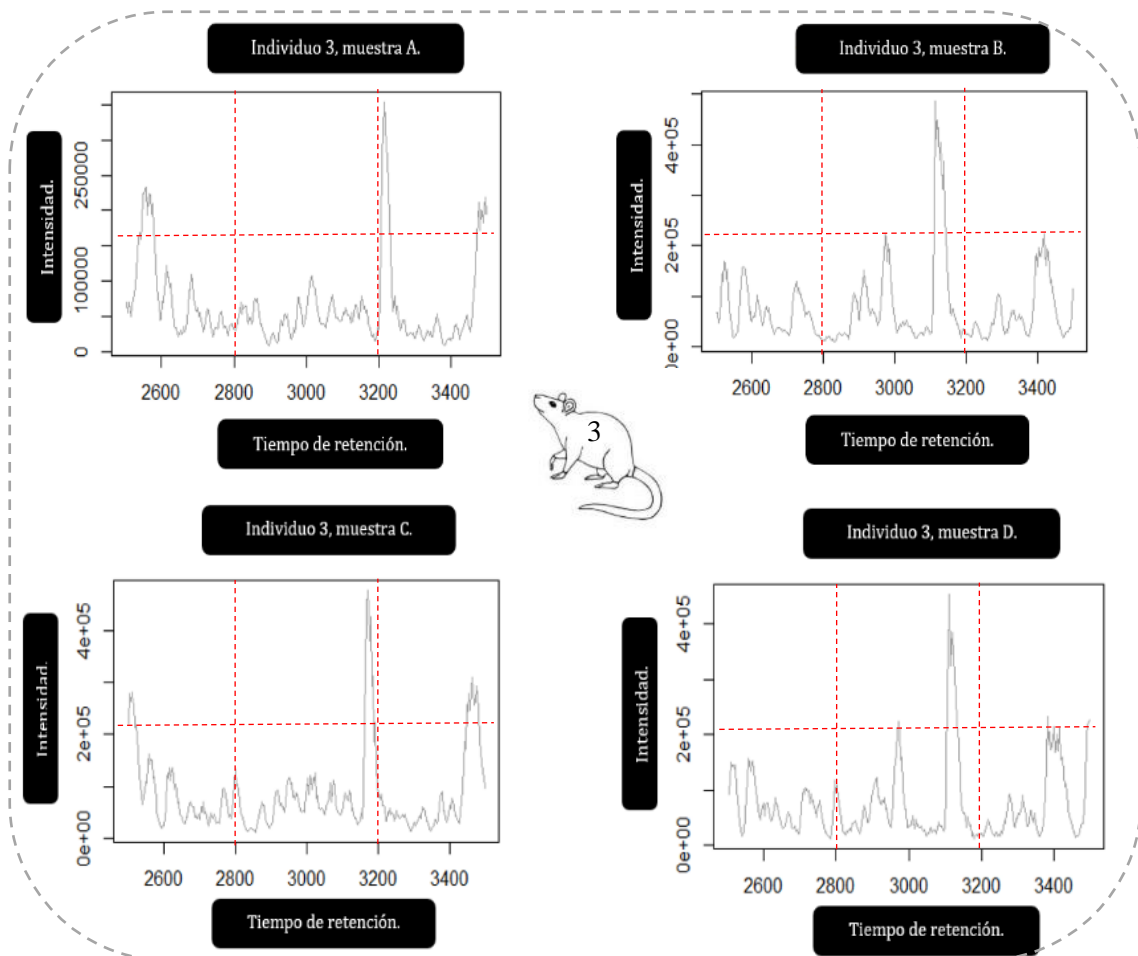


Gráfico 6. Resultados de la cromatografía, realizada con rangos de tiempo entre 2500 y 3500 segundos, del individuo número 3.

Cabe recalcar que la muestra A de los 3 individuos, es la que mayor semejanza presente entre las 3 réplicas biológicas. No solo cumple con presentar 3 picos cromatográficos que sobrepasan la línea Y, sino que la ubicación de estos se aproxima en cuanto al tiempo de retención.

- **Detección de picos cromatográficos usando el algoritmo *centWave* del paquete de R “*xcms: LC-MS and GC-MS Data Analysis*” utilizando un tiempo de retención entre 2500 y 3500 segundos y utilizando una brecha de *M/Z* entre 434.9 y 435.1.**

En el análisis de la información de este experimento vamos a centrarnos en la intensidad de los picos cromatográficos. La morfología de todos los resultados de la MS es muy similar, en este experimento, presentando un pico cromatográfico alrededor de los 2600 segundos, que se encuentra en la zona a, con una excepción, el resultado del individuo1, muestra C, presenta una morfología que difiere con las demás. Así, vamos a observar que en el siguiente gráfico la muestra A presenta un pico que llega a presentar una intensidad aproximada de 30000 cps, la muestra B de 20000 cps, la muestra C de 10000 cps y la muestra D de 30000 cps, similar a la primera. Con estos criterios, podemos decir que la muestra A y D son las que mayor semejanza presentan.

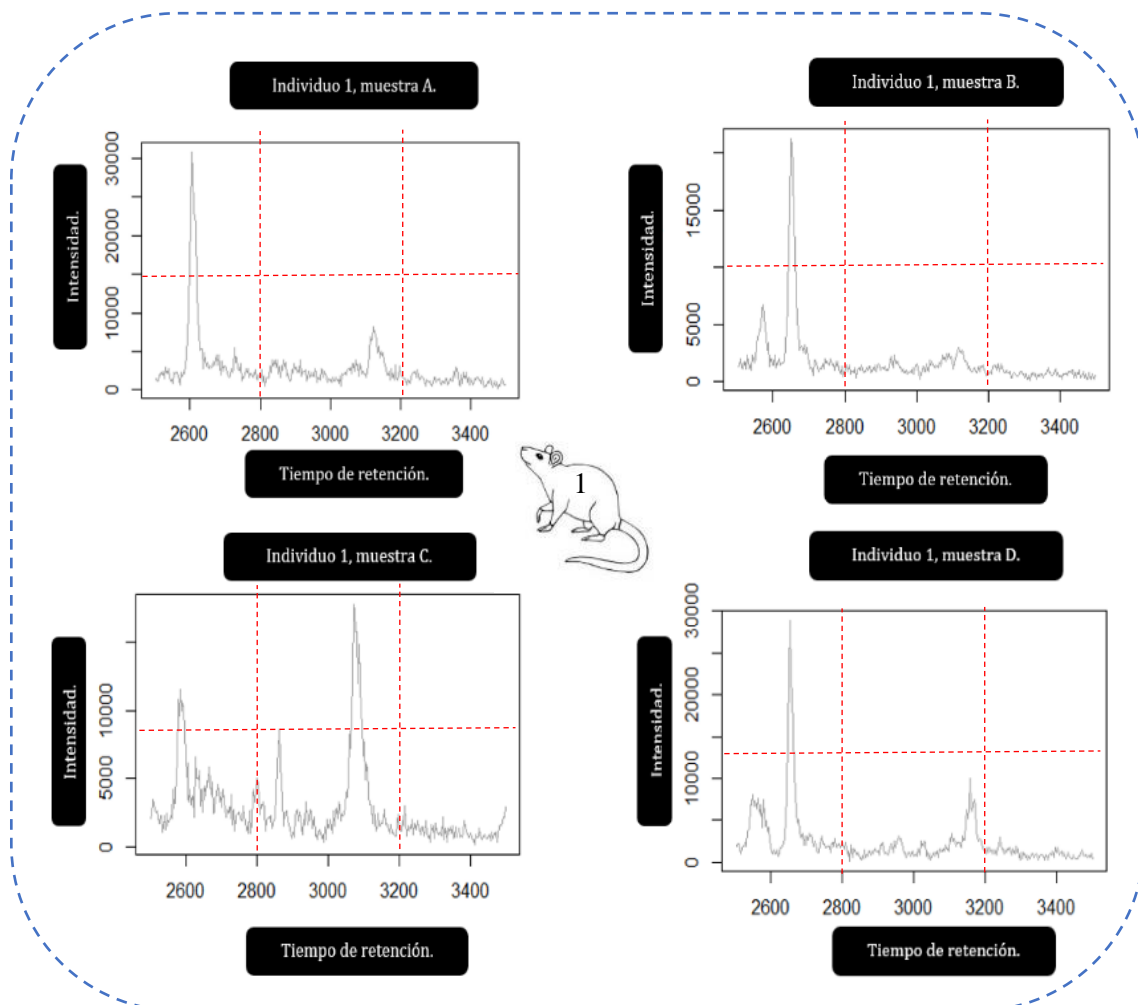


Gráfico 7. Detección de picos cromatográficos usando el algoritmo *centWave* del paquete de R “*xcms: LC-MS and GC-MS Data Analysis*” utilizando un tiempo de retención entre 2500 y 3500 segundos y utilizando una brecha de *M/Z* entre 434.9 y 435.1. Individuo número 1.

En el segundo individuo, podemos apreciar como los rangos en los que se miden las intensidades, igualmente, son distintos. El pico cromatográfico principal, que estamos estudiando en este experimento, que se encuentra alrededor de los 2600 segundos, presenta una intensidad de aproximadamente 2500 cps en el individuo A, de 2500 cps en el individuo B, de 2000 cps en el individuo C y de 3000 cps en el individuo D. En la muestra A y C podemos observar la presencia de un segundo pico cromatográfico en la zona e, así, en la muestra C y D podemos observar la ausencia de este elemento, lo que hace que estas 2 parejas, anteriormente mencionadas, sea muy similares en cuanto a la forma de sus estructuras. Aun así, por la intensidad que presentan los picos cromatográficos, podemos decir que, las 2 muestras con mayor semejanza son la A y la B, al presentar sus picos cromatográficos con una intensidad alrededor de 2500 cps.

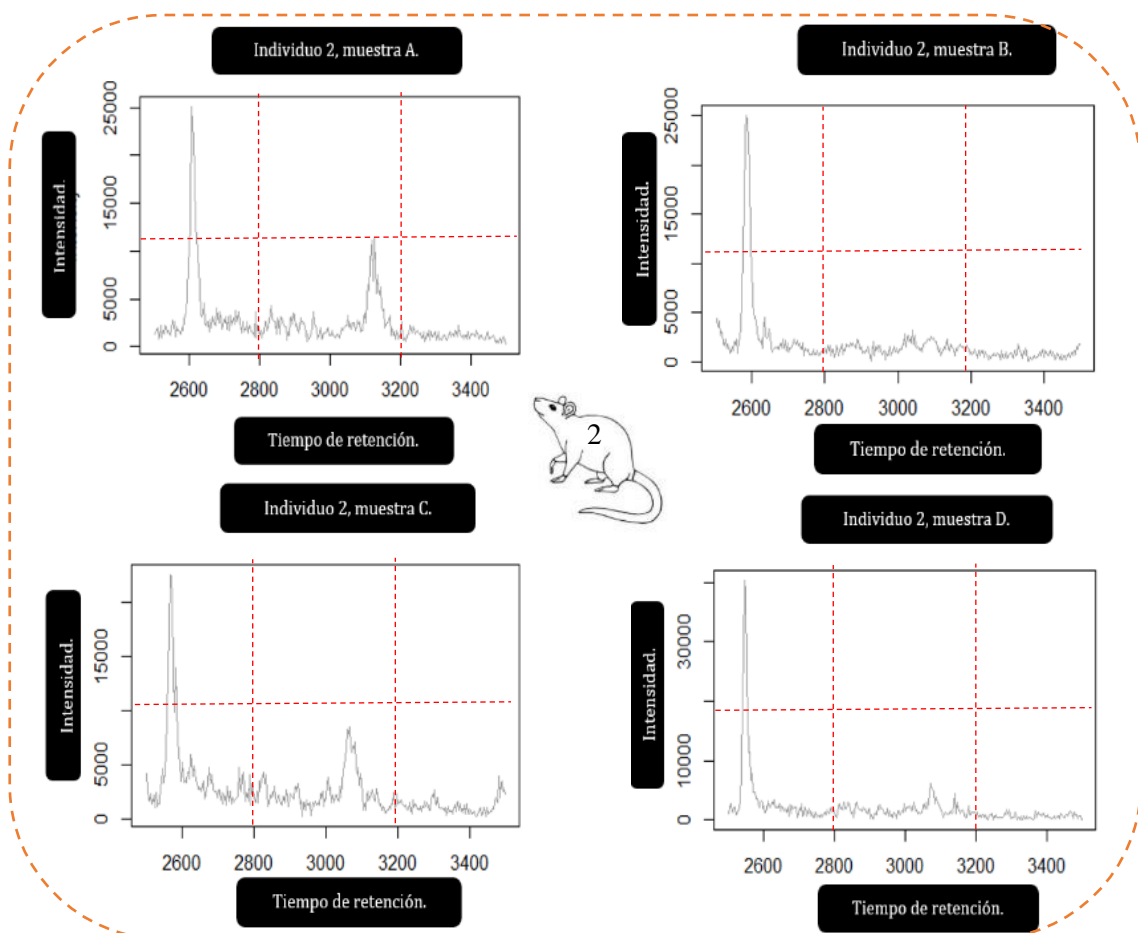


Gráfico 8. Detección de picos cromatográficos usando el algoritmo centWave del paquete de R “xcms: LC-MS and GC-MS Data Analysis” utilizando un tiempo de retención entre 2500 y 3500 segundos y utilizando una brecha de M/Z entre 434.9 y 435.1. Individuo número 1.



En el siguiente gráfico podemos observar cómo mantiene la morfología enunciada anteriormente, con ese pico inicial en la zona a en las 4 muestras de esta cuarta réplica biológica. Así podemos observar como la muestra A presenta una intensidad de aproximadamente 2500 cps, la muestra B de 3000 cps, la muestra C de 2000 cps y la muestra D de 2500 cps. Por lo que las 2 muestras que presentan mayor semejanza son la muestra A y D, en cuanto a los parámetros que estamos analizando.

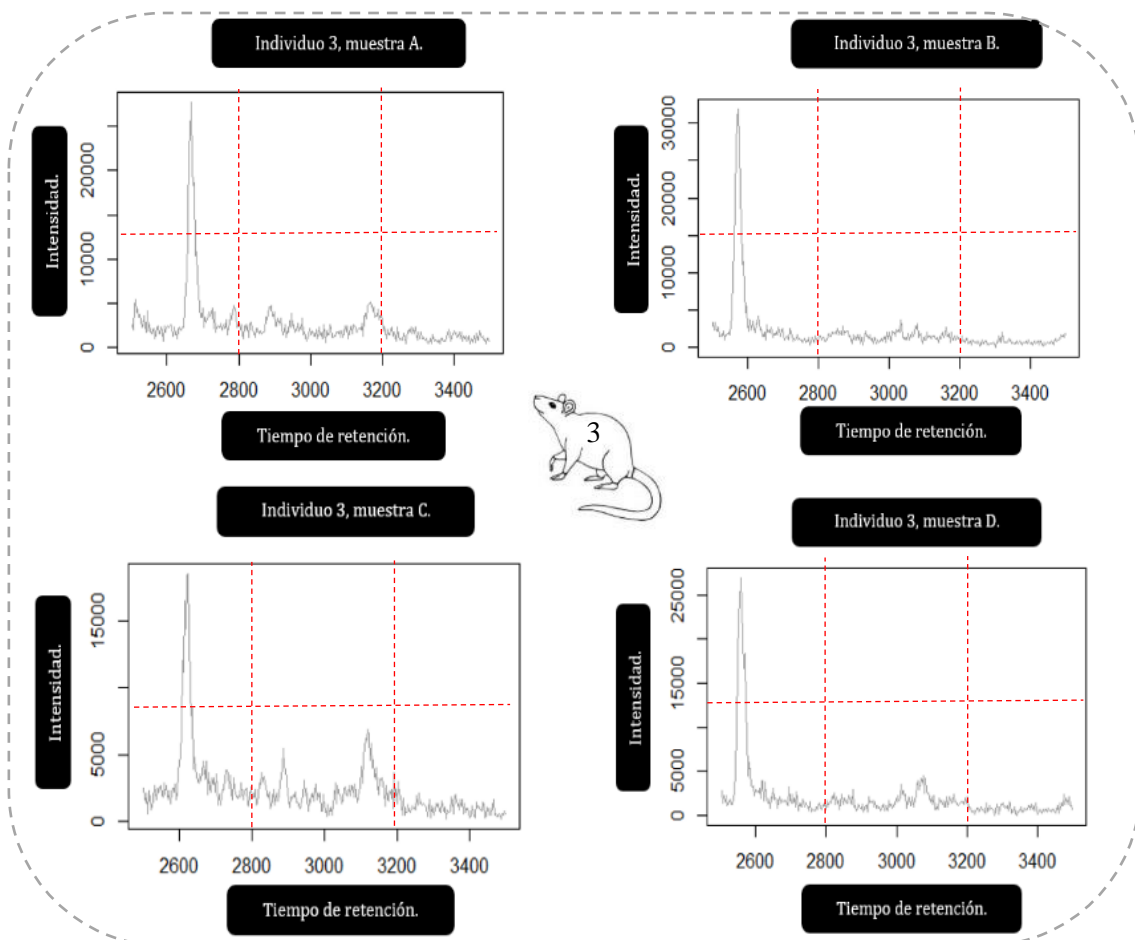


Gráfico 9. Detección de picos cromatográficos usando el algoritmo centWave del paquete de R “xcms: LC-MS and GC-MS Data Analysis” utilizando un tiempo de retención entre 2500 y 3500 segundos y utilizando una brecha de M/Z entre 434.9 y 435.1. Individuo número 3.

Un hallazgo interesante en este experimento ha sido la semejanza de la morfología del cromatograma de los datos de la MS, cabe recalcar dos aspectos importantes para tener en cuenta. El primero, se ha utilizado una muy limitada brecha de M/Z, segundo, se ha usado el algoritmo centWave para analizar la información. Estos dos aspectos parecen claves para poder presentar resultados con una consistencia más alta, que presenten resultados con mayor coherencia y que

los mismos puedan arrojar mayor claridad en cuanto a los procesos que se pueden realizar posteriormente con los resultados obtenidos. Exclusivamente el individuo 1, muestra C, es el único que presentó una anomalía en la morfología del cromatograma, lo que no deja de ser un hallazgo relevante. El resto de las muestras C, con el mismo protocolo de preparación, no presentaron esta anomalía.

- **Visualización cromatográfica de la relación que existe entre intensidad, M/Z y tiempo de retención, utilizando un tiempo de retención entre 2500 y 3500 segundos y utilizando una brecha de M/Z entre 434.9 y 435.1**

El siguiente experimento muestra una visualización cromatográfica un tanto diferente a la tradicional, se puede observar en un solo gráfico, de manera muy clara, la relación entre los tres datos centrales de la cromatografía de una MS: el tiempo de retención, la intensidad y la M/Z. Esta es una opción muy interesante a los gráficos en 3 dimensiones (3D) porque, sobre todo, ocupan muchos menos recursos computacionales que los que se necesitan para generar gráficos 3D de una información tan densa como lo son los datos masivos de una MS.

Al usar los mismos parámetros de estudio que en el experimento anterior, podemos observar la clara conformación de un patrón morfológico común, teniendo principalmente un pico cromatográfico de intensidad alrededor de los 2600 segundos y un pico cromatográfico de M/Z alrededor de los 3500 segundos. Otro patrón común que podemos observar en las mediciones de los conteos de M/Z es que la gran mayoría de conteos se ven acumulados sobre los 350.00 M/Z. Otro hallazgo interesante en este experimento es que en cada réplica biológica, al menos en una muestra, se observa un valor donde la intensidad se encuentra sobre los 4000 cps. En el caso del individuo 1, se observa en la muestra A, en el individuo 2, se observa en la muestra D, en el individuo 3, se observa en la muestra B. No se observa un claro patrón en cuanto a este resultado y el protocolo de preparación de la muestra.

En el siguiente gráfico observaremos que la intensidad del primer pico cromatográfico de la muestra A es la más elevada en comparación al resto de muestras de esta réplica biológica, llegando a los 5000 cps, mientras que la intensidad más baja del primer pico cromatográfico la podemos observar en la muestra C, donde llega aproximadamente a los 1500 cps. Las mediciones de los valores de M/Z son bastante dispersos, a excepción del pico cromatográfico mencionado previamente, no se observa un patrón claro de conteos, lo que se evidencia en forma de esa imagen dispersa. Lo que podemos apreciar en cuanto a las mediciones de la M/Z del primer individuo es que la mayoría se concentran sobre los 435.00.

