

Análisis integrativo de datos ómicos para el desarrollo de metabotipos



Universitat
Oberta
de Catalunya



UNIVERSITAT DE
BARCELONA

Claudia Arce Cuesta

MU Bioinf. i Bioest.

Statistical Bioinformatics and Machine
Learning

Nombre Tutor/a de TF

Alex Sánchez Pla

Profesor/a responsable de la asignatura

Carles Ventura

Fecha Entrega

15/01/2023



Esta obra está sujeta a una licencia de
Reconocimiento-NoComercial-

SinObraDerivada [3.0 España de Creative
Commons](https://creativecommons.org/licenses/by-nc-nd/3.0/es/)

FICHA DEL TRABAJO FINAL

Título del trabajo:	<i>Análisis integrativo de datos ómicos para el desarrollo de metabotipos</i>
Nombre del autor:	<i>Claudia Arce Cuesta</i>
Nombre del consultor/a:	<i>Alex Sánchez Pla</i>
Nombre del PRA:	<i>Carles Ventura</i>
Fecha de entrega (mm/aaaa):	<i>15/01/2023</i>
Titulación o programa:	<i>Máster universitario Bioinformática y Bioestadística</i>
Área del Trabajo Final:	<i>Statistical Bioinformatics and Machine Learning</i>
Idioma del trabajo:	<i>Castellano</i>
Palabras clave	<i>Metabotipos, Similarity network fusion, clustering</i>
Resumen del Trabajo	
<p>En este trabajo se ha dado una visión global sobre el concepto de 'metabotipo', sus usos más destacables y los métodos que habitualmente se emplean para la caracterización de estos. Se ha propuesto y desarrollado en profundidad el método Similarity Network Fusion para la creación de metabotipos más complejos que puedan integrar diversos tipos de datos ómicos. Finalmente se ha mostrado la aplicación del método sobre un conjunto de datos de pacientes.</p> <p>El trabajo se ha realizado mediante una revisión bibliográfica de estudios</p>	

relacionados con los metabotipos y el método Similarity Network Fusion, y se ha realizado el análisis en el software R mediante el paquete SNFtool. Los datos empleados han sido proporcionados por la Universidad de Barcelona.

Con esto se ha concluido que el método Similarity Network Fusion es una buena manera de crear metabotipos a partir de diferentes tipos de datos, se ha visto que es un método poco costoso a nivel computacional y se han expuesto algunos inconvenientes también que presenta dicho método, como es la falta de interpretabilidad de los metabotipos resultantes.

Abstract

In this work, we give a global vision of the concept of 'metabotype', its most notable uses and the methods that are usually used for the characterization of the metabotypes. The Similarity Network Fusion method has been proposed and further developed for the creation of more complex metabotypes that can integrate various types of omics. Finally, we show the application of the method on a set of patient data.

The work has been carried out through a bibliographic review of studies related to metabotypes and the Similarity Network Fusion method, and the analysis has been made in the R software using the SNFtool package. The data used has been provided by the University of Barcelona.

With this, we have concluded that the Similarity Network Fusion method is a good way to create metabotypes from different types of data, we show that it is not an expensive method at the computational level and some drawbacks that this method presents have also been exposed, such as the lack of interpretability of the resulting metabotypes.

Contenido

Contenido	iii
Lista de figuras	v
Lista de tablas	v
1. Introducción	1
1.1. Contexto y justificación del Trabajo.....	1
1.2. Objetivos del trabajo.....	3
1.3. Impacto en sostenibilidad, ético-social y de diversidad	3
1.4. Enfoque y método seguido.....	5
1.5. Planificación del Trabajo	6
1.6. Análisis de riesgos	8
1.7. Breve resumen de productos obtenidos	9
1.8. Breve descripción de los otros capítulos de la memoria	9
2. Estado del arte.....	11
3. Materiales y métodos.....	13
4. Resultados.....	15
4.1 Metabotipos.....	15

4.2	Metabotipos en la nutrición personalizada	17
4.3	Métodos más utilizados para el estudio de metabotipos	18
4.4	Similarity Network Fusion	22
5.	Conclusiones y trabajos futuros	31
6.	Glosario	35
7.	Bibliografía	36
6.	Anexos	39

Lista de figuras

Figura 1. Planificación de octubre y noviembre	7
Figura 2. Planificación de diciembre, enero y febrero.....	7
Figura 3. Distintos enfoques de la nutrición: General, enfocada en grupos (metabotipos) y personalizada.....	18
Figura 4. Heatmap de los datos de metabolitos (nº clústers = 2)	27
Figura 5. Heatmap de los datos de las variables relacionadas con el riesgo cardiovascular. (nº clústers = 3)	27
Figura 6. Heatmap de los datos de los recuentos de OTUS bacterianos de la microbiota intestinal. (nº clústers = 2)	27
Figura 7. Heatmap obtenido de la red resultante de la fusión de matrices de similitud mediante el método SNF (nº clústers = 3)	29

Lista de tablas

Tabla 1. Valores obtenidos para los clusters de las matrices de similitud iniciales.....	26
Tabla 2. Valores obtenidos para los clusters de la matriz resultante de la fusión de las matrices iniciales	28
Tabla 3. Matriz de concordancia.....	30

1.Introducción

1.1. Contexto y justificación del Trabajo

La medicina y la nutrición actuales tienen como uno de sus objetivos más recientes desarrollar terapias adaptadas totalmente a cada individuo, evaluando su salud y su estado para conocer dosis, dietas, tratamientos o intervenciones precisas.

La alimentación puede ser uno de los factores determinantes para el desarrollo de algunas enfermedades como la diabetes, enfermedades cardiovasculares y varios tipos de cáncer (Hillesheim y Brennan, 2020). Algunas enfermedades cardiovasculares forman parte de la lista de las principales causas de muerte no natural en el mundo, estando en los primeros puestos la cardiopatía isquémica y el accidente cerebrovascular (World Health Organization, 2020). La intervención nutricional en muchos casos puede ser crucial para evitar que un paciente desarrolle alguna enfermedad cardiovascular o para retrasar su evolución (Martinez et al., 2004).

El concepto de metabotipo se define como “la descripción probabilística multiparamétrica de un organismo dado su estado fisiológico basado en el análisis de sus células, biofluidos y tejidos” (Gavaghan et al., 2000). La clasificación de individuos en grupos según su fenotipo metabólico, lo que se conoce como ‘metabotipado’, es un camino muy reciente y en constante evolución en el ámbito de la metabolómica y con muchas aplicaciones en la nutrición y el tratamiento médico.