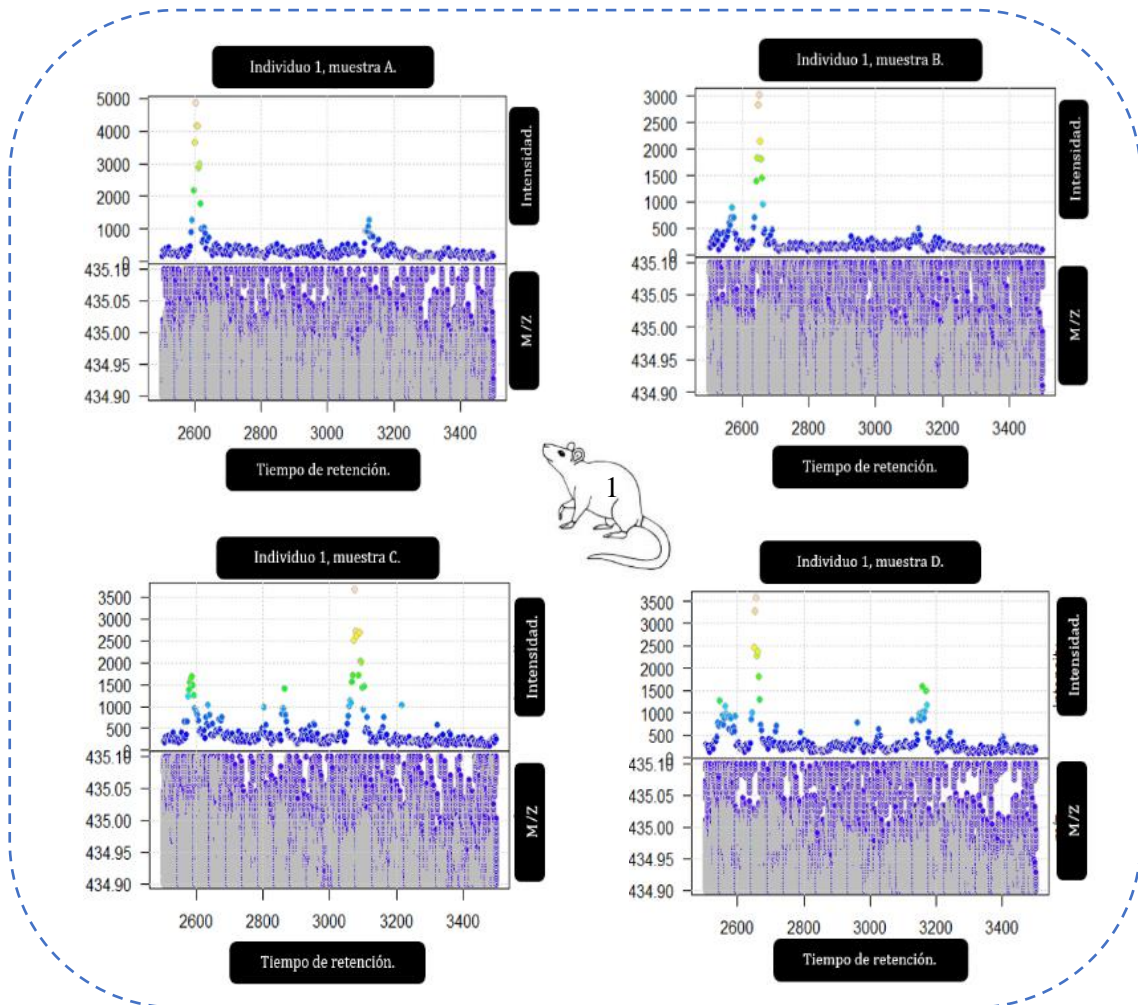


Gráfico 10. Visualización cromatográfica de la relación que existe entre intensidad, M/Z y tiempo de retención, utilizando un tiempo de retención entre 2500 y 3500 segundos y utilizando una brecha de M/Z entre 434.9 y 435.1. Individuo número 1.

En el siguiente gráfico podemos observar los resultados de las 4 muestras de la réplica biológica del individuo 2. Es evidente un mantenimiento de la morfología en los conteos de los picos cromatográficos de la MS. Podemos observar como el primer pico cromatográfico de la muestra A presenta una intensidad aproximada de 3500 cps, la muestra B presenta una intensidad aproximada de 3000 segundos, la muestra C de 3500 cps y, finalmente, la muestra D con una intensidad aproximada de 5000 cps. Así, tenemos que la muestra D es la que presenta mayor intensidad en el primer pico cromatográfico, la muestra B es la que presenta la menor intensidad en su pico cromatográfico, con 5000 y 3000 cps respectivamente.

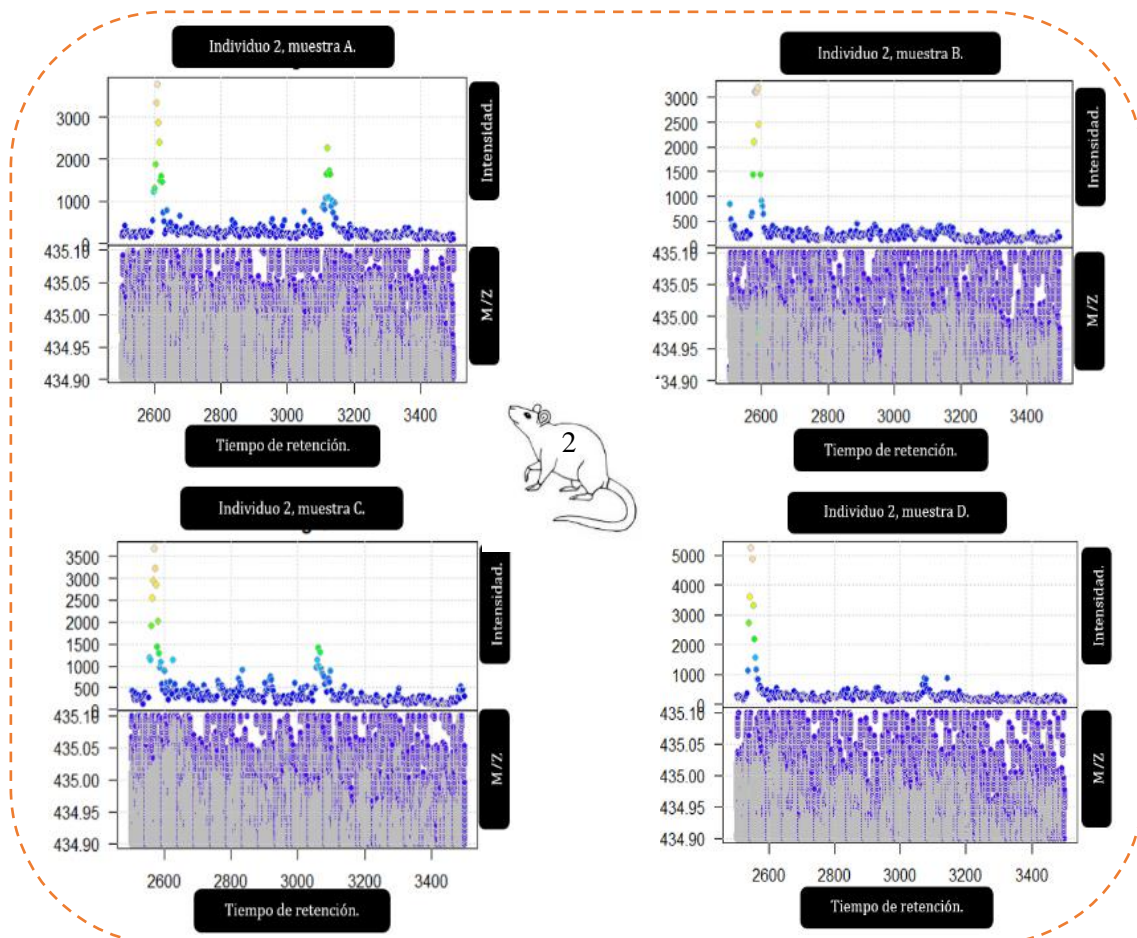


Gráfico 11. Visualización cromatográfica de la relación que existe entre intensidad, M/Z y tiempo de retención, utilizando un tiempo de retención entre 2500 y 3500 segundos y utilizando una brecha de M/Z entre 434.9 y 435.1. Individuo número 2.

En el siguiente gráfico se puede observar que el primer pico cromatográfico presenta un valor de 3500 cps en la muestra A, mientras que en la muestra B presenta aproximadamente 4000 cps, en la muestra C presenta 3000 cps y, finalmente, en la muestra D presenta aproximadamente 3500 cps, haciendo que las 2 muestras con mayor similitud sean las muestras A y D. La muestra con una intensidad más elevada es la muestra B. La muestra que presenta menor intensidad, podríamos decir que son las muestras A y D, pero, un hecho interesante es que en la muestra C existe un solo conteo que pasa el valor de 3000 cps, por lo que si comparamos el número de conteos por intensidad, la que presentaría una menor intensidad sería la muestra C.

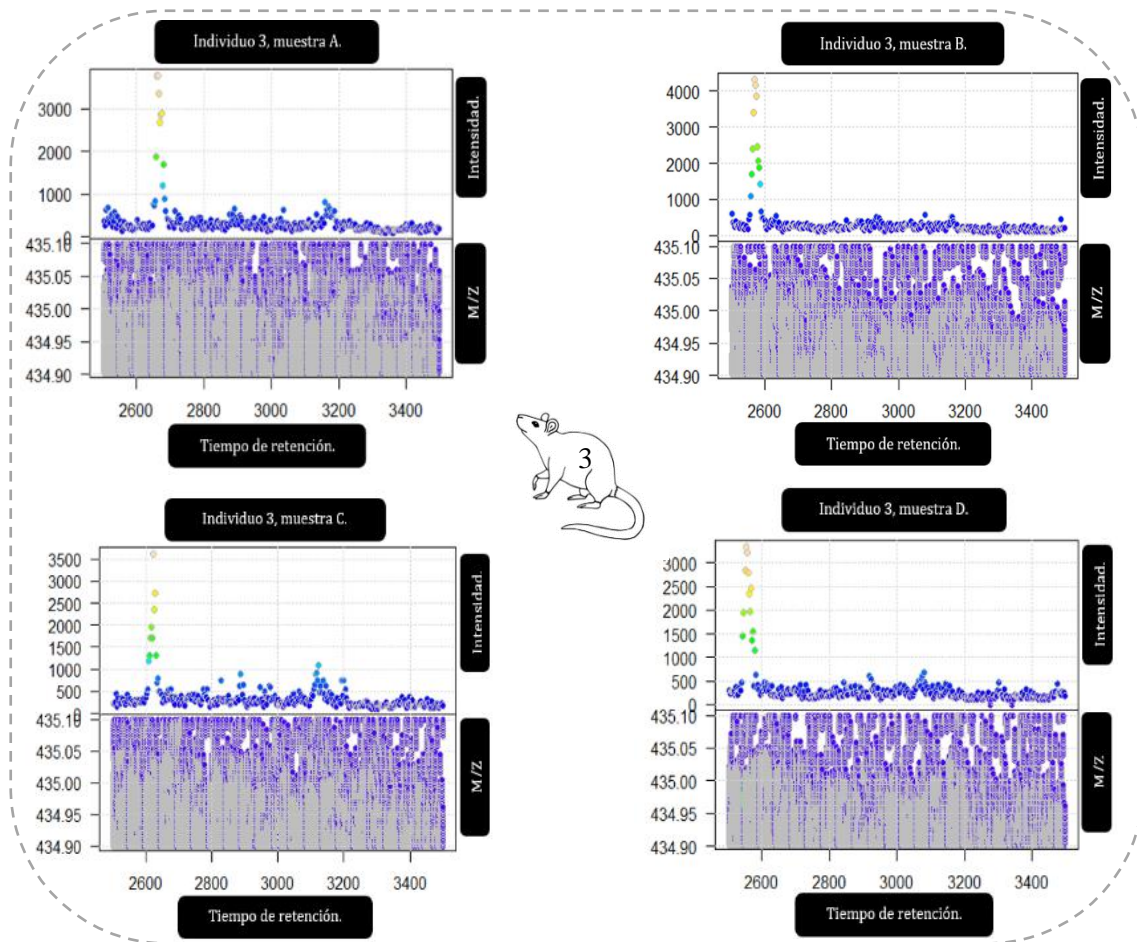


Gráfico 12. Visualización cromatográfica de la relación que existe entre intensidad, M/Z y tiempo de retención, utilizando un tiempo de retención entre 2500 y 3500 segundos y utilizando una brecha de M/Z entre 434.9 y 435.1. Individuo número 3.

- **Cromatograma realizado utilizando, exclusivamente, iones extraídos de las muestras, lo cual nos ayudará a diferenciar las distintas configuraciones de detección máxima, utilizando un tiempo de retención entre 2500 y 3500 segundos.**

El siguiente experimento presenta los resultados cromatográficos de la MS de cada muestra, con una particularidad, se marcarán los iones extraídos de las muestras, al usar el paquete “xcms” (55–57) podemos eliminar los conteos que se presenten con mayor probabilidad como “ruido”, marcando los sitios donde se condensan más los conteos, permitiendo de esta forma marcar y señalar los picos que presenten una mayor cantidad de conteos. No es que los demás conteos no se presenten en el cromatograma, pero los picos cromatográficos marcados presentaran una mayor cantidad y, por ende, mayor calidad de datos para representarlos.

En este contexto, podemos observar en el siguiente gráfico perteneciente a las 4 muestras de la réplica biológica del primer individuo de estudio, la muestra A presenta en total 4 diferentes

picos cromatográficos marcados, 1 en la zona a, 2 en la zona d y 1 en la zona. La muestra B presenta 3 diferentes picos cromatográficos, 1 en la zona a, 1 en la zona d y 1 en la zona c. El pico cromatográfico de la zona c de la muestra B es el pico más ancho de todos, a pesar de ello, se evidencia que no existe división en el mismo, lo que representa a un solo pico. Una de las ventajas de este experimento es que nos muestra las divisiones entre picos, pudiendo de esta forma contar el número total de picos cromatográficos con mayor certeza. En la muestra C evidenciamos que presenta 7 picos cromatográficos, 1 en la zona a, 1 en la zona b, 2 en la zona d, 2 en la zona e y, finalmente, 1 en la zona f. La muestra D, presenta 5 picos cromatográficos, 1 en la zona a, 2 en la zona d y 2 en la zona e. Algo muy interesante, es que esta muestra presenta 2 picos cromatográficos diferentes adyacentes, a diferencia del pico cromatográfico que encontramos en la muestra B.

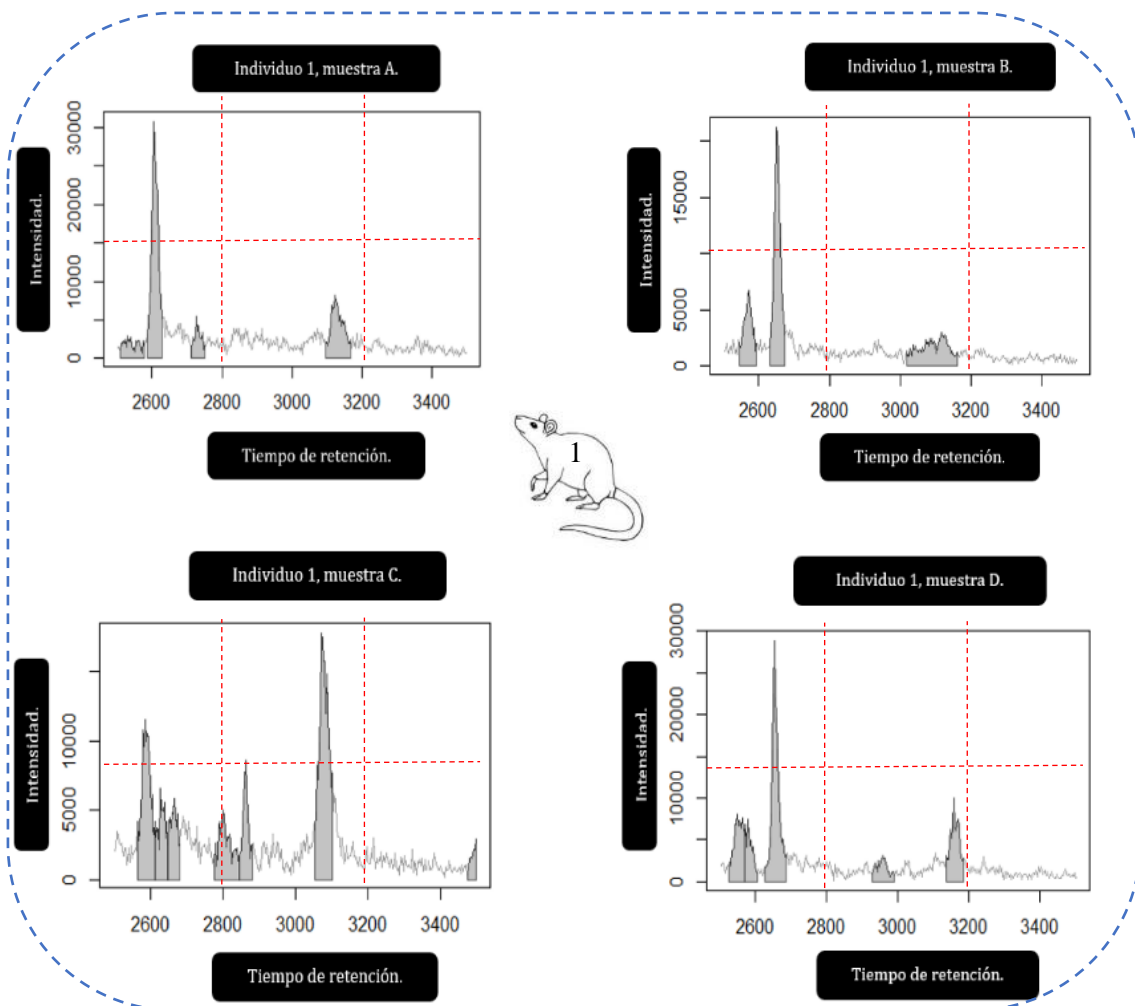


Gráfico 13. Cromatograma realizado utilizando, exclusivamente, iones extraídos de las muestras, lo cual nos ayudará a diferenciar las distintas configuraciones de detección máxima, utilizando un tiempo de retención entre 2500 y 3500 segundos, usando el algoritmo de centWave del paquete de R “xcms: LC-MS and GC-MS Data Analysis”. Individuo 1.

En el siguiente gráfico podemos observar cómo los resultados de este experimento en el individuo 2. Podemos observar como en la muestra A presenta 5 picos cromatográficos, 1 en la zona a, 2 en la zona d y 2 en la zona e. Como particularidad de esta muestra, podemos observar que los 3 primeros picos cromatográficos se encuentran unidos y, aun así, cada uno representa un pico cromatográfico individual. En la muestra B, observamos solamente 2 picos cromatográficos, 1 en la zona a y uno en la zona D. En la muestra C, se evidencia que existen 4 picos cromatográficos, 1 en la zona a, 2 en la zona e y 1 en la zona f. Mientras que en la muestra D, es posible observar 3 picos cromatográficos, 1 en la zona a y 2 en la zona e. Todas las muestras de la réplica biológica del segundo individuo presentan un pico cromatográfico en la zona a. El mismo que, a pesar de presentar una morfología común entre muestras al estar alrededor de los 2600 segundos, presenta una intensidad distinta en las muestras. De esta manera tenemos que, en la muestra A presenta 2500 cps de intensidad, en la B, también, 2500 cps de intensidad, en la C presenta 2000 cps de intensidad y en la D 3500 cps de intensidad.

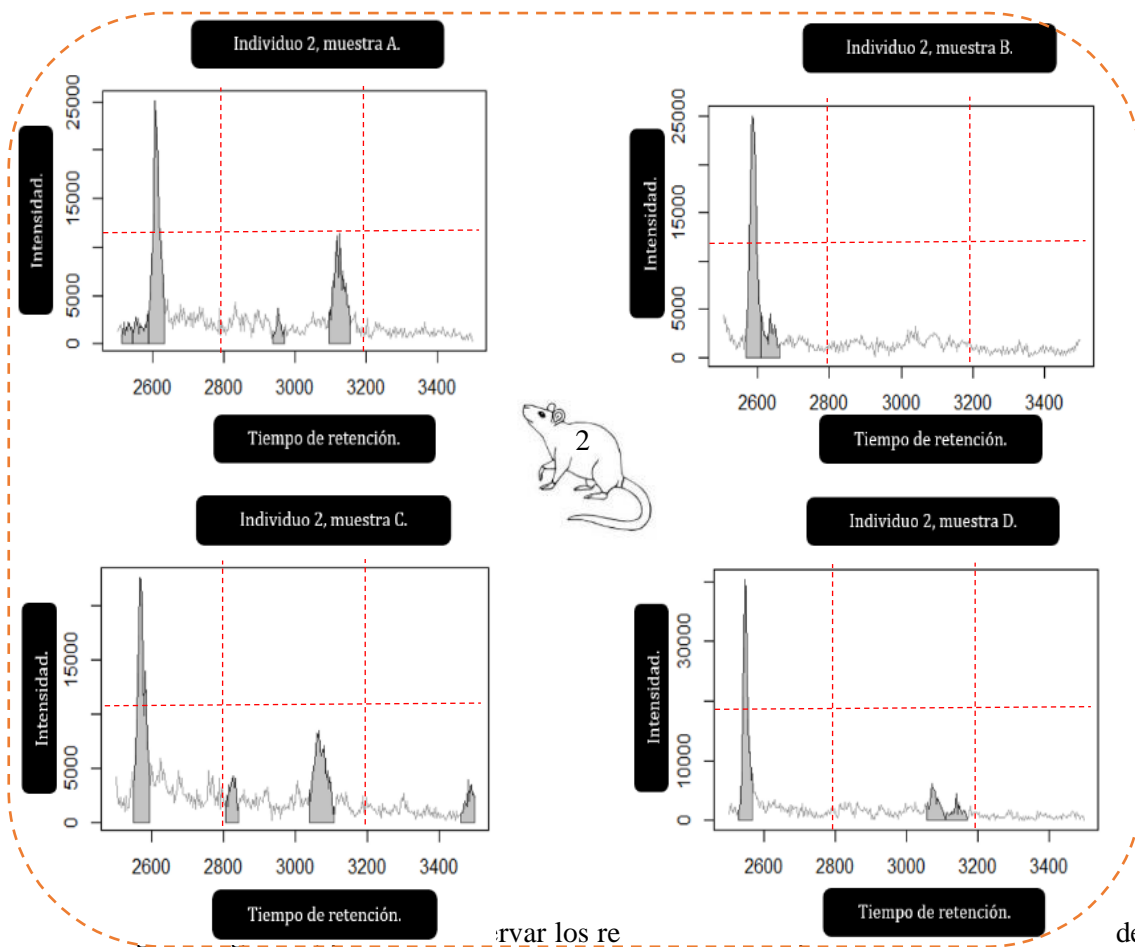


Gráfico 14. Cromatograma realizado utilizando, exclusivamente, iones extraídos de las muestras, lo cual nos ayudará a diferenciar las distintas configuraciones de detección máxima, utilizando un tiempo de retención entre 2500 y 3500 segundos, usando el algoritmo de *centWave* del paquete de R “*xcms: LC-MS and GC-MS Data Analysis*”. Individuo 2.

B presenta 4 picos cromatográficos, 1 en la zona a, 1 en la zona d y 2 en la zona e. En la muestra C se evidencian 6 picos cromatográficos, 1 en la zona a, 2 en la zona d y 3 en la zona e. En la muestra D podemos observar solamente 3 picos cromatográficos, 1 en la zona a, 1 en la zona e y 1 en la zona f. En esta muestra podemos evidenciar que el pico cromatográfico que se halla en la zona a, en cada muestra, presenta una intensidad distinta. El más intenso es el pico cromatográfico de la muestra B, que presenta 3000 cps. El menos intenso es el pico cromatográfico de la muestra C, que presenta aproximadamente 1500 cps.