La aplicación de los metabotipos para encontrar dietas precisas según las características de los individuos puede ayudar a disminuir factores de

riesgo cardiometabólicos y tratar enfermedades relacionadas con la alimentación (Hillesheim y Brennan, 2020). Estos metabotipos se construyen mediante diversos métodos estadísticos bioinformáticos a partir de datos clínicos, metabolómicos y nutricionales de pacientes, de forma que estos individuos pueden agruparse siguiendo patrones en su metabolismo y su expresión génica, con el objetivo de poder tratarlos de una forma más dirigida y precisa.

La nutrición personalizada centrada en cada paciente es un gran reto científico que debe pasar primero por una nutrición dirigida en función de grupos de metabotipos. Las nuevas tecnologías que permitan estudiar los metabotipos de forma precisa constituyen un pilar fundamental en este reto (Palmnäs et al. 2020).

Existen multitud de métodos que pueden utilizarse para la búsqueda de metabotipos. Habitualmente para esta tarea se utilizan métodos tradicionales de agrupación, pero existen otros más novedosos y menos utilizados que pueden dar muy buenos resultados. Entre ellos está 'Similarity Fusion Network', un nuevo método computacional que puede combinar muchos tipos de datos biológicos de pacientes, como datos de expresión de mRNA o metilación de DNA (Wang et al. 2014), en el que se centrará este trabajo en profundidad.

En este trabajo, se buscará clarificar en qué consisten los metabotipos, sus usos y aplicaciones, describir algunos métodos que más se utilizan en la actualidad y explicar la aplicación del método 'Similarity Network Fusion' para la creación de los metabotipos.

1.2. Objetivos del trabajo

Los objetivos principales del trabajo son los siguientes:

- Dar una visión general sobre los metabotipos.
 - Ver en qué consisten los metabotipos, cómo se crean y en qué se utilizan.
 - Identificar los diferentes métodos que se pueden aplicar en la búsqueda de metabotipos más utilizados.
- Explicar el método de “Similarity Network Fusion” utilizado para construir metabotipos.
 - Explicar el método SNF: funcionamiento del algoritmo, parámetros, datos con los que trabaja, etc.
 - Aplicar el método a un conjunto de datos de forma práctica y explicar los resultados.

1.3. Impacto en sostenibilidad, ético-social y de diversidad

Este TF no tiene ningún tipo de impacto medioambiental en su desarrollo ni en sus productos resultantes debido a que se compone de una parte realizada mediante revisión bibliográfica y una parte práctica que utiliza datos extraídos de otros estudios, por lo que no se ha requerido utilizar medios de transporte, materiales para las extracciones de muestras, ni se han generado residuos.

En cuanto a los aspectos éticos y sociales, este trabajo no contribuye a generar ningún tipo de impacto positivo ni negativo. En lo relativo a la responsabilidad social, la igualdad de género y la igualdad, los datos de los pacientes que se han utilizado son totalmente anónimos. A nivel personal solo se conoce su sexo y hay aproximadamente el mismo número de mujeres que de hombres. No se tiene registro de datos relacionados con etnias, religión, orientación sexual o ideologías de los pacientes, puesto que no tienen valor para este estudio.

En cuanto a impactos positivos, este estudio puede contribuir al bien común de la sociedad y de la ciencia ya que puede aportar información útil para la investigación en el ámbito de la nutrición y la medicina personalizada. No tiene ningún tipo de impacto beneficioso a nivel ambiental.

1.4. Enfoque y método seguido

El trabajo está dividido en una parte teórica en la que se hará una revisión sobre los metabotipos, su importancia, aplicaciones y el método de 'Similarity Network Fusion', y una parte práctica donde se aplicará el método revisado sobre un conjunto de datos utilizando paquetes específicos en R.

Se llevarán dos líneas de trabajo correspondientes a la parte teórica (redactar memoria, buscar información sobre estudios previos, entender el método principal...) y la parte más práctica donde se aplicará el algoritmo 'Similarity Network Fusion' con datos. Dichas líneas de trabajo se realizarán de manera sucesiva, comenzando por la parte teórica para tener una visión más profunda del tema y, para finalizar, aplicando lo visto de forma práctica sobre un conjunto de datos de pacientes reales.

La primera toma de contacto relacionada con los metabotipos consistirá en la búsqueda de las aplicaciones de estos y su importancia. Después, se aproximarán los métodos de metabotipado más comunes y se priorizarán los métodos más importantes.

El método de "Similarity Network fusion" se desarrollará en mayor profundidad, explicando todos los parámetros que están implicados, sus funciones dentro del método, su importancia dentro del ámbito de los metabotipos y su utilización para la creación de estos.

Finalmente, se aplicará este método a un conjunto de datos de pacientes anónimos reales que ha sido proporcionado por la Universidad de Barcelona.

1.5. Planificación del Trabajo

Para alcanzar los objetivos propuestos se plantearon las siguientes tareas:

- Objetivo 1.a: Ver en qué consisten los metabotipos, cómo se crean y en qué se utilizan
 - T1. Definir el concepto de metabotipo, aplicaciones e importancia.
- Objetivo 1.b: Identificar los diferentes métodos que se pueden aplicar en la búsqueda de metabotipos más utilizados.
 - T2. Buscar artículos recientes donde se realicen metabotipados
 - T3. Enumerar algunos de los diferentes métodos que utilizan más importantes.
- Objetivo 2.a: Explicar el método SNF
 - T4. Conocer el funcionamiento del método de Similarity Network Fusion.
 - T5. Entender y explicar cada parámetro del método.
- Objetivo 2.b: Aplicar el método a un conjunto de datos de forma práctica y explicar los resultados.
 - T6. Estudiar el conjunto de datos que se va a utilizar.
 - T7. Conocer cómo aplicar SNF en el software R. Paquetes y funciones.
 - T8. Aplicar SNF a conjuntos de datos y analizar los resultados obtenidos.

Estas tareas se organizaron en función del tiempo disponible y la carga de

trabajo que supondría cada una, aunque luego los plazos se adaptaron sobre la marcha. Para visualizar mejor la planificación se han dispuesto las tareas propuestas en un diagrama de Gantt.