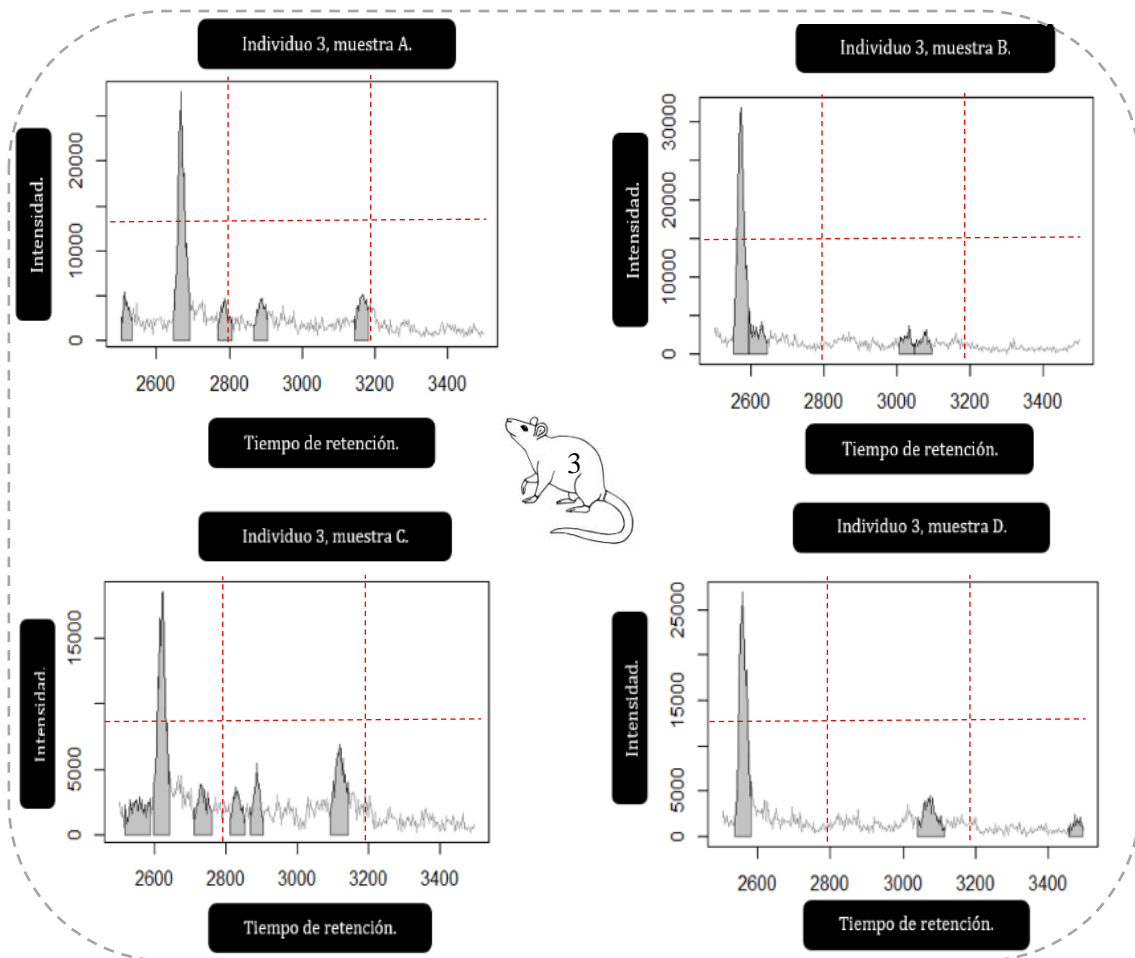


Gráfico 15. Cromatograma realizado utilizando, exclusivamente, iones extraídos de las muestras, lo cual nos ayudará a diferenciar las distintas configuraciones de detección máxima, utilizando un tiempo de retención entre 2500 y 3500 segundos, usando el algoritmo de centWave del paquete de R “xcms: LC-MS and GC-MS Data Analysis”. Individuo 3.

- **Visualización de picos cromatográficos en el conjunto de datos completo, considerando las señales con un valor de intensidad de los iones entre 6 a 5000 conteos por segundo (cps) en el paso de detección de picos, cubriendo todos los posibles picos, utilizando un tiempo de retención entre 2500 y 3500 segundos.**



El siguiente experimento tiene la particularidad de que vamos a observar los picos cromatográficos, sin restricción de M/Z, midiendo todos los picos de intensidad entre 6 y 5000 cps, lo que hará que veamos de una forma más completa el cromatograma en el lapso establecido. Uno de los problemas que se general al realizar la medición de datos de la MS y su posterior análisis cromatográfico, es que presenta demasiados picos en espacios de tiempo muy cortos, lo que dificulta su observación y la búsqueda de patrones.

En el siguiente gráfico podemos observar los datos del experimento del individuo 1. El número elevado de picos en períodos de tiempo de retención muy cortos hace muy complejo su análisis en busca de patrones. Lo más importante a recalcar en los resultados de este gráfico con los valores de la intensidad entre los que los picos se presentan. En la muestra A se presentan entre 0 y 2000 cps, en la muestra B, valores entre 0 y 1000 cps, en la muestra C, valores entre 0 y 3500 cps y finalmente en la muestra D, valores entre 0 y 1500. De esta manera, podríamos decir que las 2 muestras con mayor similitud son la muestra A y D, al presentar las mediciones entre rangos con más proximidad en comparación a las demás.

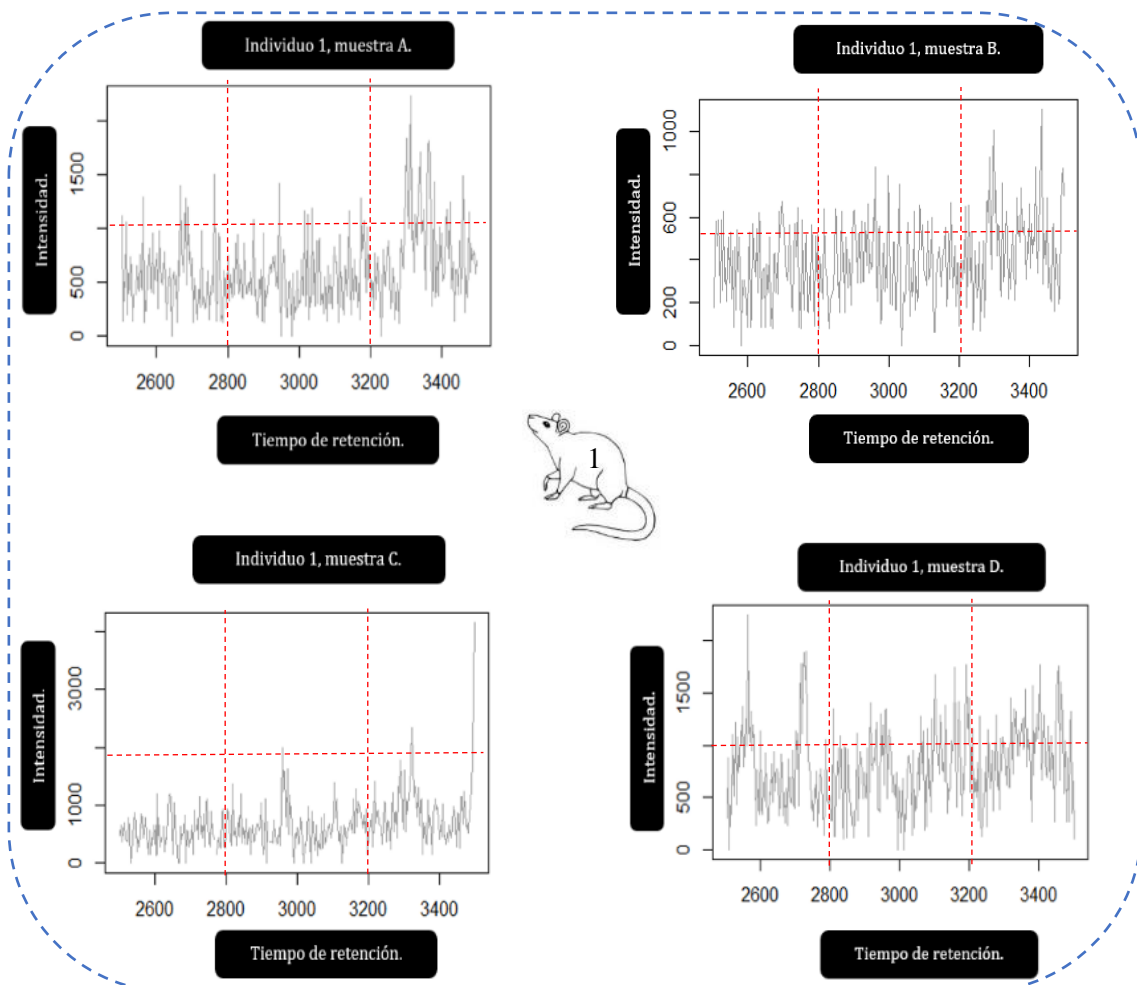


Gráfico 16. Visualización de picos cromatográficos en el conjunto de datos completo, considerando las señales con un valor de intensidad de los iones entre 6 a 5000 cps en el paso de detección de picos, cubriendo todos los posibles picos, utilizando un tiempo de retención entre 2500 y 3500 segundos. Individuo 1.

En el siguiente gráfico podemos observar los resultados del experimento de las muestras pertenecientes al segundo individuo. En estos resultados es interesante encontrar que en las muestras B, C y D podemos hallar una similitud, vemos un pico cromatográfico, que es el de mayor intensidad en cada una de esas muestras, en un tiempo específico alrededor de los 3500 segundo de tiempo de retención. Siendo los más similares las muestras C y D. En cuanto a la intensidad entre la que se encuentran los picos cromatográficos de las muestras, la muestra A se encuentra entre 0 y 2000 cps, la muestra B se encuentra entre 0 y 2000 cps, la muestra C entre 0 y 5000 cps y, finalmente, la muestra D entre 0 y 6500 cps.

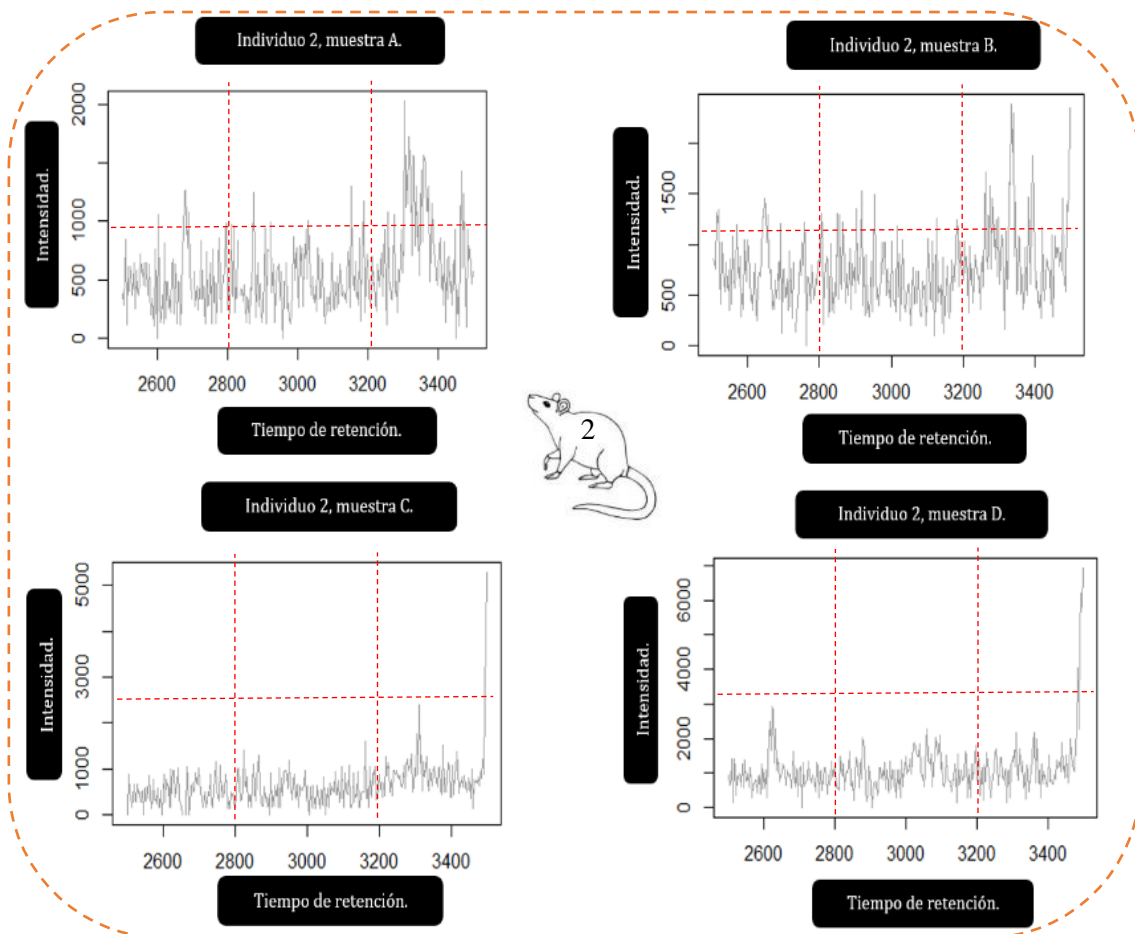


Gráfico 17. Visualización de picos cromatográficos en el conjunto de datos completo, considerando las señales con un valor de intensidad de los iones entre 6 a 5000 cps en el paso de detección de picos, cubriendo todos los posibles picos, utilizando un tiempo de retención entre 2500 y 3500 segundos. Individuo 2.

En el siguiente gráfico encontraremos los resultados de las 4 muestras pertenecientes a la réplica biológica proporcionada por el tercer individuo. Vemos que existe un patrón similar entre la muestra B y D, con un pico cromatográfico, el más elevado en cada muestra, que se halla

aproximadamente en los 3500 segundos. La intensidad de la muestra A va desde 0 a 1500 cps, la muestra B entre 0 y 5000 cps, la muestra C entre 0 y 2000 cps y, finalmente, la muestra D entre 0 y 5000 cps. Haciendo que las muestras B y C presenten muchas similitudes con estos parámetros de estudio.

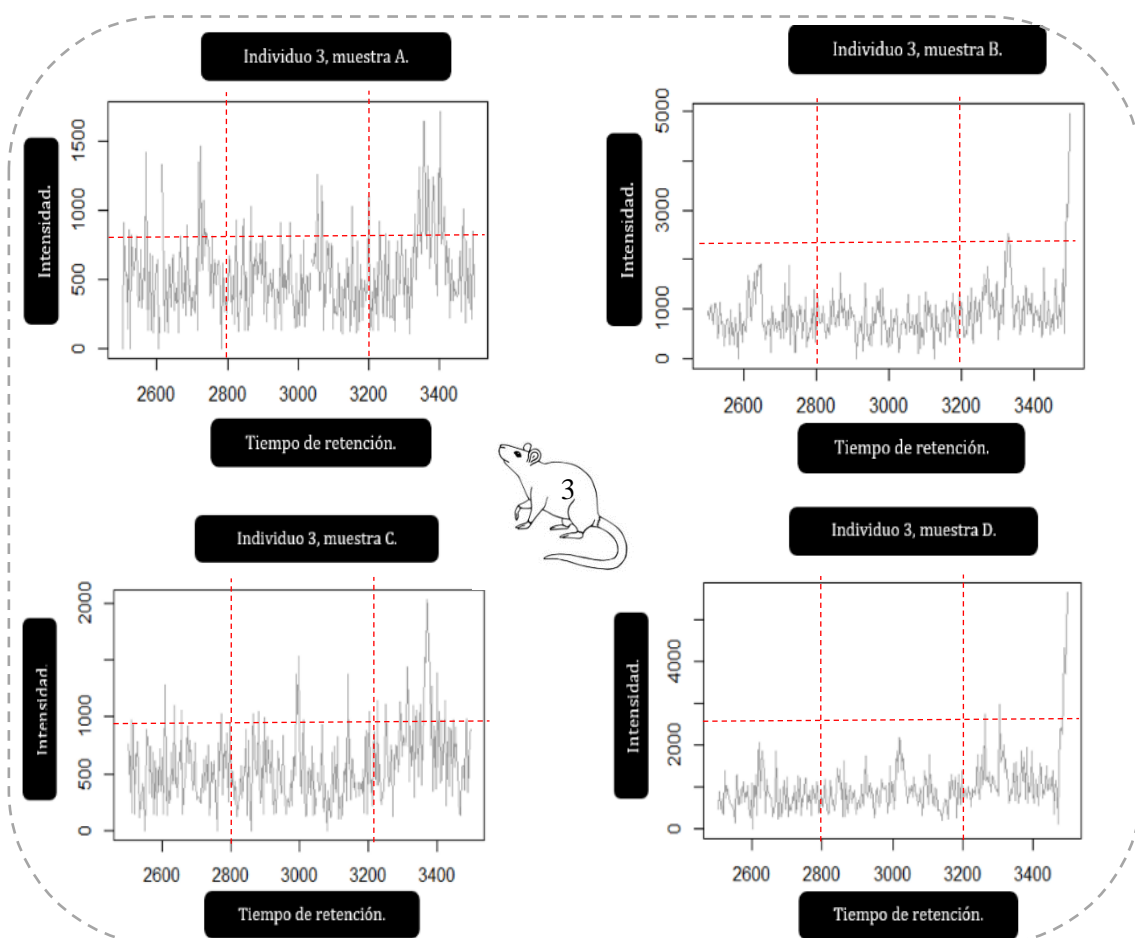


Gráfico 18. Visualización de picos cromatográficos en el conjunto de datos completo, considerando las señales con un valor de intensidad de los iones entre 6 a 5000 cps en el paso de detección de picos, cubriendo todos los posibles picos, utilizando un tiempo de retención entre 2500 y 3500 segundos. Individuo 3.

- **Tabla resumen del conteo total de picos por muestra e individuo, con un conteo del número de picos que presento un tiempo de retención relativo (0 %,25%, 50%, 75% y 100% de tiempo de retención), visualizando así, el número de picos identificados por muestra y ancho/duración de los picos cromatográficos.**

En el siguiente experimento hemos analizado el porcentaje de conteos que se hallan en todo el cromatograma de la MS, se ha contabilizado 5 puntos clave como marcadores del patrón que

puede seguir el número de conteos. Así, se ha visto los porcentajes de tiempo de retención de 0%, del 25%, 50%, 75% y del 100%. Se ha obtenido esta información de cada una de las muestras. Como lo indica la siguiente tabla 2, que podemos encontrarla en la sección de anexos.

Para que la información que hemos obtenido pueda ser visualizada de una manera más clara, se ha realizado el siguiente gráfico, donde podemos observar cómo existen algunos patrones importantes en cuanto a los resultados del experimento realizado. Lo que más llama la atención del siguiente resultado es que en los 3 individuos de estudio, las muestras D son las que presentan mayor número de conteos en todos porcentajes de tiempos de retención analizados. De la misma forma, en los 3 individuos, las muestras B son las que presentan menor número de conteos con tiempos de retención en los tiempos de retención de 100%. Podemos observar además, como los datos del segundo individuo son los que menor variación tienen entre resultados, mientras que los datos del primer individuo son los que mayor variación tienen entre resultados.

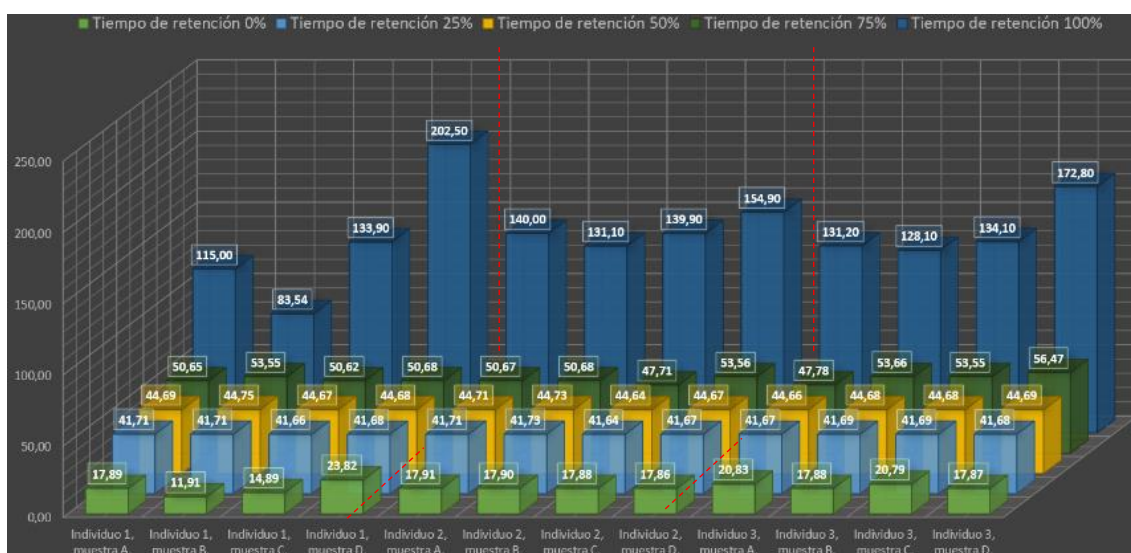


Gráfico 19. El siguiente gráfico presenta el conteo total de picos por muestra e individuo, con un conteo del número de picos que presento un tiempo de retención relativo (0%, 25%, 50%, 75% y 100% de tiempo de retención), visualizando así, el número de picos identificados por muestra y ancho/duración de los picos cromatográficos.

- **Visualización de la frecuencia de los picos cromatográficos identificados a lo largo del eje del tiempo de retención, identificando así, la frecuencia de los picos detectados a lo largo del eje de tiempo de retención. Lo que nos permite identificar períodos de tiempo a lo largo de la ejecución de la espectrometría de masa con una mayor cantidad de picos cromatográficos.**

En el siguiente experimento podemos observar la frecuencia de los picos de retención en el tiempo de retención determinado, entre 2500 y 3500 segundos. Entre más oscuro sea el color, mayor condensación de conteos de picos cromatográficos se ubicarán en ese período de tiempo en concreto. Lo cual nos va a permitir identificar períodos de tiempo en el transcurso de la MS donde se presenten un mayor número de picos cromatográficos. Esta es una forma bastante interesante de poder visualizar un aspecto importante de la MS.