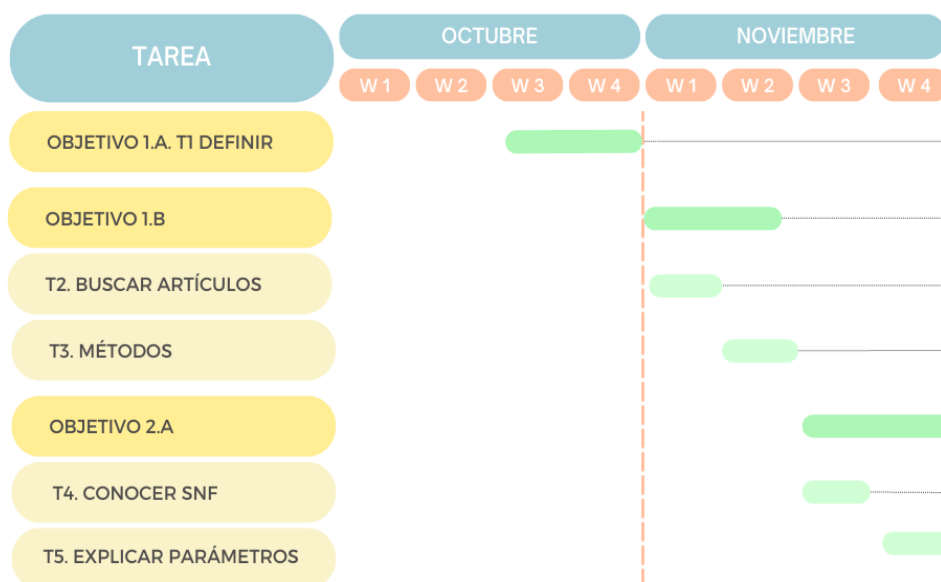


Figura 1. Planificación de octubre y noviembre

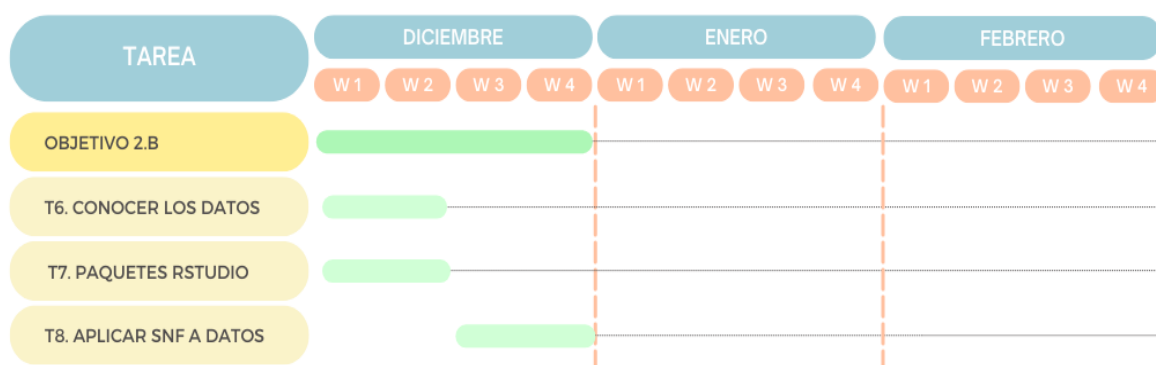


Figura 2. Planificación de diciembre, enero y febrero

Los hitos se han marcado teniendo en cuenta las fechas de entrega de cada una de las PEC. El primer objetivo (1.A y 1.B) se ha completado para

la fecha de entrega de la primera PEC correspondiente con la fase 1.

Se planificó que la fase 2 correspondiese con el objetivo 2 (2.A y 2.B) pero se terminó con una semana de retraso, por lo que la PEC 2 se entregó con lo que se había conseguido hasta el momento.

A lo largo del desarrollo de las tareas se vio que había ciertas tareas que no requerían tanto tiempo como el que se había asignado, y otras tareas que requerían algo más de tiempo de lo planificado. De esta manera, se ha compensado el tiempo faltante de unas tareas con el sobrante de otras.

1.6. Análisis de riesgos

Existen algunos factores que han podido afectar negativamente en la realización del trabajo que se han propuesto durante la etapa de planificación del proyecto.

- Tiempo reducido o posible planificación incorrecta
- Retraso en la finalización de las tareas. Un retraso en una tarea puede afectar a todo el trabajo final.
- Encontrarse demasiados métodos de creación de metabotipos y sea complicado determinar los principales.
- Dificultades a la hora de realizar la parte práctica en R por falta de información sobre el método Similarity Network Fusion.
- Los datos sanitarios utilizados para la parte práctica pueden no ser válidos por contener información personal de pacientes.

- Que el método SNF no se pueda aplicar correctamente a los datos
- El conjunto de datos puede no tener suficientes variables relacionadas entre sí o muestras insuficientes.

1.7. Breve resumen de productos obtenidos

- Memoria del trabajo final: los capítulos se describen en detalle en el siguiente epígrafe.
- Código empleado para mostrar la aplicación del método Similarity Network Fusion (en GitHub).
- Presentación virtual final compuesta por un documento donde aparecen las diapositivas y un video explicativo del trabajo final resumiendo los objetivos, resultados y conclusiones.

1.8. Breve descripción de los otros capítulos de la memoria

- Estado del arte: breve introducción sobre lo que se conoce hasta ahora sobre el ámbito de estudio de los metabotipos
- Material y métodos: metodología seguida para la elaboración del trabajo final. Origen e información del conjunto de datos y software utilizado.
- Resultados:
 - Metabotipos: qué son, aplicaciones, estudios, etc.
 - Metabotipos en la nutrición personalizada

- Métodos más utilizados para la creación de metabotipos: capítulo dedicado a introducir los algoritmos que se utilizan para el estudio de metabotipos habitualmente.
- Similarity Network Fusion:
 - Parámetros: se explica la función de cada parámetro dentro del algoritmo.
 - Similarity Network Fusion y su aplicación en los metabotipos: cómo se utiliza el algoritmo en el ámbito de los metabotipos, beneficios que aporta frente a otros algoritmos.
 - Aplicación del método Similarity Network Fusion a un grupo de pacientes mediante R: ejemplo de aplicación del algoritmo mediante un caso práctico con datos de pacientes reales.

2.Estado del arte

Los metabotipos son grupos que se construyen en base a características metabólicas similares mediante algoritmos de clasificación. Algunas variables que se han utilizado para la creación de metabotipos orientados a ofrecer una nutrición dirigida son los niveles de colesterol total, el colesterol HDL o de alta densidad y la glucosa. Agrupando a los pacientes en función de características similares pueden darse consejos dietéticos que tengan mejores resultados (Hillesheim et al, 2020).

El concepto de metabotipo se creó a principios de este siglo. Esta definición aparece en la bibliografía por primera vez en un estudio realizado en el año 2000 por C.L. Gavaghan y otros, titulado “An NMR-based metabonomic approach to investigate the biochemical consequences of genetic strain differences: application to the C57BL10J and Alpk:ApfCD mouse”. En este estudio, definieron el concepto de metabotipo como “la descripción probabilística multiparamétrica de un organismo dado su estado fisiológico basado en el análisis de sus células, biofluidos y tejidos”. Las aplicaciones que se proponen en el estudio de C.L. Gavaghan et al. son conocer las consecuencias metabólicas de la modificación genética o de patologías de forma cuantitativa y así, poder realizar los correspondientes análisis estadísticos. Mediante este nuevo concepto que habían creado estudiaron las diferencias que existían en la excreción de metabolitos en la orina en dos modelos de ratón, y vieron que según la cepa existían diferencias en varias rutas metabólicas.

Actualmente, se sigue aplicando este concepto en grupos de individuos para experimentación con animales y en el estudio de patologías en pacientes humanos, pero aún continúa desarrollándose el término de metabotipo para su utilización con fines médicos. Un ejemplo ilustrativo de metabotipos estudiados en seres humanos con fines médicos fue el

estudio realizado por M. Urpi Sarda et al. en 2019, “Non-targeted metabolomic biomarkers and metabotypes of type 2 diabetes: A cross-sectional study of PREDIMED trial participants”. Realizaron una investigación en la que buscaban caracterizar la huella metabólica urinaria y los metabolitos asociados con la diabetes tipo 2 (T2D), para clasificar la población en metabotipos relacionados con dicha enfermedad. A grandes rasgos, encontraron que había diferencias significativas en 33 metabolitos y se distinguieron dos metabotipos en personas con T2D y otros dos, en personas no diabéticas.

Una de las aplicaciones de los metabotipos más innovadoras y que está prosperando a grandes pasos es el ámbito de la nutrición para el tratamiento y la prevención de patologías. Como Hillesheim y Brennan propusieron en 2020 en su artículo “*Metabotyping and its role in nutrition research*”, el uso de los metabotipos puede ser un paso intermedio muy importante hacia la nutrición totalmente personalizada para cada paciente.

Los metabotipos son un concepto relativamente reciente y en continuo desarrollo que, con certeza se puede afirmar que tienen un gran futuro en la investigación relacionada con la salud, la medicina y la nutrición personalizada.

3. Materiales y métodos

En primer lugar, se ha realizado una revisión de varios estudios que se han publicado hasta la fecha mediante PubMed y Google Scholar sobre el ámbito de los metabotipos, los métodos bioinformáticos que se utilizan para crearlos y sus aplicaciones más destacadas. También se ha tratado de encontrar información sobre el origen del término “metabotipo” con el fin de tener una definición determinada.

Se han utilizado algunas combinaciones de palabras clave para buscar estudios y artículos. Las palabras empleadas en la búsqueda han sido principalmente en inglés, puesto que en este idioma existen más estudios publicados que en castellano. La terminología utilizada para la búsqueda ha sido la siguiente: metabotype, metabolic profiling, metabolic phenotype, targeted nutrition, metabolomics, nearest neighbor, clustering, Similarity network fusion, SNF, metabolic classification.

Para aplicar el método Similarity Network Fusion se ha empleado un conjunto de datos correspondiente a más de 400 pacientes anónimos de los que se tiene información de 14 variables relacionadas con el riesgo cardiovascular (HDL, LDL, presión sanguínea...), composición de la microbiota intestinal con más de mil variables correspondientes con categorías taxonómicas y 484 metabolitos. Los datos de la composición de la microbiota intestinal están formados por valores de OTUs o unidades taxonómicas operativas que han permitido contabilizar las diferentes especies y familias de bacterias encontradas en cada paciente. Se descartaron previamente todos los pacientes de los que faltaba información de alguna de las variables. Dichos datos fueron proporcionados por la Universidad de Barcelona y forman parte de un proyecto europeo de investigación relacionado con el síndrome metabólico y las enfermedades cardiovasculares.

Inicialmente tuvieron que tratarse los datos antes de comenzar el análisis puesto que se tenía información por triplicado de cada paciente, correspondiente con tres puntos temporales en los que se les tomaron valores para todas las variables. Se tomaron solo los valores de cada paciente para el punto temporal inicial o 'baseline' y se descartaron los otros dos.

El análisis de los datos se ha realizado mediante R, gracias al programa Rstudio (R version 4.1.1). El paquete de R que se ha utilizado para aplicar el método "Similarity network fusion" de forma práctica ha sido SNFtool en su versión 2.3.1. Para el desarrollo de esta parte, se ha decidido emplear R en lugar de Python, a pesar de que podía hacerse en cualquiera de los dos, debido a que R es un lenguaje de programación más utilizado en el máster que el lenguaje Python.

El código empleado para realizar el método Similarity Network Fusion para la creación de metabotipos puede obtenerse en [este repositorio de GitHub](#), y el workflow seguido para el análisis ha sido el siguiente:

1. Preparación de los datos: se han eliminado pacientes con datos incompletos y las variables no cuantitativas. Se han escalado las variables.
2. Cálculo de las matrices de distancia y matrices de afinidad.
3. Se calcula el número de clústers, se aplican y se visualizan mediante heatmaps.
4. Aplicación de Similarity Network Fusion sobre las matrices de afinidad creadas anteriormente. Se muestran los clústers de la red fusionada mediante heatmaps.
5. Revisión de los resultados: evaluación, visualización de matriz de concordancia.

4.Resultados

4.1 Metabotipos

La metabolómica es la rama de las ciencias ‘ómicas’ que estudia las moléculas producidas como resultado del metabolismo de una célula, tejido u organismo. Mediante la metabolómica podemos conocer el funcionamiento de las rutas metabólicas y sus alteraciones en situaciones patológicas (Ruiz Esparza-Garrido et al., 2014). Los diferentes metabolitos resultantes de las rutas metabólicas pueden estudiarse mediante diversas técnicas de análisis, pero la forma más común de obtenerlos es mediante muestras biológicas de sangre, orina, heces o saliva. Utilizando muestras tan sencillas y rápidas de obtener puede estudiarse el estado metabólico de los individuos para evaluar su salud y dar consejo personalizado (Noerman et al., 2020).

Los fenotipos están determinados por los genes junto con la interacción con el medio ambiente, lo que se conoce como la ciencia de la “epigenética”. Estos factores epigenéticos ambientales pueden ser el estilo de vida, la alimentación, la calidad del aire, factores estresantes o el nivel de actividad física. Todas estas características del ambiente pueden repercutir en los genes de un individuo modificando su expresión, activándolos o inactivándolos y, como consecuencia, pueden llegar a influir en el desarrollo o la prevención de enfermedades (Petersen et al., 2014)

La clasificación de individuos en grupos según su fenotipo metabólico se conoce como “metabotipado”, y los grupos o clases obtenidos se denominan “metabotipos” o fenotipos metabólicos. Los metabotipos agrupan individuos que mantienen características similares, de forma que puedan caracterizarse así variaciones dentro de grupos de pacientes con ciertas enfermedades y condiciones que impliquen alteraciones en rutas

metabólicas.