En el siguiente gráfico observaremos los resultados del experimento del primer individuo de estudio. Podemos evidenciar como, en las 4 muestras, alrededor de los 3000 segundos vamos a observar una condensación de picos cromatográficos. Es muy importante recalcar que la muestra C, es la que mayor número de condensación de picos cromatográficos presente en una mayor cantidad de momentos del tiempo de retención. Mientras que la muestra B, al contrario, es la que menos número de presenta.

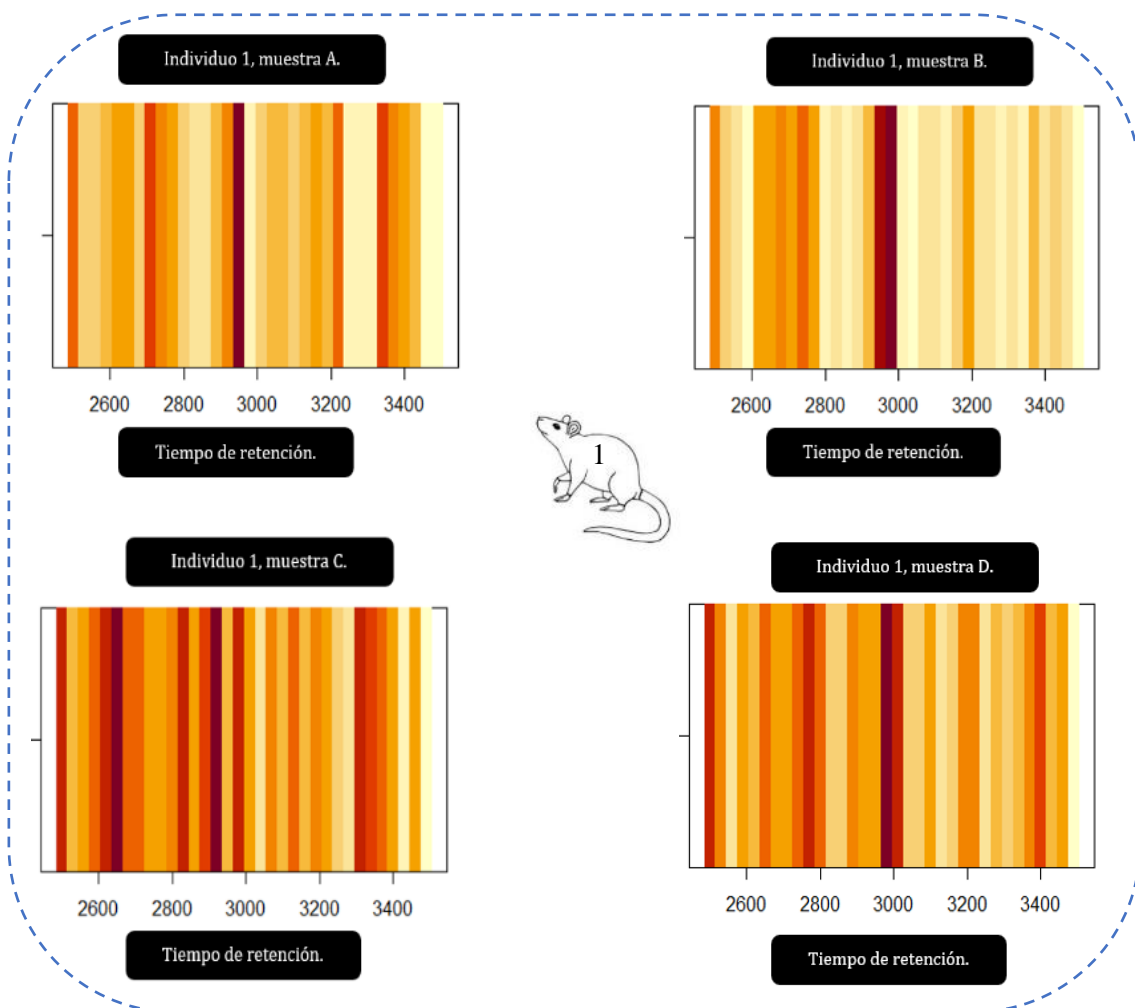
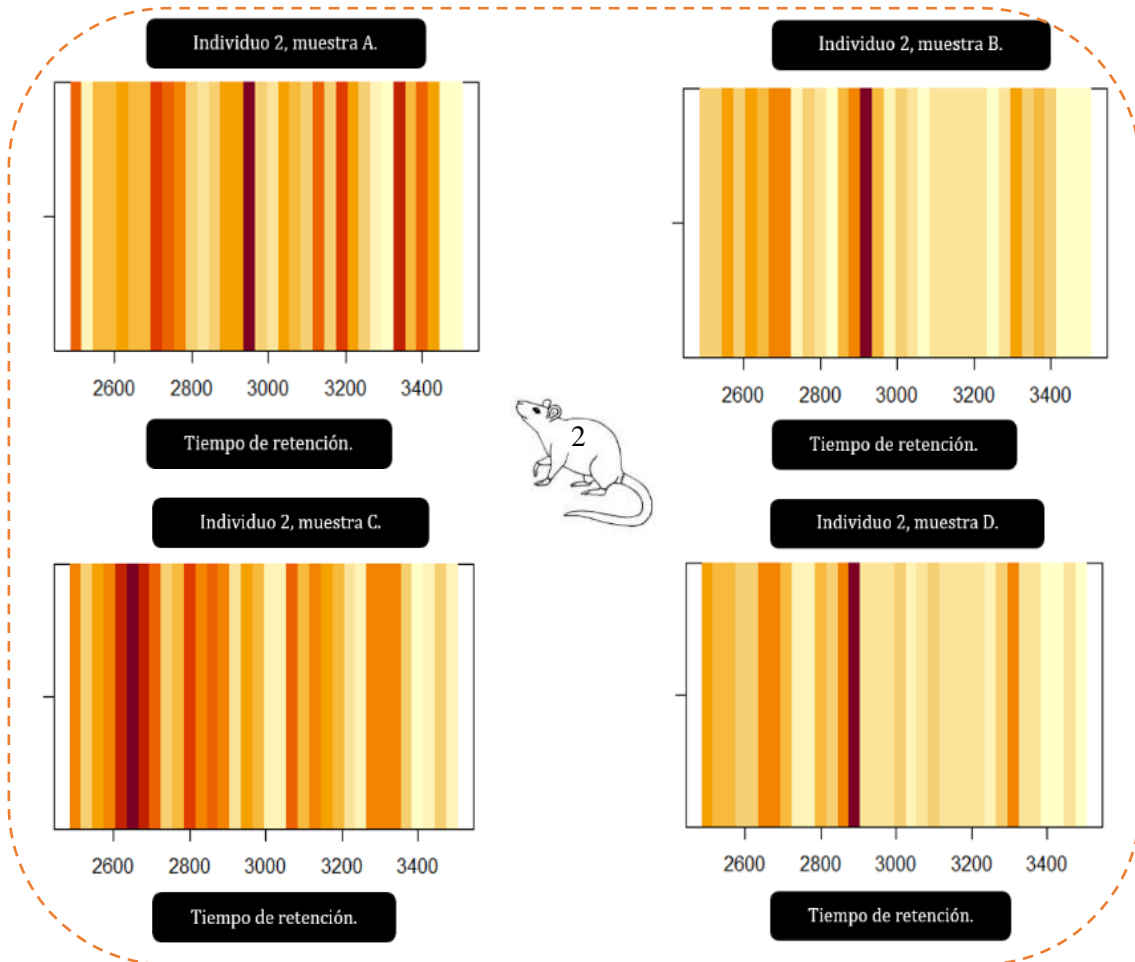


Gráfico 20. Visualización de la frecuencia de los picos cromatográficos identificados a lo largo del eje del tiempo de retención, identificando así, la frecuencia de los picos detectados a lo largo del eje de tiempo de retención. Lo que nos permite identificar períodos de tiempo a lo largo de la ejecución de la espectrometría de masa con una mayor cantidad de picos cromatográficos, visualizando de manera más comprensible las diferencias entre cada una de las muestras en cuento este aspecto. Individuo 1.

En el siguiente gráfico vamos a observar los resultados del experimento del segundo individuo. Podemos evidenciar como alrededor de los 3000 segundos se presenta una condensación de picos cromatográficos, al igual que en la imagen previa, con excepción de la muestra C. Además, la muestra C es la que más número de picos cromatográficos presenta en un mayor número diverso de momentos del tiempo de retención, que, igualmente, coincide con el gráfico anterior.



*Gráfico 21. Visualización de la frecuencia de los picos cromatográficos identificados a lo largo del eje del tiempo de retención, identificando así, la frecuencia de los picos detectados a lo largo del eje de tiempo de retención. Lo que nos permite identificar períodos de tiempo a lo largo de la ejecución de la espectrometría de masa con una mayor cantidad de picos cromatográficos, visualizando de manera más comprensible las diferencias entre cada una de las muestras en cuenta este aspecto. Individuo 2.*

En el siguiente gráfico vamos a observar los resultados del experimento del tercer individuo. En este gráfico podemos observar como la muestra C, coincidente con los anteriores dos gráficos, muestra una mayor variabilidad, con un mayor número de condensaciones de picos cromatográficos en un mayor número de momentos diferentes del tiempo de retención.

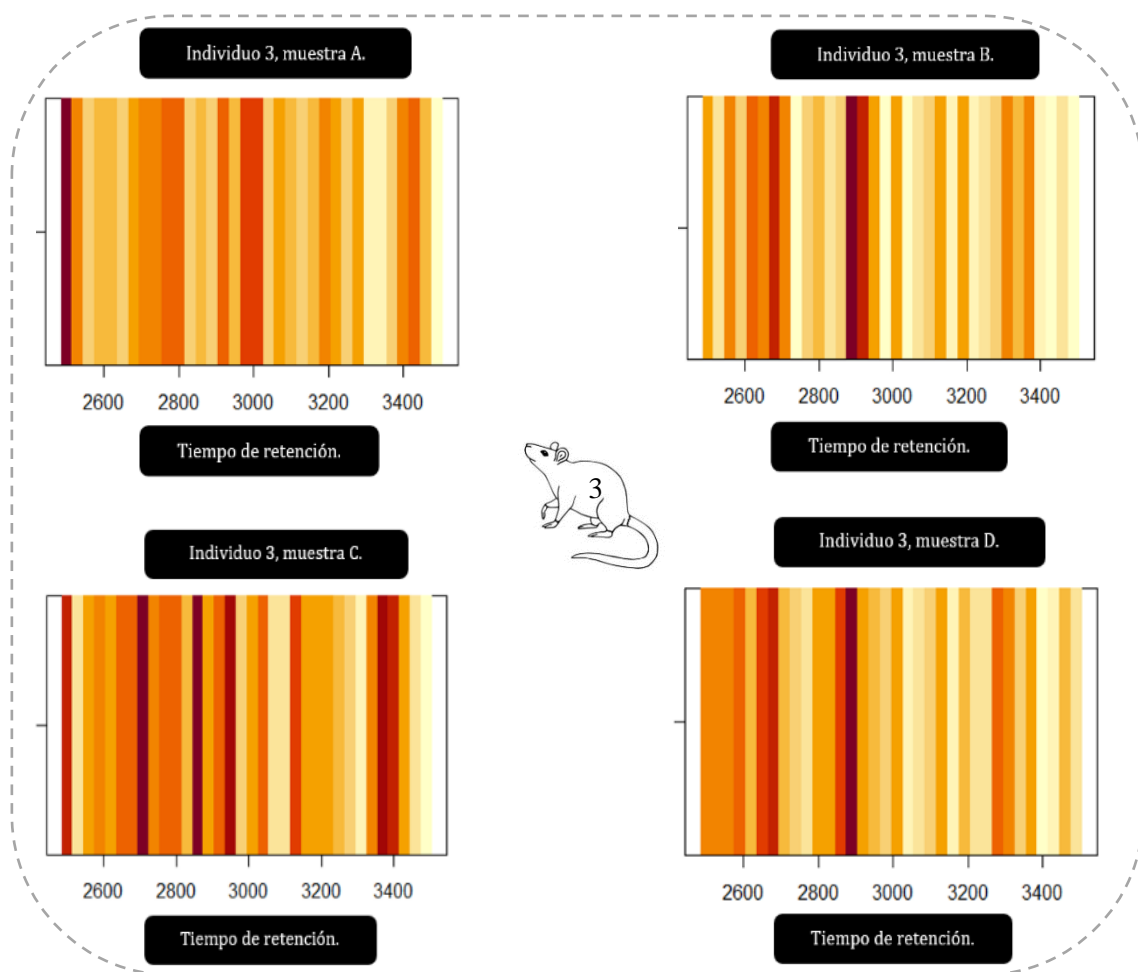


Gráfico 22. Visualización de la frecuencia de los picos cromatográficos identificados a lo largo del eje del tiempo de retención, identificando así, la frecuencia de los picos detectados a lo largo del eje de tiempo de retención. Lo que nos permite identificar períodos de tiempo a lo largo de la ejecución de la espectrometría de masa con una mayor cantidad de picos cromatográficos, visualizando de manera más comprensible las diferencias entre cada una de las muestras en cuento este aspecto. Individuo 3.

En el siguiente experimento, se ha realizado una medición de los picos cromatográficos y se ha señalado mediante escala de grises la densidad del número de mediciones de cada pico cromatográfico, entre más oscuro una mayor cantidad de mediciones. Además, se ha dividido cada pico cromatográfico para poder apreciar correctamente donde inicia y acaba cada uno. Esto, usando un valor más amplio de M/Z, entre 465 y 499.

En el siguiente gráfico podemos observar los resultados de las muestras de este experimento del primer individuo. Vemos que la morfología del cromatograma tiene una clara tendencia, mostrando mayor densidad de mediciones e intensidad al inicio del cromatograma y, como se

puede observar, disminuyendo la densidad de las mediciones e intensidad al final del cromatograma.

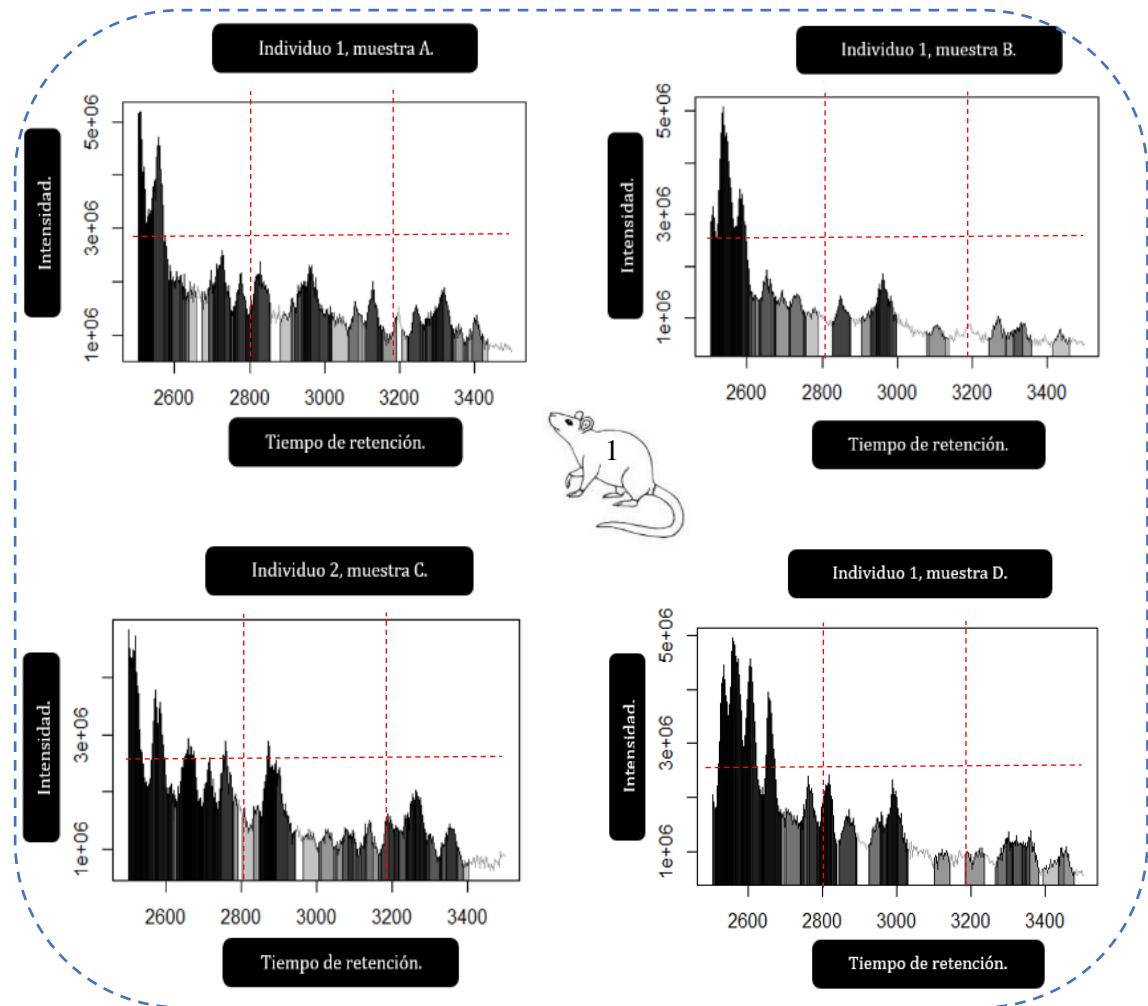


Gráfico 23. Visualización cromatográfica individualizando los picos para un valor específico de  $M/Z$  entre 465 y 499, con un tiempo de retención entre 2500 y 3500 segundos. Individuo 1.

En el siguiente gráfico podemos observar los resultados de este experimento de las muestras del segundo individuo de estudio. Vemos como el patrón que hemos mencionado anteriormente, una mayor intensidad con mayor densidad de mediciones al inicio del cromatograma, con una disminución de la densidad y de la intensidad al final del cromatograma, se repite en todas las muestras de este individuo. Algo que llama la atención es que, la muestra C, presenta densidades bastante marcadas al final del cromatograma, se lo puede apreciar en la zona e, lo cual sería una particularidad en las muestras de este individuo.



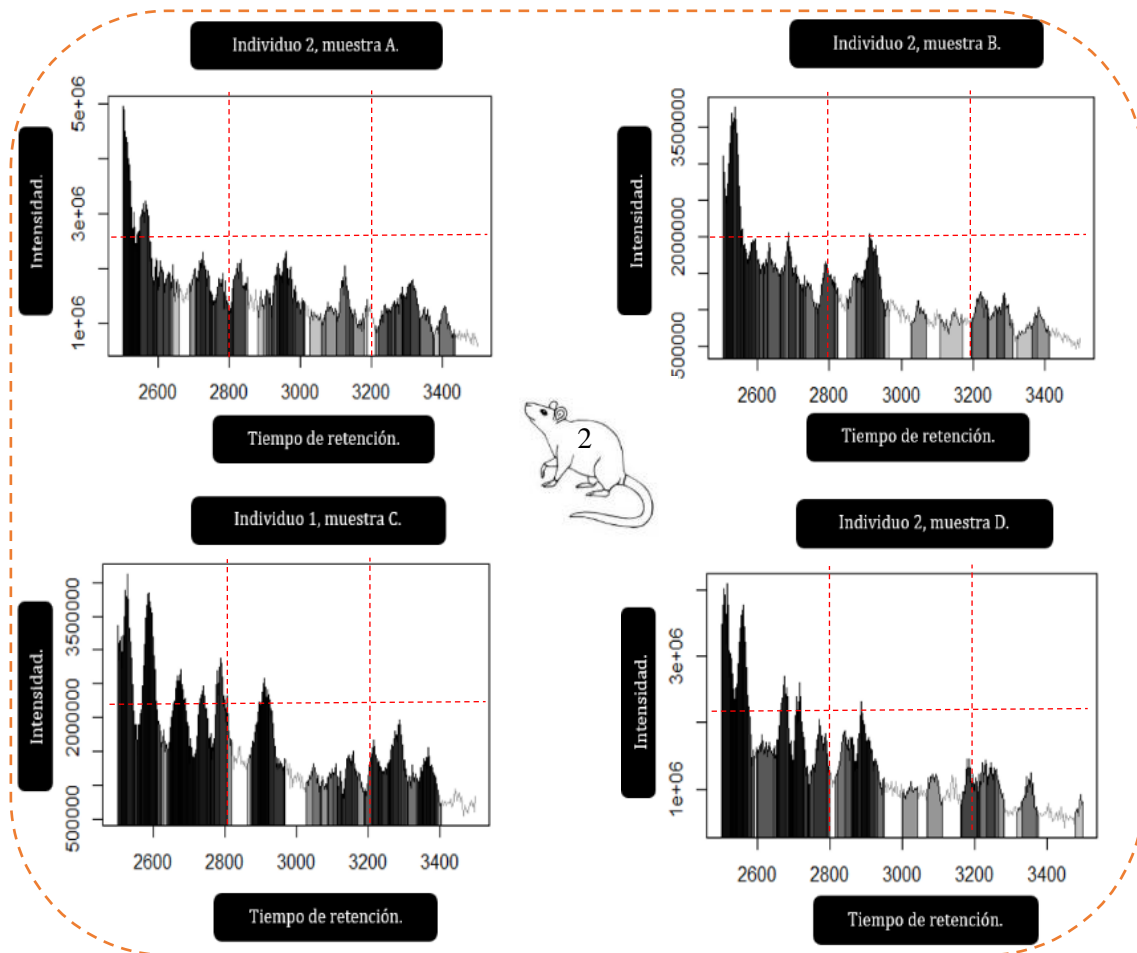


Gráfico 24. Visualización cromatográfica individualizando los picos para un valor específico de  $M/Z$  entre 465 y 499, con un tiempo de retención entre 2500 y 3500 segundos. Individuo 2.

En el siguiente gráfico podemos observar los resultados del experimento de las muestras del tercer individuo de estudio. En este gráfico, se hace evidente que en las muestras 3 réplicas biológicas se ha repetido el mismo patrón morfológico de la cromatografía. Haciendo de este experimento, uno de los que más consistencia ha presentado en cuanto a la variabilidad entre muestras y preparaciones.