Estos metabotipos se pueden construir de forma automática mediante algoritmos informáticos de clasificación, como el método de Análisis de Componentes Principales, el clustering jerárquico o el algoritmo de K – medias (Riedl et al., 2017). Para ello, se debe disponer de un conjunto de datos de pacientes donde las variables sean los valores de los metabolitos o determinadas características que se van a estudiar para crear los metabotipos.

Los metabotipos o fenotipos metabólicos habitualmente suelen venir dados por los valores metabolómicos que se pueden encontrar en los pacientes. Pero, además de los metabolitos, existen otras características que pueden utilizarse para realizar un análisis integrativo más amplio. En los últimos años, se ha estudiado la clasificación de pacientes en función de la antropometría (Palmas et al., 2020), mediciones clínicas como los datos cardiometabólicos (Riedl et al. 2018) y la composición de la microbiota intestinal, cuyos grupos de clasificación se denominan “enterotipos” (Romo-Vaquero et al. 2019). Consecuentemente, podemos deducir que, combinando todos estos datos, los metabotipos basados en biomarcadores, medidas antropométricas y la composición de la microbiota mediante métodos informáticos pioneros, podrán desarrollarse grupos metabotípicos aún más precisos y mejor enfocados en las necesidades de las personas.

Los metabotipos no solo se utilizan con pacientes humanos y animales de laboratorio como ya se ha mostrado, sino que pueden servir también para la caracterización de otros seres vivos, como animales de laboratorio o seres vivos de interés biotecnológico e industrial. Se ha visto que la caracterización de metabotipos puede ser muy útil para el estudio del metabolismo de diversas bacterias y su relación con, por ejemplo, la resistencia a fármacos bactericidas. En el estudio realizado por Oriane

Moyne et al. publicado en 2021, se muestra un ejemplo muy interesante en el que aplican el concepto de metabotipo relacionándolo con la resistencia a antibióticos y la virulencia en la especie *Pseudomonas aeruginosa*.

4.2 Metabotipos en la nutrición personalizada

La nutrición dirigida o ‘targeted nutrition’ es una forma de dar consejo nutricional enfocada en un grupo o una persona con unas características determinadas. Estas características de clasificación pueden ser los metabotipos, los cuales son grupos establecidos en función de fenotipos metabólicos. Gracias a estas clasificaciones pueden aconsejarse dietas más específicas para obtener respuestas metabólicas en los individuos con la finalidad de, por ejemplo, tratar enfermedades o prevenirlas (Hillesheim y Brennan, 2020).

La nutrición personalizada consiste en dar consejo nutricional enfocado totalmente en las características y requisitos de la persona a la que se está tratando, su anatomía, su genética y su metabolismo. Este tipo de nutrición es muy compleja y requiere gran investigación y algunos hitos previos que deben alcanzarse primero. En otras palabras, el camino hacia la nutrición totalmente personalizada debe pasar primero por una nutrición según metabotipos que permitan caracterizar el comportamiento metabólico y, por tanto, la respuesta metabólica de grupos de pacientes concretos (Palmnäs et al. 2020).

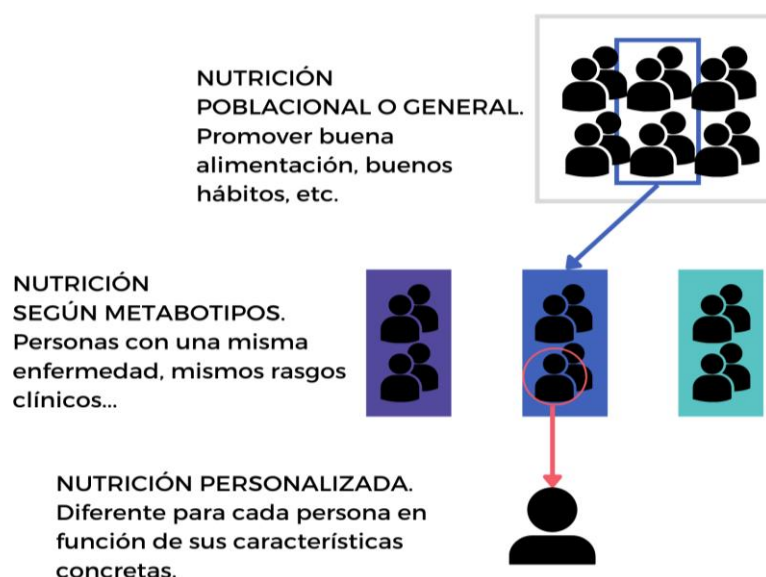


Figura 3. Distintos enfoques de la nutrición: General, enfocada en grupos (metabotipos) y personalizada.

4.3 Métodos más utilizados para el estudio de metabotipos

Los metabotipos se construyen mediante métodos informáticos de análisis de clustering o agrupación. Los métodos de aprendizaje no supervisado que, según se ha observado, se utilizan más frecuentemente en el estudio de metabotipos (Riedl et al., 2017) son los siguientes:

K – medias

El método de K-medias, K-means o algoritmo de Lloyd es un método no jerárquico para agrupar datos en el que se divide un conjunto en k grupos. Estos grupos están determinados por valores medios o *centroides* en torno a los que se aproximan las observaciones, de forma que, cada observación se asigna a un grupo o clúster mediante iteraciones.

El método realiza iteraciones, es decir, repeticiones de un proceso, para ir ajustando cada grupo observación por observación.

La iteración se realiza en varios pasos: primero, se selecciona un número de clústers que se quieren formar (k), se seleccionan k puntos al azar que serán los primeros *centroides*. Después, se asigna cada observación al punto o *centroide* más cercano y, finalmente, se recalculan los *centroides* realizando la media de las observaciones del grupo que se asignaron en el primer paso. Finalmente, se eligen los *centroides* donde las distancias de las observaciones a dichos puntos sean mínimas (Rencher, & Christensen, W. F., 2012).

Un estudio que ejemplifica el uso de este método para el desarrollo de metabotipos fue el realizado por M. Urpi-Sarda et al. en 2019, titulado “Non-targeted metabolomic biomarkers and metabotypes of type 2 diabetes: A cross-sectional study of PREDIMED trial participants”. En dicho estudio, buscaron clasificar los metabotipos de un conjunto de pacientes entre los que había personas con Diabetes tipo 2 y personas sanas. Trataron de caracterizar grupos en función de metabolitos encontrados en la orina. Mediante el uso del método de K-medias, encontraron que había cuatro metabotipos diferentes en la población de estudio.

Hierarchical Cluster Analysis (Análisis de conglomerados jerárquico)

El análisis de conglomerados jerárquicos o Hierarchical Cluster Analysis (HCA) es un algoritmo de clasificación jerárquica que agrupa según un dendrograma que está ordenado. Para crear dicho dendrograma se agrupan los atributos de las observaciones según similitud de clústers y las distancias entre observaciones (como la Euclidiana o la distancia de Manhattan).

Este método puede funcionar de forma aglomerativa o divisiva, yendo desde las “ramas” hacia la “raíz” del dendrograma o al revés. De esta

forma puede considerarse el conjunto total de datos un solo grupo e ir dividiéndose o, por el contrario, ir agrupando observaciones individuales (Zhang et al., 2017).

Un estudio en el que se utiliza el método de clustering jerárquico es el realizado por M. Urpi-Sarda et al. en 2019, mencionado en el apartado anterior, en el que utilizan este algoritmo para ordenar de forma jerárquica y visualizar posteriormente las relaciones y asociaciones entre los grupos obtenidos mediante el algoritmo de K-medias. Es común encontrar estos dos métodos combinados para la creación de metabotipos.

Principal Component Analysis (Análisis de componentes principales)

Análisis de Componentes Principales o Principal Component Analysis (PCA) es un método de visualización de datos que permite agrupar y relacionar datos en análisis multivariante. Se representan los datos en un espacio con K dimensiones que corresponden con el número de variables, y mediante PCA se consiguen reducir a un número menor de variables llamadas “componentes principales”. Esta es una técnica para que sirve para reducir la dimensionalidad de conjuntos de datos, de forma que sean más fáciles de interpretar. El método PCA permite simplificar los datos para su estudio, pero no afecta a la pérdida de información (Jolliffe y Cadima, 2016).

PCA se realiza mediante un proceso iterativo donde se descomponen linealmente las variables para convertirlas en componentes principales. La primera componente principal es la combinación lineal de variables que tienen la máxima varianza. De esta manera lo que se busca es la máxima dispersión de las observaciones. Después, la segunda componente también será una combinación lineal que tenga máxima varianza, pero en

una dirección ortogonal a la primera componente. De esta forma, el número máximo de componentes principales que podemos obtener es igual al número de muestras o al número de variables menos uno (Rencher, & Christensen, W. F., 2012).

Este método fue el utilizado por C.L. Gavaghan et al. en el año 2000 en su estudio “An NMR-based metabonomic approach to investigate the biochemical consequences of genetic strain differences: application to the C57BL10J and Alpk: ApfCD mouse”. Dicho artículo es el primero que sugirió el concepto inicial de ‘metabotipo’, y en él obtuvieron dos grupos de ratones estudiando como variables varios metabolitos encontrados en las excreciones, donde se obtuvo que los grupos coincidían con las dos variantes de ratones.

4.4 Similarity Network Fusion

Con el creciente desarrollo de nuevas técnicas para la obtención y procesamiento de macrodatos o Big Data han surgido nuevas aplicaciones dentro de la biología y, en concreto, en el ámbito de las ciencias ómicas (Stephens et al., 2015). Dentro de los algoritmos de clasificación no supervisada que existen para la clasificación de datos a gran escala encontramos el método Similarity Network Fusion (SNF) creado por Wang et al. en el año 2014, enfocado en sus inicios en su uso dentro del ámbito de la genómica y la transcriptómica. Este método basado en el clustering consiste, inicialmente, en crear redes (networks) individuales de muestras u observaciones según matrices de similitud o afinidad. Posteriormente, fusionan de forma iterativa dichas matrices de similitud de las muestras en una sola matriz o red que representa todo el conjunto mediante combinaciones no lineales. Sobre dicha red se realiza una clusterización espectral que permite ver qué muestras pueden estar asociadas por ser similares o por seguir determinados patrones. Las matrices de similitud creadas en el primer paso se obtienen a partir de matrices de distancia mediante el método de k-vecinos cercanos y deben tener una medida de similitud (σ) que determina cómo se realizan las agrupaciones de las muestras.

El método creado en 2014 en su inicio se desarrolló con el objetivo de que fuese una herramienta para la integración de datos ómicos, de manera que pudiesen combinarse diferentes tipos de datos. Similarity Network Fusion permite integrar distintos tipos de datos en una misma red, por lo que puede crearse una red compleja que combine diferentes datos ómicos y de otros tipos.

Gracias a este algoritmo computacional, pueden analizarse en conjunto diferentes datos clínicos como niveles de colesterol o nivel de riesgo cardiovascular, datos de metabolómica y transcriptómica e, incluso,

composición de la microbiota intestinal, todo ello mediante un mismo algoritmo.

Parámetros

Los parámetros que deben determinarse antes de ejecutar el algoritmo son los siguientes:

- Número de vecinos cercanos: debe determinarse el número de vecinos o muestras cercanas que se tomarán para crear las matrices de similitud primarias. También debe determinarse un número de vecinos cercanos a la hora de fusionar las matrices de similitud en la red compleja. Habitualmente comprendido entre 10 y 30 (Wang et al., 2014).
- Numero de clústers: a la hora de realizar heatmaps se emplea un clustering espectral en el que debe indicarse el número de grupos que se quieren crear. El número de clústers óptimo se puede determinar mediante una función incluida en el paquete de R SNFtool (`estimateNumberOfClustersGivenGraph`).
- Sigma: es el kernel de similitud que se utiliza para realizar el cálculo de afinidad. Se suele utilizar entre 0.3 y 0.8 (Wang et al., 2014).
- Número de iteraciones: se determinan cuantas repeticiones se harán para obtener la red fusionada global.

Similarity Network Fusion y su aplicación en los metabotipos

Actualmente, este método está en auge en las ciencias ómicas. Se ha aplicado en el campo de la genómica computacional para el estudio de la expresión de genes, la metilación de DNA y la expresión de miRNA.

Similarity Network Fusion permite crear redes de muestras o pacientes por lo que es aplicable a la creación de metabotipos, ya que lo que se busca con esto es agrupar pacientes con características similares. Como este método puede aplicarse también a expresión de genes, permite incorporar y combinar datos de composición de la microbiota intestinal mediante el estudio de Unidades Taxonómicas Operativas (OTUS) como la proteína 16s bacteriana (He et al., 2015) con variables utilizadas habitualmente en el estudio de metabotipos, como la cuantificación de metabolitos. De esta manera, podemos integrar diversos datos ómicos mediante un único método.