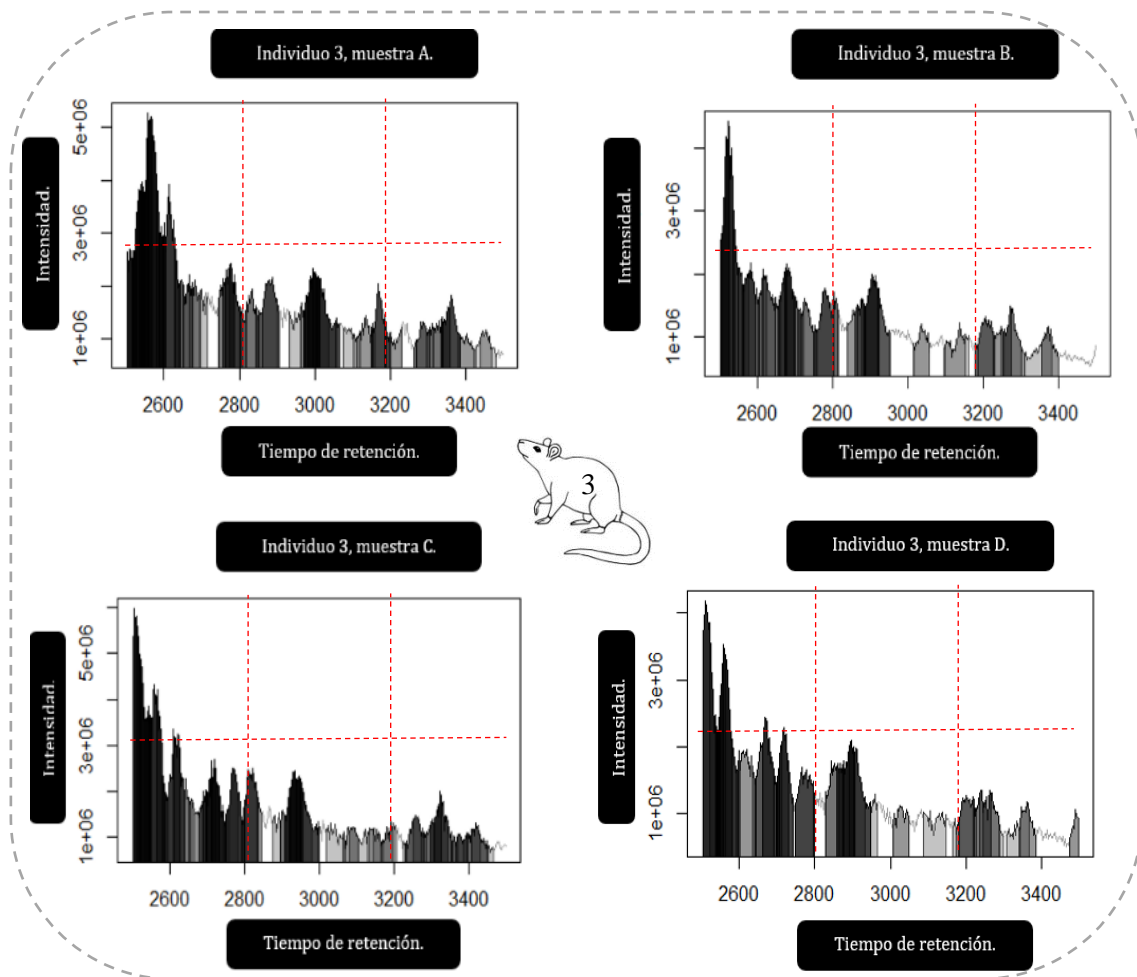


Gráfico 25. Visualización cromatográfica individualizando los picos para un valor específico de  $M/Z$  entre 465 y 499, con un tiempo de retención entre 2500 y 3500 segundos. Individuo 3.

- **Análisis estadístico de los datos de  $M/Z$ , tiempo de retención, tiempo de retención máximo y mínimo e intensidad por cada muestra, con un  $M/Z$  entre 750 y 1500, con un tiempo de retención entre 2500 y 3500 segundos.**

Para el siguiente experimento se ha medido en cada muestra de las 3 réplicas biológicas:  $M/Z$ , tiempo de retención, tiempo de retención máximo, tiempo de retención mínimo e intensidad. Se ha realizado usando una mayor amplitud de  $M/Z$ , entre 750 u 1500. Esto permitirá registrar una mayor cantidad de información, al mismo tiempo que se optimizan los recursos computacionales. Una parte importante de este experimento es que no se ha graficado los datos obtenidos, sino que

se ha obtenido toda la información de cada uno de los conteos, para lo que luego fue necesario el uso del paquete “jamovi ”(62) para procesar toda esta cantidad de información, así, obtener la media, mediana, desviación estándar, mínimo y máximo de cada uno de los datos mencionados anteriormente. Con lo que se ha obtenido los siguientes datos que muestra la siguiente tabla.

Algo que se puede apreciar en la siguiente tabla es la diferencia entre el número de mediciones que se realiza en cada muestra, teniendo como el número mínimo de conteo de mediciones: 76 mediciones, en los datos del individuo 1 , muestra B , y el número máximo de conteo de mediciones: 313 mediciones, en los datos del individuo 2, muestra C.

La tabla 3, resumen de las medidas de dispersión, desviación estándar, valor máximo y mínimo de los parámetros: M/Z, tiempo de retención, tiempo de retención máximo y mínimo, e intensidad en cps, se puede encontrar en la sección de anexos.

- **Huella de datos del análisis estadístico de los datos de M/Z, tiempo de retención, tiempo de retención máximo y mínimo e intensidad por cada muestra, con un M/Z entre 750 y 1500. Muestra la relación espacial de los diferentes datos.**

Para este experimento hemos usado los datos, en bruto, obtenidos en el experimento anterior. Uno de los retos de la gran cantidad de datos de la MS, no retratados en un cromatograma, es poder darles sentido mediante la capacidad de procesamiento del ser humano. Las tablas, aunque muy útiles en muchas ocasiones, no son una herramienta que permita entender un gran número de datos per se, mucho menos su relación entre sí. Por lo que se ha buscado una manera, sin usar la cromatografía, de explicar la relación que los datos pueden presentar.

Se ha creado un tipo de gráfico, que el autor del presente trabajo de investigación no ha logrado encontrar en otro estudio de investigación sobre el tema, que pueda ser de utilidad para expresar los datos obtenidos en una MS. El autor se ha inspirado en la capacidad que el cerebro humano tiene por encontrar patrones y poder encontrar pequeñas diferencias en formas básicas. (64–67) Es impensable pensar que la habilidad humana de reconocer patrones sea superior a la “habilidad” de una “maquina” en reconocer los patrones que le pedimos que busque. Lo que se pretende con es modelo de gráfico expuesto a continuación, es dar una imagen que el cerebro humano logre entender de los datos masivos que una MS puede presentar, y que pueda ver la relación que esos datos puedan presentar. Inspirado en la similitud con las huellas digitales del ser humano, el nombre que se ha usado para nominar a este tipo de gráfico es “huella de datos”.

Para poder entender la huella de datos es importante comprender que es lo que está mostrando. En la parte externa podemos observar los números que corresponden a cada una de las mediciones de la MS, dentro encontramos los datos de cada una de las mediciones que ha realizado la MS, podemos ver la M/Z, el tiempo de retención (rt), el tiempo de retención mínimo (rtmin), tiempo de retención máximo (rtmax) y la intensidad dividida para 10000 (into/10000), esto se ha hecho para normalizar las cantidades y que sea posible ver claramente la relación de los datos entre sí, sin alterar su estructura.

Con la estructura creada, podemos observar claramente no solo la relación entre datos, sino la variabilidad de cada tipo de dato. Así, una mayor variabilidad de datos nos mostrará muchas zonas con “picos”, un resultado con poca variabilidad nos mostrará un gráfico más aplanado. Algo muy importante a destacar es que cada pico se presentará e la sección correspondiente para el número de medición que produjo el dato. Así, podemos observar la intensidad de cada uno de los diferentes datos de resultado: M/Z, rt, rtmin, rtmax e into/10000, en cada una de las mediciones correspondientes, observando de esta forma la relación que presentan. Esto produce formas similares, con patrones reconocibles, pero al mismo tiempo formas únicas en cada una de las muestras, como si de una huella digital se tratara.

En el siguiente gráfico podemos observar la huella de datos que producen las muestras de la réplica biológica del primer individuo de estudio. Vemos como la muestra B es la que menor variabilidad presente, mientras que la muestra D es la que mayor variabilidad presenta en cuanto a los tiempos de retención. Sí dividiéramos cada huella de datos en 4 cuadrantes, podemos observar que, como patrón común en todas las muestras, existe un pico en la intensidad en el cuadrante superior izquierdo, que sería, en todos los casos, el de mayor intensidad. Este pico tendría una intensidad de 2500 cps en la muestra A, de 1000 cps en la muestra B, de 2500 cps en la muestra C y de 2000 cps en la muestra D.

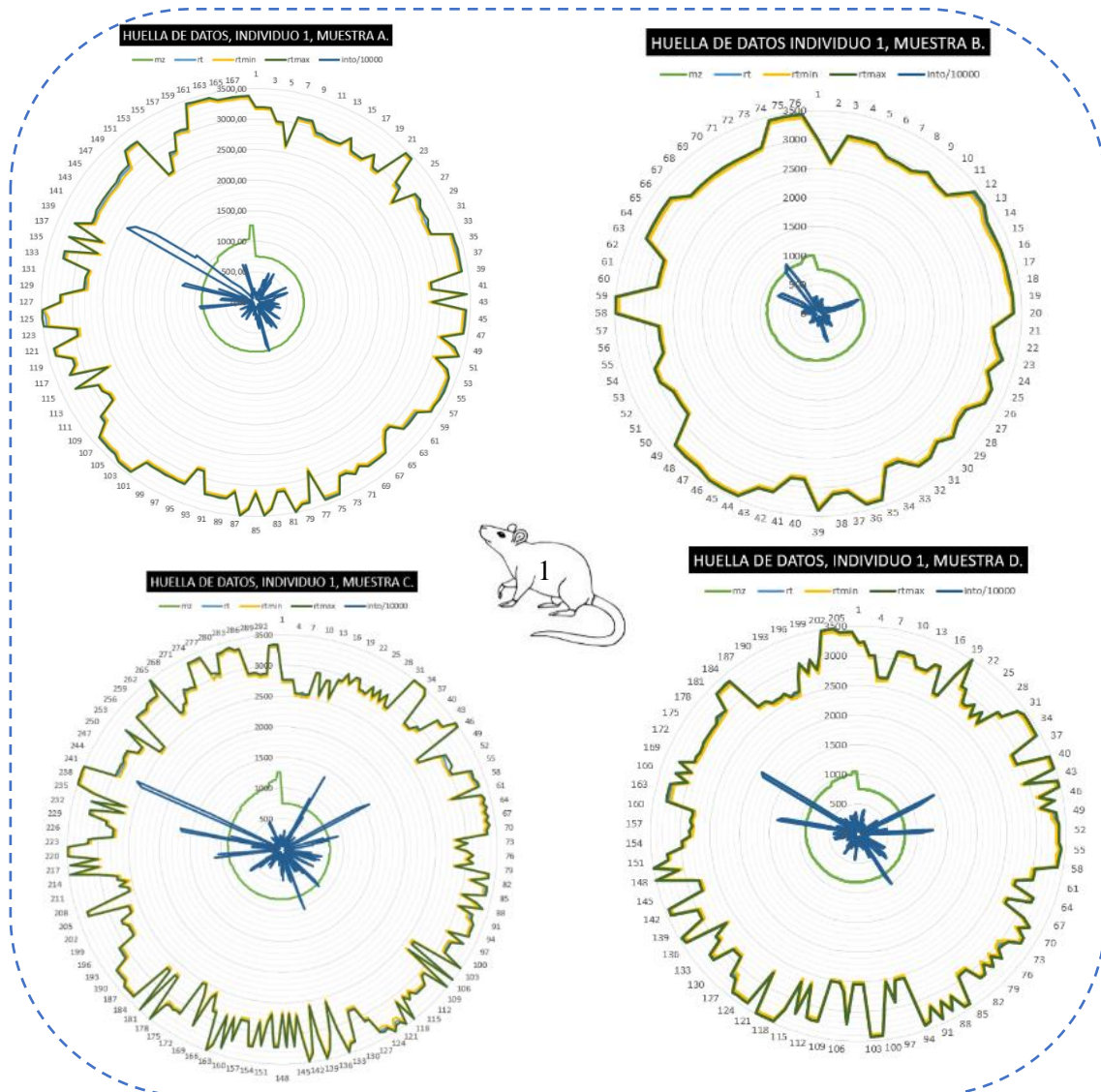


Gráfico 26. Huella de Datos del análisis estadístico de M/Z, tiempo de retención, tiempo de retención máximo y mínimo e intensidad por cada muestra, con un M/Z entre 750 y 1500. Muestra la relación espacial de los diferentes datos. Individuo 1.

En el siguiente gráfico podemos observar que, al igual que en gráfico anterior, la muestra B es la que menor variabilidad presenta, mientras que la muestra C es la que mayor variabilidad presenta en cuanto a los tiempos de retención. Se observa, en las cuatro muestras, ese característico pico en la intensidad en el cuadrante superior izquierdo, con diferentes intensidades en cada muestra. En la muestra A presenta una intensidad de 2000 cps, en la muestra B de aproximadamente 2000 cps, en la muestra C de aproximadamente 3000 cps y en la muestra D aproximadamente 2500 cps.

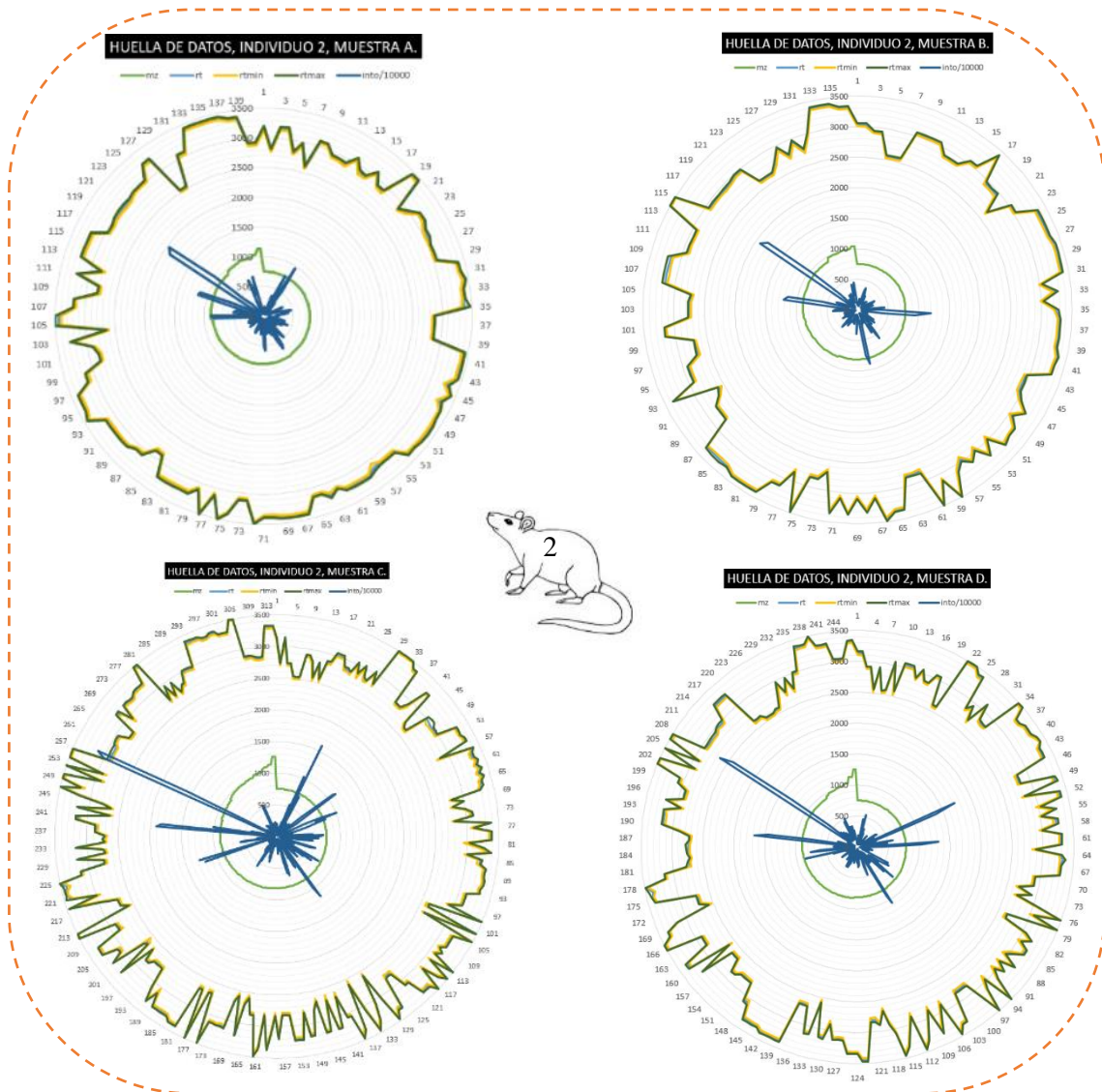


Gráfico 27. Huella de Datos del análisis estadístico de M/Z, tiempo de retención, tiempo de retención máximo y mínimo e intensidad por cada muestra, con un M/Z entre 750 y 1500. Muestra la relación espacial de los diferentes datos. Individuo 2.

En el siguiente gráfico observamos como los patrones, que hemos evidenciado en las 2 réplicas biológicas anteriores, se repiten nuevamente. Con ese pico de intensidad en el cuadrante superior izquierdo. Así, la muestra A presenta un pico con una intensidad de aproximadamente 2300 cps, la muestra B de, también, 2300 cps, la muestra C de 2400 cps y la muestra D de 2300 cps, lo que hace que las 4 muestras de este individuo de estudio sean las que menos variabilidad entre muestras presenten los picos de mayor intensidad que se hallan, en todos los casos, en el cuadrante superior izquierdo. Lo que llama la atención, es que en este caso, es la muestra A la que presenta una menor variabilidad en cuanto a los tiempos de retención.

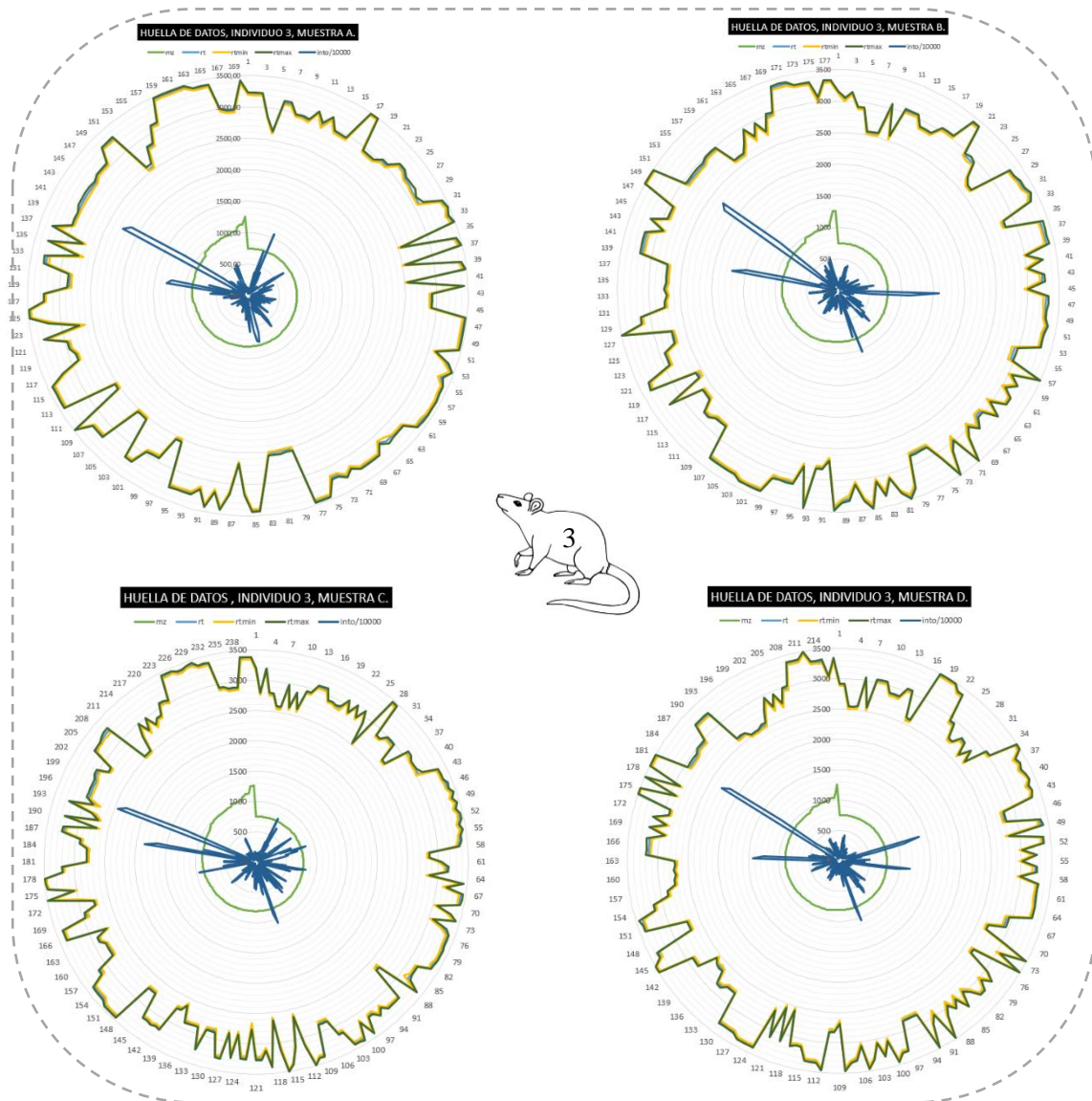


Gráfico 28. Huella de Datos del análisis estadístico de M/Z, tiempo de retención, tiempo de retención máximo y mínimo e intensidad por cada muestra, con un M/Z entre 750 y 1500. Muestra la relación espacial de los diferentes datos. Individuo 3.