El método SNF, debido a su versatilidad y su capacidad para ser aplicado en extensos y complejos conjuntos de datos, es una buena opción con mucho potencial para el análisis de metabotipos, de forma que se pueden integrar gran cantidad de variables de diversos tipos, como datos de metabolómica, pero también de transcriptómica o incluso datos antropométricos. Así, crear metabotipos complejos se vuelve una tarea más sencilla puesto que analizar los conjuntos de variables en su inicio por separado puede ayudar a tratar con el ruido, las escalas diferentes o posibles sesgos.

Aplicación del método Similarity Network Fusion a un grupo de pacientes mediante R

El paquete SNFtool permite aplicar el algoritmo de Similarity Network Fusion en el software R. Existen paquetes como SNFpy que sirven para aplicar Similarity Network Fusion en Python, pero este trabajo se centrará en su aplicación en R.

Para mostrar la aplicación del método Similarity Network Fusion para el desarrollo de metabotipos se utiliza a continuación un conjunto de datos correspondiente a pacientes anónimos de los que se tiene información de la composición bacteriana de su microbiota intestinal, más de 400 metabolitos y variables relacionadas con el riesgo cardiovascular como HDL, LDL y triglicéridos. Gracias a este dataset tan amplio y variado se mostrará a continuación cómo combinar diferentes datos con una cantidad bastante alta de variables mediante el método SNF, correspondientes con diferentes tipos de datos ómicos y de otra naturaleza.

En el inicio, se prepararon los datos eliminando dos de los tres puntos temporales en los que se recogió información sobre cada paciente, dejando solo los datos del primer momento en el que se tomó información.

Tras este preprocesado externo de los datos, se realizan las matrices de similitud. Al ver dichas matrices de similitud de forma gráfica mediante el uso de heatmaps, se observó que podía existir algún tipo de agrupamientos dentro de cada conjunto de datos por separado, aunque en los datos de composición bacteriana de la microbiota aparecen los clústers bastante poco definidos. Estos heatmaps nos indican que, desde el inicio, podemos tener indicios de agrupamientos por similitud entre muestras, lo que facilita que al realizar la fusión de matrices de similitud encontremos buenas agrupaciones. Los valores de los clústers fueron calculados mediante una función contenida en el paquete SNFtool de R

que proporciona, según varios métodos ('eigen-gap' y 'rotation cost'), los posibles valores que puede tomar el número de clústers. Los valores óptimos para los clústers que se han obtenido para cada conjunto de variables han sido los que se muestran en [Tabla 1. Valores obtenidos para los clusters de las matrices de similitud iniciales.](#)

	Eigen-gap best	Eigen-gap 2nd best	Rotation cost best	Rotation cost 2nd best	Valor final escogido
Metabolitos	2	4	2	4	2
Relacionadas con cardiovascular	3	2	3	2	3
Composición microbiota intestinal (OTUs)	2	3	3	2	2

Tabla 1. Valores obtenidos para los clusters de las matrices de similitud iniciales

Los heatmaps resultantes de la matriz de similitud de cada conjunto de variables se muestran en la [Figura 4](#), la [Figura 5](#) y la [Figura 6](#). Los clústers en el heatmap son visibles, pero no están demasiado marcados y se ven muchos puntos correspondientes con similitudes entre muestras no clasificadas fuera de ellos.

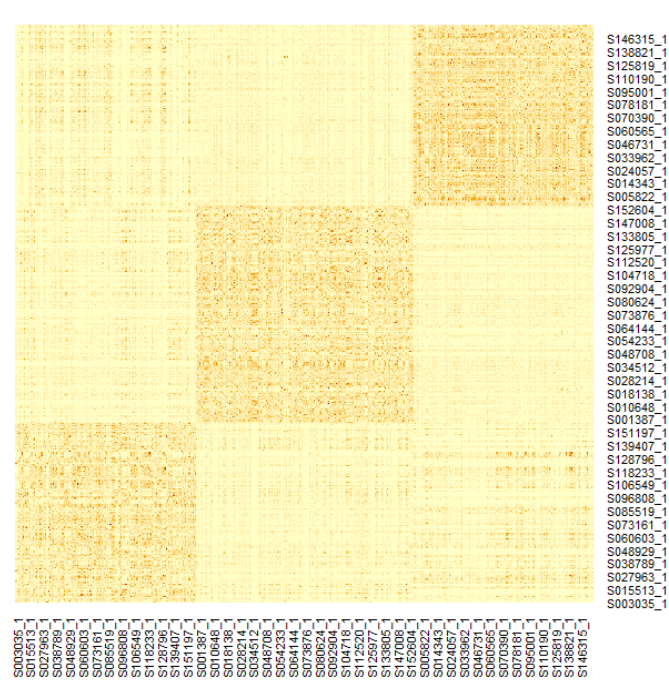
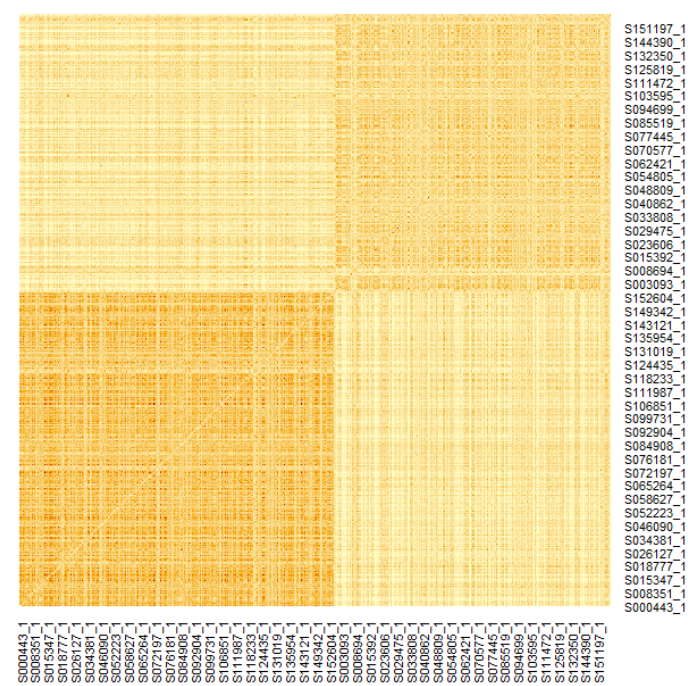


Figura 4. Heatmap de los datos de metabolitos (n° clústers = 2)

Figura 5. Heatmap de los datos de las variables relacionadas con el riesgo cardiovascular. (n° clústers = 3)