- **Representación gráfica de la relación entre tiempo de retención con intensidad y M/Z, con su media, desviación estándar y la densidad del conteo en los dos ejes: X y Y.**

A continuación podremos observar los resultados de los dos últimos experimentos. Estos se han realizado utilizando los mismos datos que para el experimento anterior. Se ha usado el

paquete “jamovi” (62), para graficar la información de la MS. La diferencia es que se ha buscado darles un valor más importante a otros aspectos de esta. En este caso, se ha graficado los datos de la intensidad en relación con los tiempos de retención y de la M/Z en relación con los tiempos de retención. Se ha graficado la línea de la media de los datos suavizados y se ha presentado la desviación estándar que presenta esa media. Además, se ha graficado en la parte superior y en la parte derecha del gráfico, una curva de densidad de conteos tanto para el eje de las X: intensidad o M/Z, correspondientemente, y de las Y: tiempo de retención.

Esta es una manera diferente de analizar los datos de una MS. Nos da información que de otra forma no podríamos obtener. Nos indica, para cada eje, en que valores son donde más se condensan los conteos. Además, nos indica en que zonas existe una mayor desviación estándar, lo que nos indicaría una mayor variabilidad en los valores, o viceversa.

De esta manera, podemos ver en el siguiente gráfico los resultados de este experimento de las muestras de la réplica biológica del individuo 1. En este gráfico veremos en el eje de las X: tiempo de retención y en el eje de las Y: intensidad, medida en cps. Observamos como en la muestra A, al inicio de la curva de densidad de conteos del tiempo de retención, presenta una escasa densidad de conteos, hecho que se repite, aunque no de manera tan marcada, en las otras 3 muestras. Se puede observar también como, en las curvas de densidad de conteos de la intensidad, en todas las muestras, presenta una densidad de conteos marcada en las zonas de intensidad más bajas, siendo escasas en las zonas de altas intensidades. Esto es muy interesante, ya que muchos de los gráficos que se han realizado anteriormente mostraban la densidad de conteos, pero no mostraban en que zona de su medición presentaban ese acúmulo de conteos. De esta forma podemos observar con mayor claridad, donde se encontraban los acúmulos de densidad de conteos para cada caso en este individuo.



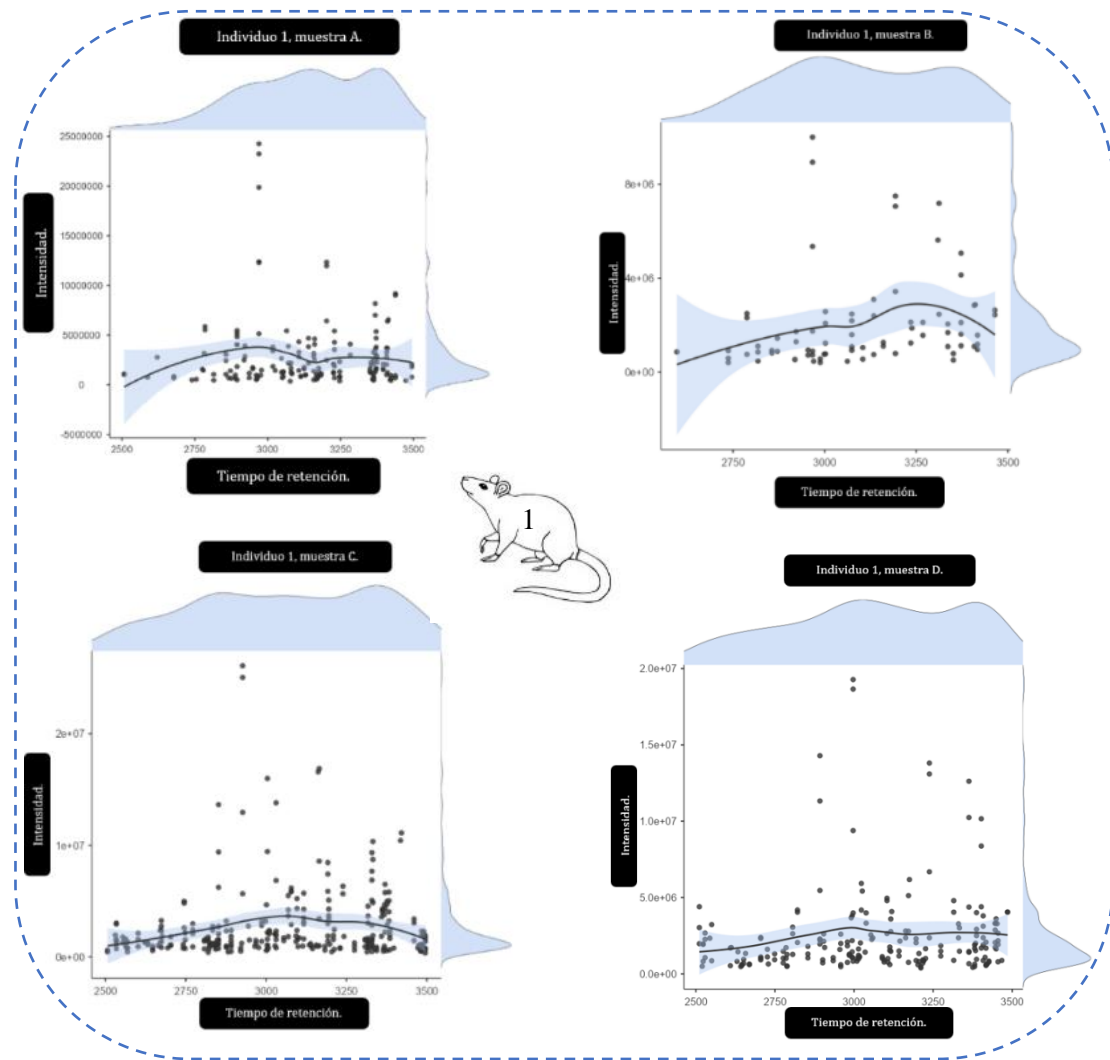


Gráfico 29. Representación gráfica de la relación entre tiempo de retención con intensidad, con su media, desviación estándar y la densidad del conteo en los dos ejes: X y Y. Individuo 1.

En el siguiente gráfico podemos observar los resultados de este experimento de las muestras de la réplica biológica del individuo 1. En este gráfico veremos en el eje de las X: tiempo de retención y en el eje de las Y: los valores de M/Z. Al igual que en el gráfico anterior, podemos observar que la curva de densidad de conteos de tiempos de retención presenta un número más escaso de conteos al inicio de la curva. Podemos observar también como la curva de densidad de conteos de M/Z presenta una mayor cantidad de conteos en los valores más bajos de este. Algo particular que llama la atención, es que en la muestra B se presenta, a diferencia de las demás muestras, una línea media que termina en un valor de M/Z muy superior que el resto coincidente con una segunda joroba de acumulo de densidad de conteos de M/Z. Algo que cabe recalcar, es que la muestra D es la que presenta una mayor dispersión en los conteos, a diferencia de las demás muestras, donde se tiende a acumular los conteos en los valores más bajos de M/Z.

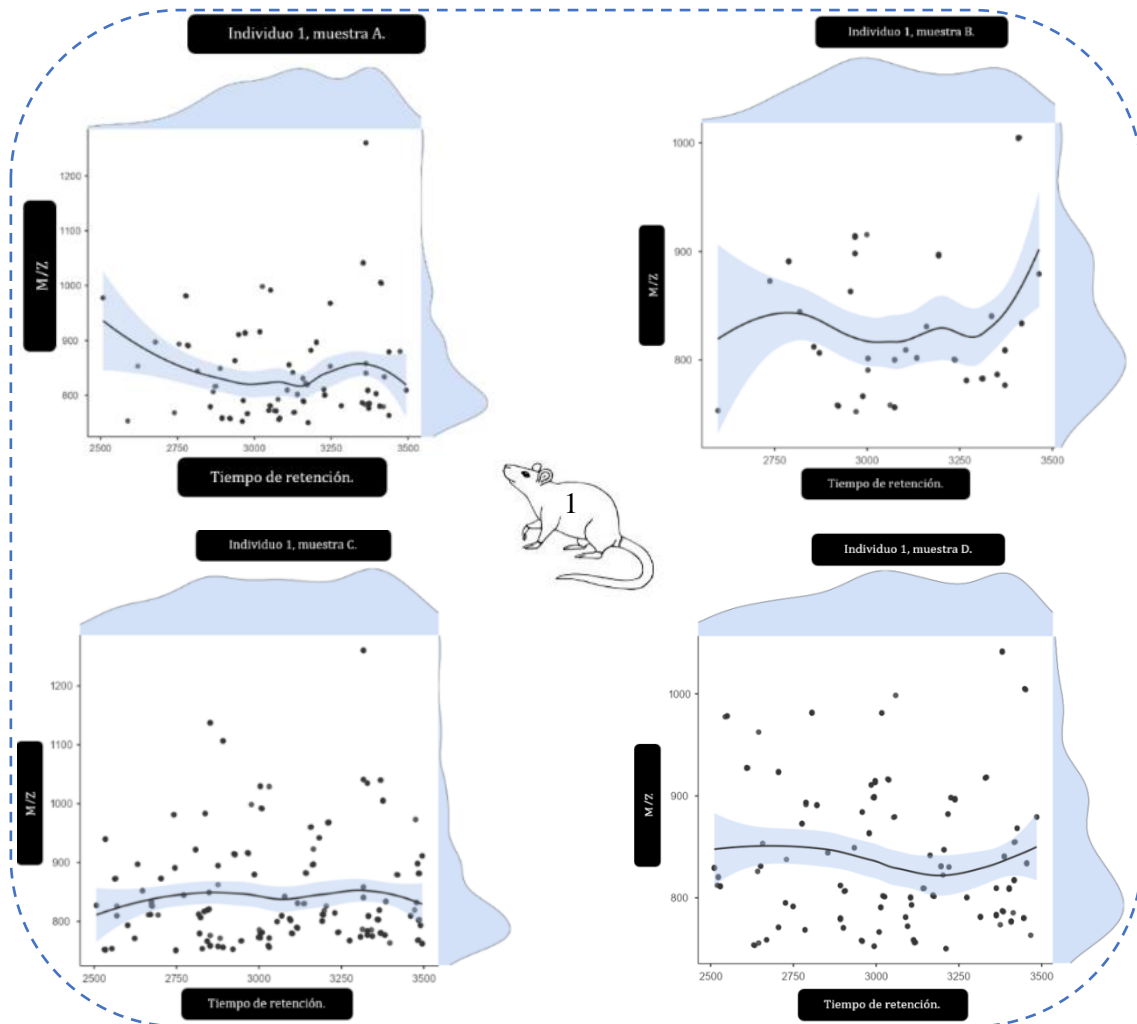
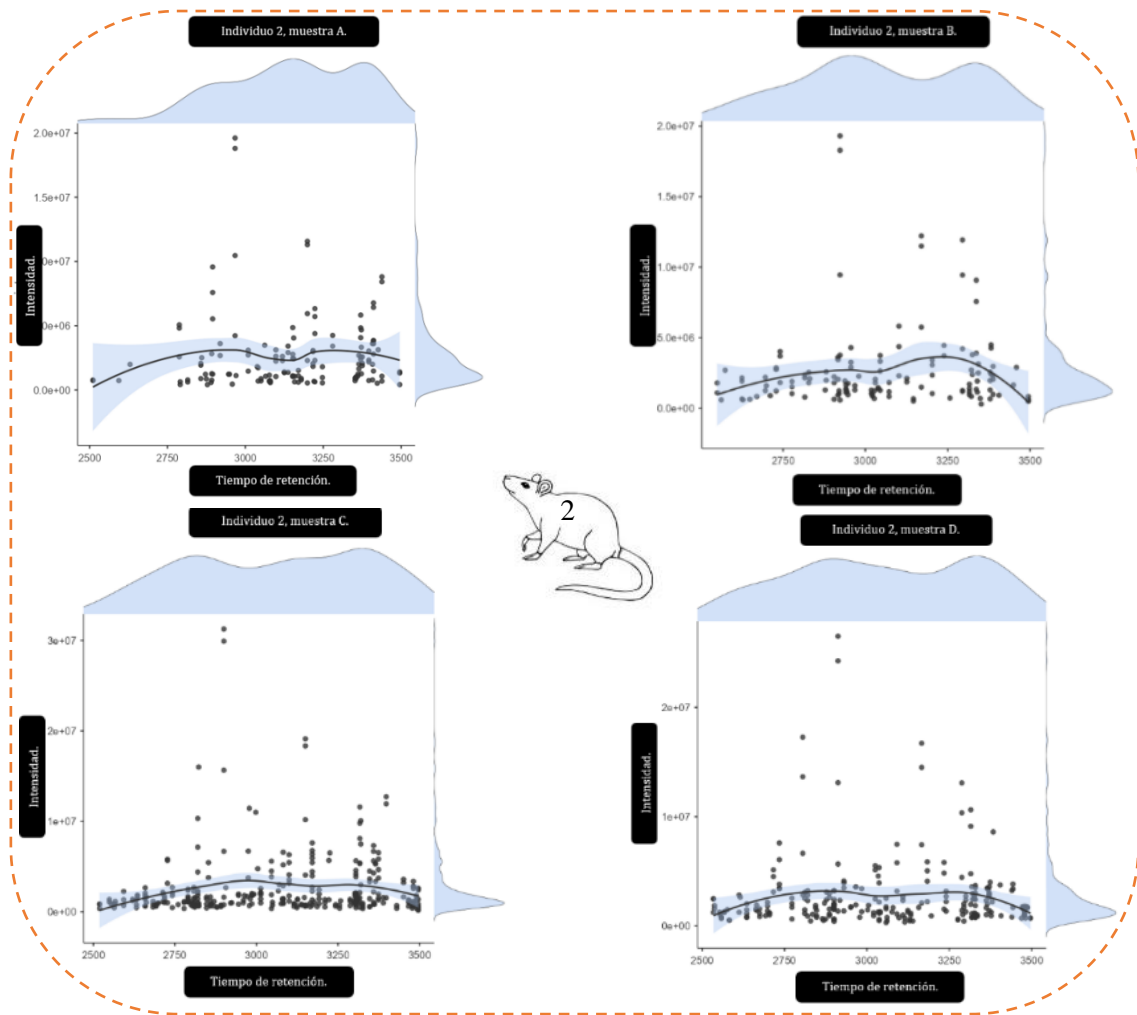


Gráfico 30. Representación gráfica de la relación entre tiempo de retención con M/Z, con su media, desviación estándar y la densidad del conteo en los dos ejes: X y Y. Individuo 1.

En el siguiente gráfico evidenciamos los resultados de este experimento de las muestras de la réplica biológica del individuo 2. En este gráfico veremos en el eje de las X: tiempo de retención y en el eje de las Y: los valores de intensidad, medidos en cps. Aquí podemos observar algunos patrones que se repiten en los resultados de la réplica biológica del individuo anterior. La poca cantidad de conteos al inicio de la curva de densidad de conteos en el tiempo de retención, la acumulación de conteos en la curva de densidad de conteos en los valores más bajos de la intensidad, unos valores de medias que se hallan siempre en valores inferiores de la intensidad. Es interesante observar el comportamiento de la media de las muestras C y D, las dos muestran un mayor número total de conteos, y las dos presentan una media en zonas más bajas de los valores de la intensidad con una desviación estándar más estrecha que las otras dos muestras.



*Gráfico 31. Representación gráfica de la relación entre tiempo de retención con intensidad, con su media, desviación estándar y la densidad del conteo en los dos ejes: X y Y. Individuo 2.*

En el siguiente gráfico evidenciamos los resultados de este experimento de las muestras de la réplica biológica del individuo 2. En este gráfico veremos en el eje de las X: tiempo de retención y en el eje de las Y: los valores de M/Z. Podemos ver que las medias en todos los momentos del tiempo de retención de todas las muestras se encuentran alrededor de los valores de M/Z entre 800 y 900. Se evidencia que a ante una mayor cantidad total de conteos más “recta” es la media y menor es su desviación estándar.

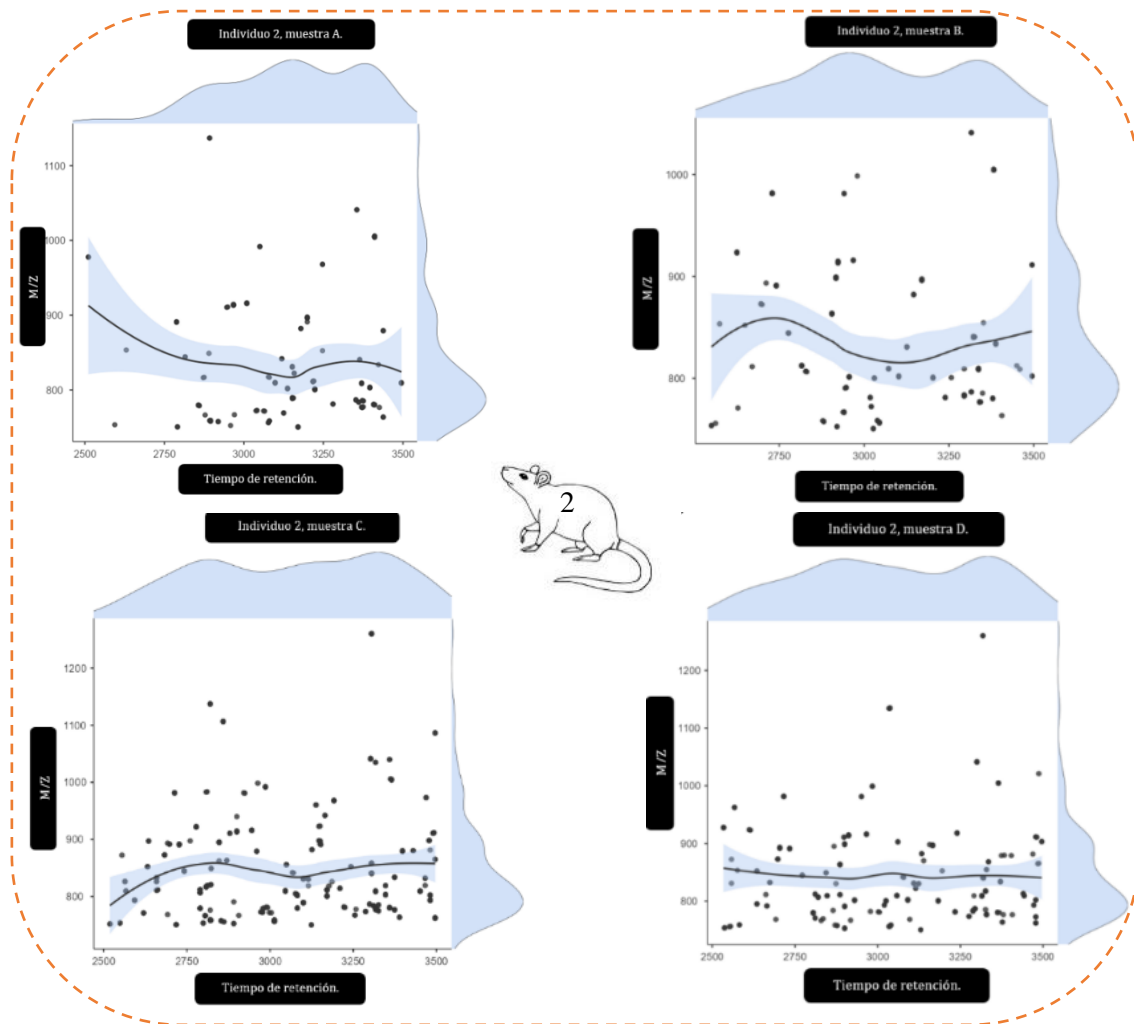
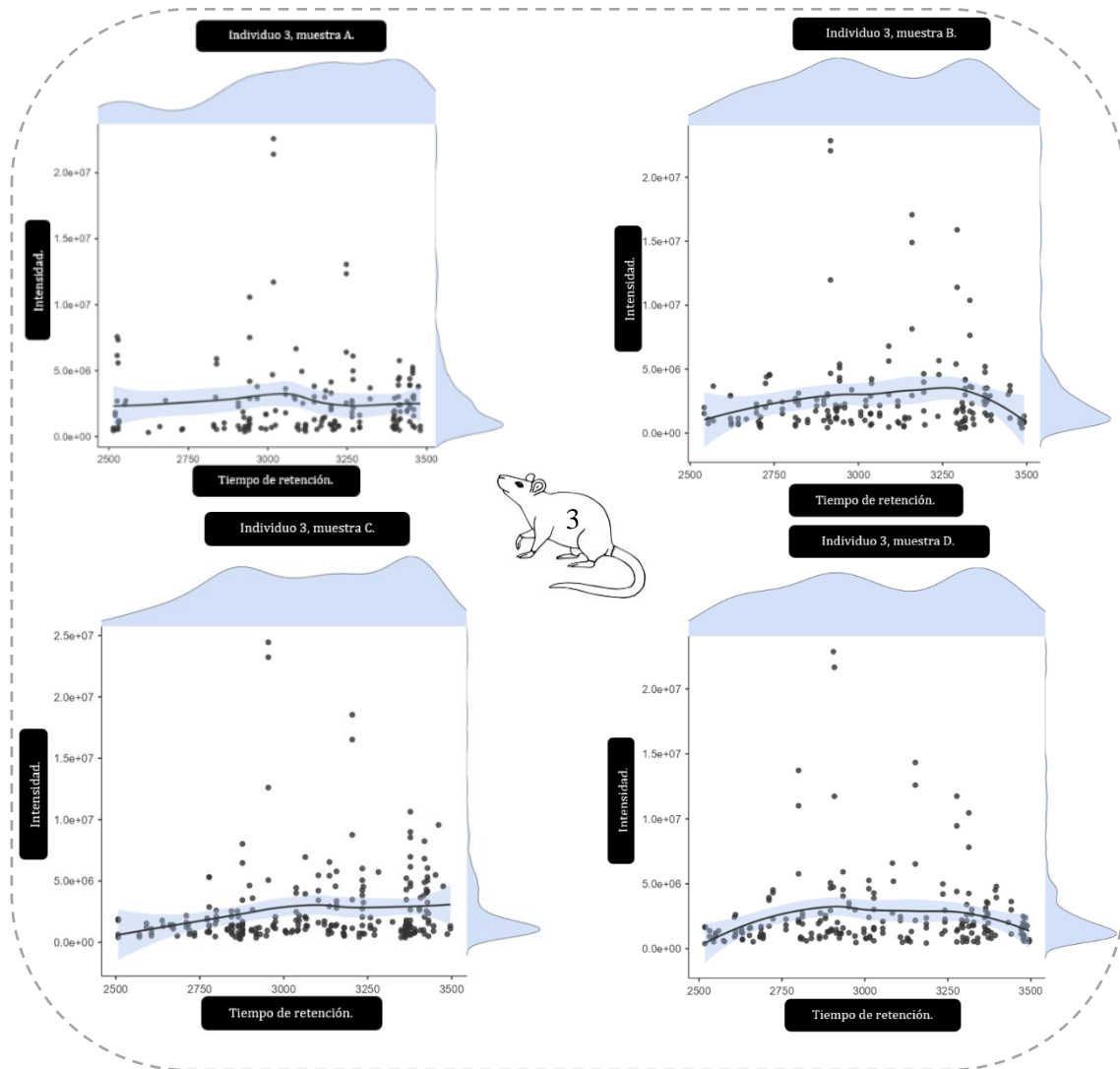


Gráfico 32. Representación gráfica de la relación entre tiempo de retención con M/Z, con su media, desviación estándar y la densidad del conteo en los dos ejes: X y Y. Individuo 2.

En el siguiente gráfico evidenciamos los resultados de este experimento de las muestras de la réplica biológica del individuo 3. En este gráfico veremos en el eje de las X: tiempo de retención y en el eje de las Y: los valores de intensidad, medido en cps. Se puede observar en el siguiente gráfico como la mayoría de las mediciones a través del tiempo de retención se acumula en las intensidades más bajas, a pesar de tener mediciones altas alrededor de los 3000 segundos. Esto se puede apreciar mejor en la curva de densidad de mediciones.



*Gráfico 33. Representación gráfica de la relación entre tiempo de retención con intensidad, con su media, desviación estándar y la densidad del conteo en los dos ejes: X y Y. Individuo 3.*

En el siguiente gráfico evidenciamos los resultados de este experimento de las muestras de la réplica biológica del individuo 3. En este gráfico veremos en el eje de las X: tiempo de retención y en el eje de las Y: los valores de M/Z. Así, podemos observar que los principales patrones vistos anteriormente en los otros dos individuos de estudio se repiten en este individuo. Es importante tener en cuenta que la media en todas las muestras de este individuo, se presentan alrededor de los 800 M/Z, que el número de conteos de la densidad de conteos de M/Z se acumula preferentemente alrededor de los 750 M/Z, lo que tiene sentido ya que en todas las muestras de este individuo tenemos al menos una medición sobre los 1200 M/Z. Es muy importante tener en cuenta que a diferencia del clásico cromatograma, este tipo de gráfico presenta información complementaria que no podemos observar en un cromatograma clásico.

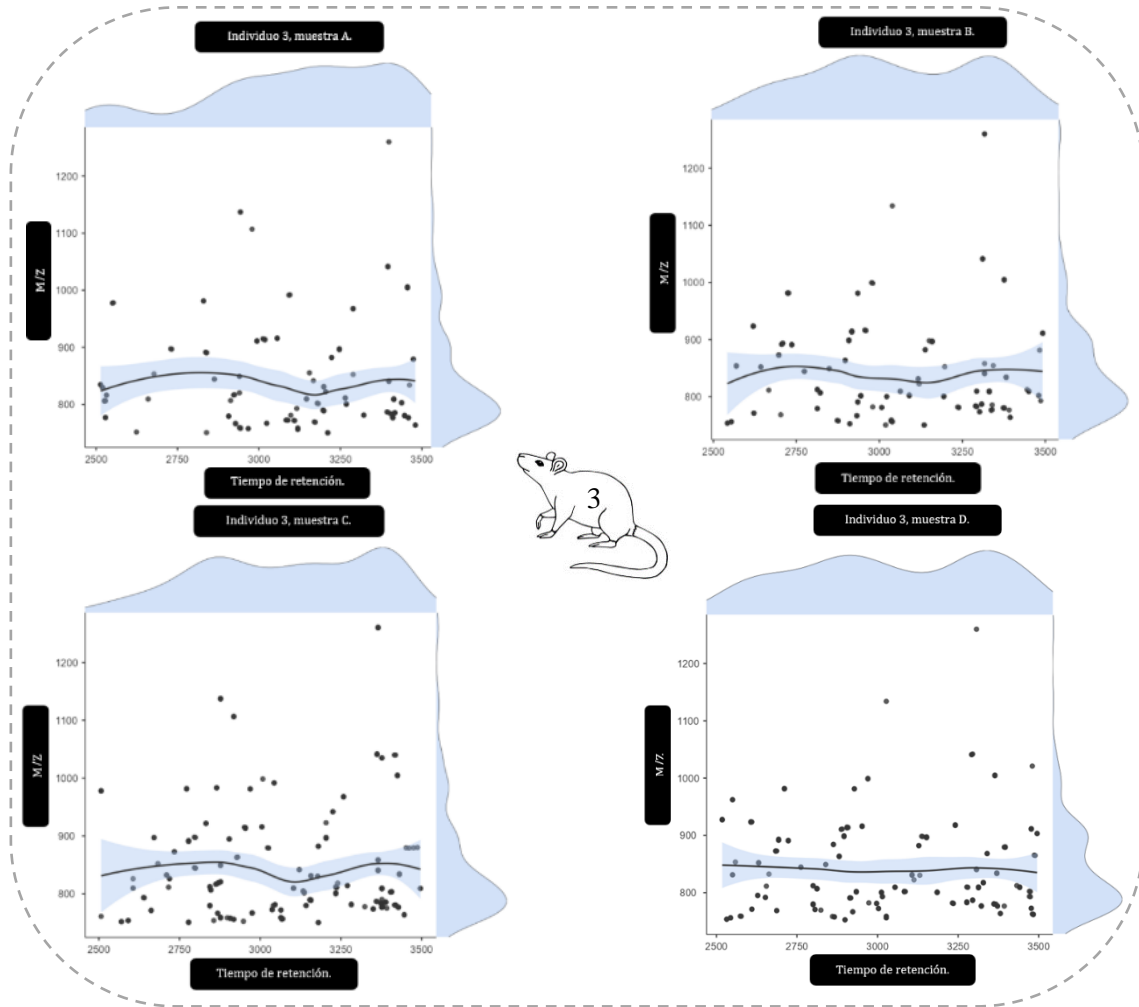


Gráfico 34. Representación gráfica de la relación entre tiempo de retención con M/Z, con su media, desviación estándar y la densidad del conteo en los dos ejes: X y Y. Individuo 3.

## **Discusión.**

### **- Réplicas biológicas.**

Dentro del campo de la investigación biomédica, un aspecto clave es poder entender la diferencia entre las réplicas biológicas y las réplicas técnicas. Para entender este concepto debemos basarnos en que, por lo general, cuando usamos estos términos estamos hablando de algún tipo de estudio que incluye tejidos biológicos. Así, debemos definir a las réplicas biológica como mediciones de muestras biológicamente distintas que muestran variación biológica (13,17), por el contrario, las réplicas técnicas son mediciones repetidas de la misma muestra que muestra que son independientes del ruido asociado con el equipo y los protocolos. (13)

Una de las preguntas que se presenta ante los resultados obtenidos es, ¿por qué hay tanta variabilidad entre los resultados de la MS en las réplicas biológicas de cerebro? La respuesta es mucho más compleja de lo que se puede pensar. De hecho, es un debate abierto hoy en día. (68) En un breve resumen de la evidencia actual sobre las réplicas biológicas, las réplicas técnicas y las herramientas de análisis. La solución que más firmemente se plantea como respuesta es, como sería obvio de esperar, realizar una mezcla, es decir, analizar al menos 3 réplicas biológicas, analizar todas las muestras, al menos una vez más cada una, creando réplicas técnicas y utilizando herramientas cada vez más precisas, tanto para analizar las muestras como para procesar y analizar la información. (15,16,68,69)

La variabilidad de los resultados de puede deberse a que se están analizando réplicas biológicas, de 3 individuos diferentes. Aunque se ha tratado de guardar lo más exhaustivamente posible la similitud entre réplicas, es muy difícil que ante un proceso tan exacto como la MS, los resultados sea iguales. Esto, debido a que incluso si el órgano afectado fuese un clon, los cortes histológicos, la preparación, los cambios ambientales, entre demás variables, pueden hacer que la información proporcionada por la MS sea diferente entre individuos.

Es importante comprender que, más allá de que los datos obtenidos sean diferentes, lo que importa es que guarden los patrones cromatográficos. Esto, debido a que es la forma en la que se analiza una muestra en busca de su composición, como se explicará más adelante. Lo realmente crítico es que no se guarden patrones básicos, y como hemos evidenciado en los resultados del estudio que se ha realizado, la gran mayoría de muestras guarda un patrón morfológico en la cromatografía. Sin embargo, a pesar de que las muestras han sido preparadas siguiendo correctamente los procedimientos y protocolos de preparación, existen algunas muestras que rompen el patrón. No se ha logrado identificar un patrón común para la disociación de patrones que presentan los resultados cromatográficos.

- **Patrones dentro de los datos masivos.**

Es muy importante entender la forma en la que los softwares de cotejamiento para cromatografías pueden obtener los elementos, en este caso proteínas, que conforman los resultados de la MS. Lo que hacen es analizar los datos puros de la matriz de resultados, algo que resulta contraintuitivo y complejo para la forma de ver los datos para el cerebro del ser humano. Una forma que tenemos de entender esos datos masivos es presentarlos en forma de un gráfico que nuestro cerebro logre interpretar, en nuestro caso es la cromatografía de la MS. Cuando un software que analiza los datos en busca de similitudes, lo que hace es aproximar los datos encontrados, exponiendo la probabilidad de certeza, para interpretar ese patrón existente con un patrón ya estudiado de su base de datos. Esta forma de crear los resultados hace que muchas veces los datos obtenidos no sean óptimos, que los resultados obtenidos no presenten una certeza alta o que exista falsas mediciones, los denominados falsos positivos. Este campo del análisis de datos es uno de los que mayores saltos ha dado en las últimas décadas, haciendo que este proceso sea cada vez sea más exacto. (70–73)

Como se ha explicado anteriormente, analizar los datos en busca de coincidencias para las proteínas no siempre es confiable y no siempre está exento de errores, errores que son difíciles de corregir. Para corregir estos errores existen varias estrategias, por ejemplo, al momento de preparar la muestra se puede usar un compuesto específico que marque proteínas de interés. Esto, aunque muy eficaz, tiene sus desventajas, ya que se distorsiona la información del resto de elementos no marcados.

Lo que propone el presente trabajo, en cuanto a este particular, es buscar nuevas herramientas de análisis, encontrar formas de cotejar los patrones con la base de datos en busca de similitudes usando otros métodos de análisis, o incluso, medir los datos de una manera diferente, ya que como se ha visto en el apartado de resultados, existen alternativas para conocer y analizar los resultados de la MS.

Es importante entender, que al tener varios métodos de preparación, además, con distintos individuos de estudio para cada réplica biológica, es un riesgo conocido el que los resultados de la búsqueda de coincidencias en el análisis de cotejamiento puedan presentar resultados distintos y no siempre verdaderos.

- **Nuevas maneras de representar los datos masivos.**

Es importante visualizar la información de un modo diverso, así, es más probable que encontremos soluciones que antes pudieron no habérsenos ocurrido y explicar eventos de una forma que sea más comprensible para todos, incluso, observar nuevas perspectivas de los datos antes no estudiadas. (74–76)



En el presente trabajo se ha analizado de múltiples maneras de información obtenida de los resultados de la MS de cada una de las muestras de las 3 réplicas biológicas. Se ha analizado cada gráfico en busca de patrones, o su ausencia. El campo del conocimiento es muy vasto, siempre se debe analizar por la vía “clásica” sin dejar de lado las contribuciones “de novo” que se pueda aportar como investigadores. Al igual que existe la técnica artística, existe la inspiración científica. Los datos masivos son de los elementos más complejos de entender, sobre todo porque nuestro cerebro no está hecho para entender, sin más, una tabla de datos. Es mucho más probable que analizar con mayor eficacia cuando observamos los patrones, o la ausencia de estos, en una imagen que es capaz de representar correctamente el significado de la información que representa.

## **Conclusiones.**

El estudio de la neuroproteómica es un vasto campo del conocimiento que está en constante desarrollo. El estudio del sistema más complejo que conocemos es, sin lugar a duda, un reto que merece la pena enfrentar. Las células que componen en cerebro, que ha sido el objeto de nuestro estudio, no se comportan siempre de la misma manera que las células del resto del cuerpo. (1,10,11) Por lo que la reacción de estas a diferentes protocolos de preparación, aún más observando los datos obtenidos en el presente estudio, es un paso que hay que tomar muy en cuenta el momento de analizar el cerebro. Sobre este aspecto se sabe muy poco todavía, y la poca información que se tiene no presenta una robustez suficiente, por el momento. El autor de este trabajo espera que los conocimientos adquiridos a través de este sencillo trabajo de investigación aporten su granito de arena en la investigación del órgano irremplazable. Sin embargo, son necesarios más estudios para corroborar los resultados obtenidos en el presente estudio.

Se ha evidenciado como el uso de diferentes experimentos con varios diseños son fundamentales para poder observar la información de los resultados de una manera más amplia y desde diferentes perspectivas. Se espera que la manera en la que se ha analizado la información sea útil para que, en el futuro, se puedan crear nuevos diseños de estudio que aporten un conocimiento más profundo sobre el tema en cuestión, así como para que sirva de inspiración en el análisis de otros temas importantes en la neuroproteómica y la bioinformática en general.

### **Agradecimiento.**

Se agradece a la Universidad Europea de Madrid y a todo el personal, tanto docente como administrativo, en verdad han hecho un gran trabajo, se nota el esfuerzo y dedicación que le han puesto a la creación, organización y ejecución de este Máster Oficial, se nota en la excelencia. En especial agradezco a María del Rocío González Soltero, en toda mi vida dentro de las aulas universitarias, tanto como estudiante así como docente, nunca he visto un director tan entregado y cercano a sus estudiantes, me llevo no solo admiración de la forma en la que hace su trabajo, sino también un ejemplo de cómo se debe llevar el liderazgo dentro de una institución. Agradezco a Verónica Moral Dardé por haberme guiado en cada paso, con calidez y claridad, en la realización de mi Trabajo de Fin de Máster. Agradezco al personal especializado de la Unidad de Proteómica del Hospital Nacional de Paraplégicos de Toledo (España), en especial a Gemma Borroso García y Alba González Arandilla, por haberme proporcionado la materia prima, los datos, para realizar mi investigación, cada vez que un investigador ayuda a otro hacemos que la investigación se democratice, estos nobles gestos hacen que la ciencia progrese. Agradezco a mi amadísima esposa y a mi pequeño motor, mi hijo, por haberme dado las fuerzas, así como su amor y paciencia durante todo este año. Agradezco a mis padres por su amor incondicional, por el ejemplo de sus vidas y la entereza de su apoyo. Y agradezco a Dios por darme la oportunidad de formarme; en mi corazón llevo gravado el mensaje, que a la voz de ese gran santo español pronunciaba, ese mensaje que guía mi vida y que espero que siempre lo siga haciendo: “ser más para servir mejor”.

Al cosmos y al cerebro se los estudiarán por siempre, son inabarcables.

**Bibliografía.**

1. Dieterich DC, Kreutz MR. Proteomics of the Synapse – A Quantitative Approach to Neuronal Plasticity. *Mol Cell Proteomics* [Internet]. 2016 feb 1 [citado 2022 sep 25];15(2):368. Available from: [/pmc/articles/PMC4739661/](#)
2. Watrous JD, Dorrestein PC. Imaging mass spectrometry in microbiology. *Nat Rev Microbiol* [Internet]. 2011 sep [citado 2022 sep 25];9(9):683. Available from: [/pmc/articles/PMC3710447/](#)
3. Unsihuay D, Mesa Sanchez D, Laskin J. Quantitative Mass Spectrometry Imaging of Biological Systems. <https://doi.org/10.1146/annurev-physchem-061020-053416> [Internet]. 2021 abr 20 [citado 2022 sep 25];72:307-29. Available from: <https://www.annualreviews.org/doi/abs/10.1146/annurev-physchem-061020-053416>
4. Sinha A, Mann M. A beginner's guide to mass spectrometry-based proteomics. *Biochem (Lond)* [Internet]. 2020 oct 21 [citado 2022 sep 25];42(5):64-9. Available from: [/biochemist/article/42/5/64/226371/A-beginner-s-guide-to-mass-spectrometry-based](#)
5. Chen C, Hou J, Tanner JJ, Cheng J. Bioinformatics Methods for Mass Spectrometry-Based Proteomics Data Analysis. *Int J Mol Sci* [Internet]. 2020 abr 2 [citado 2022 sep 26];21(8). Available from: [/pmc/articles/PMC7216093/](#)
6. Buchberger AR, DeLaney K, Johnson J, Li L. Mass Spectrometry Imaging: A Review of Emerging Advancements and Future Insights. *Anal Chem* [Internet]. 2018 ene 1 [citado 2022 sep 25];90(1):240. Available from: [/pmc/articles/PMC5959842/](#)
7. Nairn AC, Williams KR. Neuroproteomics. 2020 feb 12 [citado 2022 sep 26]; Available from: <https://www.mdpi.com/books/book/2022>
8. Alzate O. Neuroproteomics. *Neuroproteomics* [Internet]. 2010 ene 1 [citado 2022 sep 26];1-16. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK56022/>
9. Santa C, Anjo SI, Mendes VM, Manadas B. Neuroproteomics — LC-MS Quantitative Approaches. *Recent Advances in Proteomics Research* [Internet]. 2015 nov 11 [citado 2022 sep 26]; Available from: [undefined/state.item.id](#)
10. van Gelder CAGH, Altelaar M. Neuroproteomics of the Synapse: Subcellular Quantification of Protein Networks and Signaling Dynamics. *Mol Cell Proteomics* [Internet]. 2021 may 1 [citado 2022 sep 25];20:100087. Available from: [/pmc/articles/PMC8167277/](#)

11. Haenig C, Atias N, Taylor AK, Mazza A, Schaefer MH, Russ J, et al. Interactome Mapping Provides a Network of Neurodegenerative Disease Proteins and Uncovers Widespread Protein Aggregation in Affected Brains. *Cell Rep* [Internet]. 2020 ago 18 [citado 2022 sep 25];32(7):108050. Available from: <http://www.cell.com/article/S2211124720310354/fulltext>
12. Schneider C. Wie offen ist die Open Souce Software? TATuP - Zeitschrift für Technikfolgenabschätzung in Theorie und Praxis. 2016 nov 1;25(3):77-8.
13. Tsvetkov D, Kolpakov E, Kassmann M, Schubert R, Gollasch M. Distinguishing Between Biological and Technical Replicates in Hypertension Research on Isolated Arteries. *Front Med (Lausanne)* [Internet]. 2019 jun 20 [citado 2022 sep 27];6:126. Available from: </pmc/articles/PMC6595250/>
14. Bell G. Replicates and repeats. *BMC Biol* [Internet]. 2016 abr 7 [citado 2022 sep 27];14(1). Available from: </pmc/articles/PMC4825082/>
15. Goren E, Liu P, Wang C, Wang C. BinQuasi: a peak detection method for ChIP-sequencing data with biological replicates. *Bioinformatics* [Internet]. 2018 sep 1 [citado 2022 sep 27];34(17):2909-17. Available from: <https://academic.oup.com/bioinformatics/article/34/17/2909/4978050>
16. Squair JW, Gautier M, Kathe C, Anderson MA, James ND, Hutson TH, et al. Confronting false discoveries in single-cell differential expression. *Nat Commun* [Internet]. 2021 dic 1 [citado 2022 sep 27];12(1). Available from: </pmc/articles/PMC8479118/>
17. Blainey P, Krzywinski M, Altman N. Points of significance: Replication. *Nat Methods*. 2014 sep 1;11(9):879-80.
18. Thiourea/Urea Lysis Buffer | Nectagen [Internet]. [citado 2022 sep 27]. Available from: <https://www.nectagen.com/reference-data/solutions/thiourea-urea-lysis-buffer>
19. Mellick AS, Rodgers L. *A Handbook of Recipes, Reagents, and Other Reference Tools for Use at the Bench*. Vol.2. 2007;248.
20. Arruda SCC, Barbosa HDS, Azevedo RA, Arruda MAZ. Two-dimensional difference gel electrophoresis applied for analytical proteomics: Fundamentals and applications to the study of plant proteomics. *Analyst*. 2011 oct 21;136(20):4119-26.
21. RIPA Lysis Buffer. *Cold Spring Harb Protoc* [Internet]. 2017 dic [citado 2022 sep 27];2017(12):undefined-undefined. Available from: <https://www.mendeley.com/catalogue/b2f0f9d4-5266-350c-98e1-6dbf77e3cdbd/>

22. Shevchenko A, Tomas H, Havliš J, Olsen J v., Mann M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc* [Internet]. 2007 ene [citado 2022 sep 27];1(6):2856-60. Available from: [https://www.researchgate.net/publication/232803487\\_Shevchenko\\_A\\_Tomas\\_H\\_Havlis\\_J\\_Olsen\\_JV\\_Mann\\_M\\_In-gel\\_digestion\\_for\\_mass\\_spectrometric\\_characterization\\_of\\_proteins\\_and\\_proteomes\\_Nat\\_Protoc\\_1\\_2856-2860](https://www.researchgate.net/publication/232803487_Shevchenko_A_Tomas_H_Havlis_J_Olsen_JV_Mann_M_In-gel_digestion_for_mass_spectrometric_characterization_of_proteins_and_proteomes_Nat_Protoc_1_2856-2860)
23. Shevchenko A, Wilm M, Vorm O, Mann M. Mass Spectrometric Sequencing of Proteins from Silver-Stained Polyacrylamide Gels. *Anal Chem* [Internet]. 1996 [citado 2022 sep 27];68(5):850-8. Available from: <https://pubs.acs.org/doi/abs/10.1021/ac950914h>
24. Gao M, Thielhorn R, Rentsch J, Honigmann A, Ewers H. Expansion STED microscopy (ExSTED). *Methods Cell Biol.* 2021 ene 1;161:15-31.
25. Hughes CS, Moggridge S, Müller T, Sorensen PH, Morin GB, Krijgsveld J. Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nat Protoc.* 2019 ene 1;14(1):68-85.
26. Sielaff M, Kuharev J, Bohn T, Hahlbrock J, Bopp T, Tenzer S, et al. Evaluation of FASP, SP3, and iST Protocols for Proteomic Sample Preparation in the Low Microgram Range. *J Proteome Res* [Internet]. 2017 nov 3 [citado 2022 sep 27];16(11):4060-72. Available from: <https://www.mendeley.com/catalogue/5e650f47-5229-388f-8d24-5d81f7030061/>
27. Mikulášek K, Konečná H, Potěšil D, Holánková R, Havliš J, Zdráhal Z. SP3 Protocol for Proteomic Plant Sample Preparation Prior LC-MS/MS. *Front Plant Sci.* 2021 mar 10;12:369.
28. Ihaka R, Gentleman R. R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics.* 1996;5(3):299-314.
29. Chambers JM. S, R, and data science. *Proceedings of the ACM on Programming Languages.* 2020 jun 12;4(HOPL).
30. Team undefined Rs. RStudio | Open source & professional software for data science teams - RStudio. 2020 [citado 2022 sep 27];undefined-undefined. Available from: <https://www.mendeley.com/catalogue/238bfb92-c8e2-3982-84d1-14d0c591e31c/>
31. Team undefined Rs. Citing RStudio – RStudio Support. 2020 [citado 2022 sep 27];undefined-undefined. Available from: <https://www.mendeley.com/catalogue/0d720a28-670f-3938-945a-876efbaec114/>

32. project undefined T jamovi. jamovi (Version 2.2) [Computer Software]. 2021 [citado 2022 sep 27];undefined-undefined. Available from: <https://www.mendeley.com/catalogue/c7ad085e-d63b-37cb-a4ff-51cd4487c0fb/>
33. Annie goleman daniel; boyatzis, RM, Perdana undefined. Jamovi. 2018 [citado 2022 sep 27];53(9):1689-99. Available from: <https://www.mendeley.com/catalogue/262787d0-d297-3d33-bc35-08c9bd960191/>
34. Ma H, Li Y, Zhang H, Li Q, Chen F, Cui Z. A virtual laboratory based on full-field crystal plasticity simulation to characterize the multiscale mechanical properties of AHSS. *Sci Rep*. 2022 dic 1;12(1).
35. Guo L, Vengalil M, Abdul NMM, Wang K. Design and implementation of virtual laboratory for a microgrid with renewable energy sources. *Computer Applications in Engineering Education*. 2022 mar 1;30(2):349-61.
36. Varsos C, Patkos T, Oulas A, Pavloudi C, Gougousis A, Ijaz UZ, et al. Optimized R functions for analysis of ecological community data using the R virtual laboratory (RvLab). *Biodivers Data J*. 2016 nov 1;4(1).
37. Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, et al. Orchestrating high-throughput genomic analysis with Bioconductor. *Nature Methods* 2015 12:2 [Internet]. 2015 ene 29 [citado 2022 sep 27];12(2):115-21. Available from: <https://www.nature.com/articles/nmeth.3252>
38. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, et al. Software for computing and annotating genomic ranges. *PLoS Comput Biol*. 2013;9(8):e1003118.
39. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology* 2004 5:10 [Internet]. 2004 sep 15 [citado 2022 sep 27];5(10):1-16. Available from: <https://genomebiology.biomedcentral.com/articles/10.1186/gb-2004-5-10-r80>
40. Gatto L, Lilley KS. Msnbase-an R/Bioconductor package for isobaric tagged mass spectrometry data visualization, processing and quantitation. *Bioinformatics*. 2012 ene;28(2):288-9.
41. Gatto L, Gibb S, Rainer J. MSnbase, Efficient and Elegant R-Based Processing and Visualization of Raw Mass Spectrometry Data. *J Proteome Res* [Internet]. 2021 ene 1 [citado 2022 sep 27];20(1):1063-9. Available from: <https://pubmed.ncbi.nlm.nih.gov/32902283/>

42. Neumann S RLg. Bioconductor - msdata: Various Mass Spectrometry raw data example files. R package version 0.36.0. [Internet]. 2022 [citado 2022 sep 27]. Available from: <https://bioconductor.org/packages/release/data/experiment/html/msdata.html>
43. Kessner D, Chambers M, Burke R, Agus D, Mallick P. ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics* [Internet]. 2008 nov 1 [citado 2022 sep 27];24(21):2534-6. Available from: <https://academic.oup.com/bioinformatics/article/24/21/2534/190743>
44. Pedrioli PGA, Eng JK, Hubley R, Vogelzang M, Deutsch EW, Raught B, et al. A common open representation of mass spectrometry data and its application to proteomics research. *Nature Biotechnology* 2004 22:11 [Internet]. 2004 nov 4 [citado 2022 sep 27];22(11):1459-66. Available from: <https://www.nature.com/articles/nbt1031>
45. Martens L, Chambers M, Sturm M, Kessner D, Levander F, Shofstahl J, et al. mzML - A community standard for mass spectrometry data. *Molecular and Cellular Proteomics* [Internet]. 2011 ene 1 [citado 2022 sep 27];10(1):R110.000133. Available from: <http://www.mcponline.org/article/S1535947620313876/fulltext>
46. Chambers MC, MacLean B, Burke R, Amodei D, Ruderman DL, Neumann S, et al. A cross-platform toolkit for mass spectrometry and proteomics. *Nature Biotechnology* 2012 30:10 [Internet]. 2012 oct 10 [citado 2022 sep 27];30(10):918-20. Available from: <https://www.nature.com/articles/nbt.2377>
47. Eddelbuettel D, Balamuta JJ. Extending R with C++: A Brief Introduction to Rcpp. <https://doi.org/10.1080/0003130520171375990> [Internet]. 2018 ene 2 [citado 2022 sep 27];72(1):28-36. Available from: <https://www.tandfonline.com/doi/abs/10.1080/00031305.2017.1375990>
48. Eddelbuettel D. Seamless R and C++ integration with Rcpp. *Seamless R and C++ Integration with Rcpp*. 2013 ene 1;1-220.
49. Eddelbuettel D, François R. Rcpp: Seamless R and C++ Integration. *J Stat Softw* [Internet]. 2011 abr 13 [citado 2022 sep 27];40(8):1-18. Available from: <https://www.jstatsoft.org/index.php/jss/article/view/v040i08>
50. Deepayan Sarkar, Felix Andrews, LastNameKevin Wright, Neil Klepeis, Paul Murrell. Trellis Graphics for R [R package lattice version 0.20-45] [Internet]. cran.r-project. Comprehensive R Archive Network (CRAN); 2021 [citado 2022 sep 27]. Available from: <https://CRAN.R-project.org/package=lattice>



51. Baptiste Auguie, Anton Antonov. Miscellaneous Functions for «Grid» Graphics [R package gridExtra version 2.3] [Internet]. Comprehensive R Archive Network (CRAN); 2017 [citado 2022 sep 27]. Available from: <https://CRAN.R-project.org/package=gridExtra>
52. Huang T, Choi M, Tzouros M, Golling S, Pandya NJ, Banfai B, et al. MSstatsTMT: Statistical Detection of Differentially Abundant Proteins in Experiments with Isobaric Labeling and Multiple Mixtures. *Mol Cell Proteomics* [Internet]. 2020 oct 1 [citado 2022 sep 27];19(10):1706-23. Available from: <https://pubmed.ncbi.nlm.nih.gov/32680918/>
53. Rainer J, Vicini A, Salzer L, Stanstrup J, Badia JM, Neumann S, et al. A Modular and Expandable Ecosystem for Metabolomics Data Annotation in R. *Metabolites* 2022, Vol 12, Page 173 [Internet]. 2022 feb 11 [citado 2022 sep 27];12(2):173. Available from: <https://www.mdpi.com/2218-1989/12/2/173/htm>
54. LastNameGatto L (2022). rpx: R Interface to the ProteomeXchange Repository. R package version 2.4.1 <https://github.com/lgatto/rpx>. Bioconductor - rpx [Internet]. bioconductor.org; 2022 [citado 2022 sep 27]. Available from: <https://bioconductor.org/packages/release/bioc/html/rpx.html>
55. Benton HP, Want EJ, Ebbels TMD. Correction of mass calibration gaps in liquid chromatography-mass spectrometry metabolomics data. *Bioinformatics* [Internet]. 2010 jul 29 [citado 2022 sep 27];26(19):2488-9. Available from: <https://pubmed.ncbi.nlm.nih.gov/20671148/>
56. Tautenhahn R, Bottcher C, Neumann S. Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinformatics* [Internet]. 2008 nov 28 [citado 2022 sep 27];9(1):1-16. Available from: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-9-504>
57. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem* [Internet]. 2006 feb 1 [citado 2022 sep 27];78(3):779-87. Available from: <https://pubs.acs.org/doi/abs/10.1021/ac051437y>
58. Neuwirth E. ColorBrewer Palettes [R package RColorBrewer version 1.1-3] [Internet]. Comprehensive R Archive Network (CRAN); 2022 [citado 2022 sep 27]. Available from: <https://CRAN.R-project.org/package=RColorBrewer>

59. Gergely Daróczi, Roman Tsegelskyi. An R «Pandoc» Writer [R package pandoc version 0.6.5] [Internet]. Comprehensive R Archive Network (CRAN); 2022 [citado 2022 sep 27]. Available from: <https://CRAN.R-project.org/package=pandoc>
60. Stefan Milton Bache, Hadley Wickham, Lionel Henry. A Forward-Pipe Operator for R [R package magrittr version 2.0.3]. 2022 mar 30 [citado 2022 sep 27]; Available from: <https://CRAN.R-project.org/package=magrittr>
61. Perry Malcolm. Heatmaps: Flexible Heatmaps for Functional Genomics and Sequence Features. R package version 1.20.0. [Internet]. 2022 [citado 2022 sep 27]. Available from: <https://bioconductor.org/packages/release/bioc/html/heatmaps.html>
62. Ravi Selker, Jonathon Love, Damian Dropmann, Victor Moreno. CRAN - Package jmv [Internet]. 2022 [citado 2022 sep 29]. Available from: <https://cran.r-project.org/web/packages/jmv/index.html>
63. Morgan Martin, Obenchain Valerie, Hester Jim, Pagès Hervé. SummarizedExperiment: SummarizedExperiment container. R package version 1.26.1, <https://bioconductor.org/packages/SummarizedExperiment>. [Internet]. 2022 [citado 2022 sep 27]. Available from: <https://bioconductor.org/packages/release/bioc/html/SummarizedExperiment.html>
64. Logan RK, Tandoc M. Thinking in patterns and the pattern of human thought as contrasted with AI data processing. Information (Switzerland) [Internet]. 2018 abr 8 [citado 2022 sep 29];9(4). Available from: [https://www.researchgate.net/publication/324313060\\_Thinking\\_in\\_Patterns\\_and\\_the\\_Pattern\\_of\\_Human\\_Thought\\_as\\_Contrasted\\_with\\_AI\\_Data\\_Processing](https://www.researchgate.net/publication/324313060_Thinking_in_Patterns_and_the_Pattern_of_Human_Thought_as_Contrasted_with_AI_Data_Processing)
65. Mattson MP. Superior pattern processing is the essence of the evolved human brain. Front Neurosci. 2014;(8 AUG).
66. Chan SW, Looi CK, Ho WK, Huang W, Seow P, Wu L. Learning number patterns through computational thinking activities: A Rasch model analysis. Heliyon [Internet]. 2021 sep 1 [citado 2022 sep 29];7(9). Available from: <http://www.cell.com/article/S2405844021020259/fulltext>
67. Aboitiz F, García RR, Bosman C, Brunetti E. Cortical memory mechanisms and language origins. Brain Lang [Internet]. 2006 jul [citado 2022 sep 29];98(1):40-56. Available from: <https://pubmed.ncbi.nlm.nih.gov/16481035/>

68. Caesar LK, Kvalheim OM, Cech NB. Hierarchical cluster analysis of technical replicates to identify interferences in untargeted mass spectrometry metabolomics. *Anal Chim Acta*. 2018 ago 27;1021:69-77.
69. Tsvetkov D, Kolpakov E, Kassmann M, Schubert R, Gollasch M. Distinguishing Between Biological and Technical Replicates in Hypertension Research on Isolated Arteries. *Front Med (Lausanne)*. 2019 jun 20;6:126.
70. Yu L, Sun Y. Recent advances in protein chromatography with polymer-grafted media. *J Chromatogr A* [Internet]. 2021 feb 8 [citado 2022 sep 28];1638. Available from: <https://pubmed.ncbi.nlm.nih.gov/33453656/>
71. Hao P, Adav SS, Gallart-Palau X, Sze SK. Recent advances in mass spectrometric analysis of protein deamidation. *Mass Spectrom Rev* [Internet]. 2017 nov 1 [citado 2022 sep 28];36(6):677-92. Available from: <https://pubmed.ncbi.nlm.nih.gov/26763661/>
72. Li Z, Beeram SR, Bi C, Suresh D, Zheng X, Hage DS. High-Performance Affinity Chromatography: Applications in Drug-Protein Binding Studies and Personalized Medicine. *Adv Protein Chem Struct Biol*. 2016;102:1-39.
73. Yu B, Yufan F, Guangbo G, Fangjun W. [Advances in chromatography in the study of drug-plasma protein interactions]. *Se Pu* [Internet]. 2021 [citado 2022 sep 28];39(10):1077-85. Available from: <https://pubmed.ncbi.nlm.nih.gov/34505429/>
74. Casanova H. Graficación Estadística y Visualización de Datos. *Statistical Graphing and Data Visualization*. 2017 [citado 2022 sep 30];21(3):54-75. Available from: <https://www.mendeley.com/catalogue/001ab19c-1b27-3230-a927-78ef3de350ec/>
75. Haddaway NR, Feierman A, Grainger MJ, Gray CT, Tanriver-Ayder E, Dhaubanjhar S, et al. EviAtlas: A tool for visualising evidence synthesis databases. *Environ Evid*. 2019 jun 4;8(1).
76. Berengueres J, Fenwick A, Sandell M. Visualización de datos & storytelling. 2020 [citado 2022 sep 30]; Available from: [https://www.amazon.com/-/es/Jose-Berengueres-ebook/dp/B085B8M7XH/ref=tmm\\_kin\\_swatch\\_0?\\_encoding=UTF8&qid=&sr=](https://www.amazon.com/-/es/Jose-Berengueres-ebook/dp/B085B8M7XH/ref=tmm_kin_swatch_0?_encoding=UTF8&qid=&sr=)

**Anexos.**

<b>Individuo y muestra.</b>	<b>Número total de picos.</b>	<b>Tiempo de retención 0%</b>	<b>Tiempo de retención 25%</b>	<b>Tiempo de retención 50%</b>	<b>Tiempo de retención 75%</b>	<b>Tiempo de retención 100%</b>
<b>Individuo 1, muestra A.</b>	1156	17.89	41.71	44.69	50.65	115
<b>Individuo 1, muestra B.</b>	554	11.91	41.71	44.75	53.55	83.54
<b>Individuo 1, muestra C.</b>	1624	14.89	41.66	44.67	50.62	133.9
<b>Individuo 1, muestra D.</b>	1149	23.82	41.68	44.68	50.68	202.5
<b>Individuo 2, muestra A.</b>	1035	17.91	41.71	44.71	50.67	140
<b>Individuo 2, muestra B.</b>	787	17.9	41.73	44.73	50.68	131.1
<b>Individuo 2, muestra C.</b>	1688	17.88	41.64	44.64	47.71	139.9
<b>Individuo 2, muestra D.</b>	1205	17.86	41.67	44.67	53.56	154.9
<b>Individuo 3, muestra A.</b>	1193	20.83	41.67	44.66	47.78	131.2
<b>Individuo 3, muestra B.</b>	927	17.88	41.69	44.68	53.66	128.1
<b>Individuo 3, muestra C.</b>	1428	20.79	41.69	44.68	53.55	134.1
<b>Individuo 3, muestra D.</b>	1069	17.87	41.68	44.69	56.47	172.8

*Tabla 2. Tabla resumen del conteo total de picos por muestra e individuo, con un conteo del número de picos que presento un tiempo de retención relativo (0 %,25%, 50%, 75% y 100% de tiempo de retención), visualizando así, el número de picos identificados por muestra y ancho/duración de los picos cromatográficos .*

	M/Z	Tiempo de retención	Tiempo de retención mínimo	Tiempo de retención máximo	Intensidad en cps
<b>INDIVIDUO 1, MUESTRA A</b>					
<i>Conteo</i>	168	168	168	168	168
<i>Media</i>	838	3138	3115	3163	2.84e+6
<i>Mediana</i>	810	3158	3134	3181	1.67e+6
<i>Desviación estándar</i>	85.8	224	223	222	3.54e+6
<i>Mínimo</i>	750	2508	2502	2538	358032
<i>Máximo</i>	1260	3494	3471	3497	24262784
<b>INDIVIDUO 1, MUESTRA B.</b>					
<i>Conteo</i>	76	76	76	76	76
<i>Media</i>	832	3103	3080	3128	2.02e+6
<i>Mediana</i>	809	3074	3053	3098	1.24e+6
<i>Desviación estándar</i>	60.1	218	216	220	1.99e+6
<i>Mínimo</i>	752	2597	2573	2618	398949
<i>Máximo</i>	1005	3464	3438	3494	10007625
<b>INDIVIDUO 1, MUESTRA C.</b>					
<i>Conteo</i>	294	294	294	294	294
<i>Media</i>	844	3060	3036	3084	2.68e+6
<i>Mediana</i>	811	3069	3046	3090	1.53e+6
<i>Desviación estándar</i>	93.0	281	280	281	3.30e+6
<i>Mínimo</i>	750	2507	2501	2531	357542
<i>Máximo</i>	1261	3495	3484	3498	26090611
<b>INDIVIDUO 1, MUESTRA D.</b>					
<i>Conteo</i>	206	206	206	206	206
<i>Media</i>	839	3060	3036	3085	2.51e+6
<i>Mediana</i>	820	3058	3034	3094	1.69e+6
<i>Desviación estándar</i>	67.8	280	279	280	2.87e+6
<i>Mínimo</i>	750	2511	2502	2532	384590

<i>Máximo</i>	1042	3485	3449	3500	19278086
<b>INDIVIDUO 2, MUESTRA A.</b>					
<i>Conteo</i>	140	140	140	140	140
<i>Media</i>	833	3147	3123	3172	2.74e+6
<i>Mediana</i>	809	3157	3136	3179	1.91e+6
<i>Desviación estándar</i>	80.3	220	219	221	2.99e+6
<i>Mínimo</i>	750	2511	2502	2541	414578
<i>Máximo</i>	1137	3495	3474	3498	19614238
<b>INDIVIDUO 2, MUESTRA B.</b>					
<i>Conteo</i>	136	136	136	136	136
<i>Media</i>	836	3059	3035	3085	2.62e+6
<i>Mediana</i>	809	3035	3014	3060	1.82e+6
<i>Desviación estándar</i>	70.1	252	252	251	2.94e+6
<i>Mínimo</i>	750	2550	2530	2574	290882
<i>Máximo</i>	1041	3495	3480	3498	19304608
<b>INDIVIDUO 2, MUESTRA C.</b>					
<i>Conteo</i>	313	313	313	313	313
<i>Media</i>	849	3065	3042	3090	2.62e+6
<i>Mediana</i>	818	3082	3064	3112	1409975
<i>Desviación estándar</i>	92.3	276	276	275	3.61e+6
<i>Mínimo</i>	750	2519	2501	2540	171797
<i>Máximo</i>	1261	3496	3481	3499	31264819
<b>INDIVIDUO 2, MUESTRA D.</b>					
<i>Conteo</i>	246	246	246	246	246
<i>Media</i>	844	3050	3027	3075	2.70e+6
<i>Mediana</i>	815	3039	3019	3067	1.71e+6
<i>Desviación estándar</i>	82.7	276	276	275	3.31e+6
<i>Mínimo</i>	750	2534	2513	2555	304787
<i>Máximo</i>	1260	3496	3478	3499	26509030
<b>INDIVIDUO 3, MUESTRA A.</b>					
<i>Conteo</i>	169	169	169	169	169
<i>Media</i>	839	3115	3091	3141	2.64e+6

<i>Mediana</i>	809	3170	3143	3202	1805262
<i>Desviación estándar</i>	87.9	284	284	283	3.08e+6
<i>Mínimo</i>	750	2514	2502	2535	312122
<i>Máximo</i>	1260	3480	3462	3495	22572482
<b>INDIVIDUO 3, MUESTRA B.</b>					
<i>Conteo</i>	177	177	177	177	177
<i>Media</i>	842	3052	3027	3079	2.76e+6
<i>Mediana</i>	810	3036	3016	3060	1851065
<i>Desviación estándar</i>	84.7	262	262	262	3.33e+6
<i>Mínimo</i>	750	2542	2521	2563	322844
<i>Máximo</i>	1260	3492	3471	3498	22858446
<b>INDIVIDUO 3, MUESTRA C.</b>					
<i>Conteo</i>	240	240	240	240	240
<i>Media</i>	846	3084	3061	3110	2.56e+6
<i>Mediana</i>	811	3103	3079	3127	1.41e+6
<i>Desviación estándar</i>	96.6	263	263	264	3.17e+6
<i>Mínimo</i>	750	2507	2502	2525	281810
<i>Máximo</i>	1261	3496	3475	3499	24452364
<b>INDIVIDUO 3, MUESTRA D.</b>					
<i>Conteo</i>	216	216	216	216	216
<i>Media</i>	841	3045	3021	3070	2.52e+6
<i>Mediana</i>	811	3027	3006	3051	1.62e+6
<i>Desviación estándar</i>	77.3	291	291	290	3.01e+6
<i>Mínimo</i>	752	2518	2503	2538	394471
<i>Máximo</i>	1260	3495	3477	3498	22866814

Tabla 3. Resumen de las medidas de dispersión, desviación estándar, valores máximo y mínimo de los valores más importantes de cada muestra de cada réplica biológica.