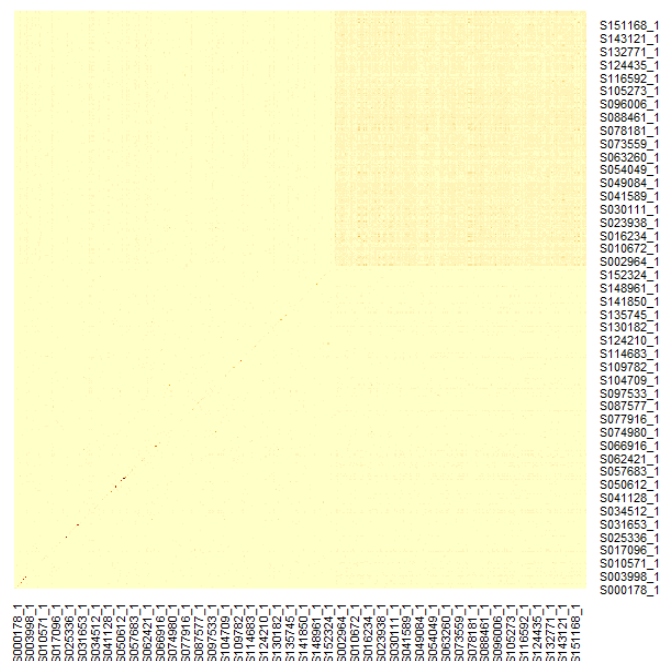


Figura 6. Heatmap de los datos de los recuentos de OTUS bacterianos de la microbiota intestinal. (n° clústers = 2)

Una vez visualizadas las matrices de similitud de forma individual mediante heatmaps, se fusionan mediante SNF y se realiza de nuevo una clusterización para crear los metabotipos. El número óptimo de clústers, igual que en el paso anterior, se ha calculado mediante una función integrada en el paquete SNFtool, y se muestran los valores obtenidos en la [Tabla 2](#).

	Eigen-gap best	Eigen-gap 2nd best	Rotation cost best	Rotation cost 2nd best	Valor final escogido
Matriz SNF	3	4	3	5	3

Tabla 2. Valores obtenidos para los clusters de la matriz resultante de la fusión de las matrices iniciales

Como heatmap resultante de la matriz global de la fusión se obtiene la [Figura 7](#), donde se aprecian tres clústers bastante bien diferenciados. Dichos clústers corresponderían con 3 metabotipos encontrados en el grupo de pacientes.

Se puede ver que, al igual que en los heatmaps individuales obtenidos de los datos antes de ser fusionados, se ve algo de ruido, pero se pueden diferenciar bien los clústers correspondientes con los tres metabotipos que podemos encontrar en el grupo de pacientes. Es importante visualizar los clústers mediante heatmap tras calcular el número óptimo mediante la función específica del paquete SNFtool, ya que en ocasiones nos da varios valores diferentes posibles y deberemos elegir uno de todos.



Figura 7. Heatmap obtenido de la red resultante de la fusión de matrices de similitud mediante el método SNF (n° clústers = 3)

	Red final	Metabolitos	Riesgo cardiovascular	Microbiota intestinal
Red final	1.0000	0.0538	0.47202	0.00998
Metabolitos	0.05383	1.0000	0.02016	0.00195
Riesgo cardiovascular	0.47202	0.02016	1.0000	0.00228
Microbiota intestinal	0.00998	0.00195	0.00228	1.0000

Tabla 3. Matriz de concordancia

Finalmente, tras realizar la matriz de concordancia se obtiene la [Tabla 3](#). Dicha matriz nos indica en qué proporciones concuerdan los clústers realizados entre las diferentes matrices de similitud. Lo más interesante a la hora de observar la tabla es la concordancia de la matriz o red final con el resto de las matrices de similitud individuales. Observando esto, podemos ver que los clústers creados en la matriz final concuerdan casi al 50% con los clústers de los datos de variables relacionadas con el riesgo cardiovascular de los pacientes, mientras que la concordancia es mucho más baja con los demás datos. Esto puede deberse a que dichas variables tengan más importancia a la hora de realizar los clústers, a la existencia de posible ruido en las muestras o al alto número de variables que contienen dichos datos.

5. Conclusiones y trabajos futuros

Mediante este trabajo final de máster, se ha tratado de dar una visión global de qué son los metabotipos, sus usos más importantes y cómo se aplican a nivel bioinformático. El estudio de los metabotipos es un camino muy reciente y en continua evolución. No obstante, queda mucha investigación por delante en cuanto a los métodos informáticos óptimos para su desarrollo y su aplicación para poder alcanzar una nutrición completamente personalizada. Es necesario que el desarrollo de nuevas técnicas de integración de datos ómicos siga alcanzando nuevas metas a medida que se avanza en la investigación en el ámbito de las ciencias ómicas.

De los resultados obtenidos y lo visto durante el proceso podemos concluir los siguientes puntos:

- Similarity Network Fusion es un método aplicable a la creación de metabotipos y se puede usar en conjuntos de datos que contengan un número alto de variables.
- Es un método rápido y poco costoso a nivel computacional, aunque los datos sean muy extensos. Esto es algo ciertamente positivo, puesto que realizar análisis con numerosas variables, en ocasiones y dependiendo del método que se esté empleando, puede dar lugar a tiempos de espera que ralenticen el trabajo.
- A la hora de utilizar el método SNF, puede encontrarse cierto ruido al visualizar los datos mediante heatmaps. Esto puede deberse a la naturaleza de las variables, al número alto de variables o al método en sí. A pesar de esto, este método consigue reducir ciertos problemas en estos aspectos realizando las matrices de similitud

iniciales de manera aislada para cada conjunto.

- El número de clústers o metabotipos óptimo para el agrupamiento de las muestras puede calcularse, pero es importante visualizar el heatmap para comprobarlo. Se ha visto que la función que facilita este proceso que provee el paquete SNFtool de R puede dar varios números de clústers óptimos distintos. En este caso, el usuario es el que debe escoger el número mas óptimo, y la forma de hacerlo es visualizando los heatmaps para encontrar el que presente unos clústers más definidos y menos puntos de similitud fuera de los clústers, siempre teniendo en cuenta que no haya un sobreajuste.
- El método Similarity Network Fusion no permite interpretar el porqué de que los clústers o metabotipos sean realizados de cierta manera, qué variables son las más decisivas y cuáles no tienen tanta importancia. Esto es un gran inconveniente ya que puede ser de gran utilidad conocer qué define a cada metabotipo. Aún así, cuando se realiza un análisis con un numero de variables demasiado alto, la interpretabilidad de los agrupamientos pasa a ser una tarea que se deja en manos de la computación, puesto que suele ser un trabajo demasiado complejo.

Los resultados obtenidos se han ajustado a lo que se esperaba, ya que lo que se quería ver principalmente era que el método Similarity Network Fusion fuese aplicable a la creación de metabotipos. En cuanto a la elaboración del proyecto, esta ha sido algo más compleja de lo que se esperaba, aun así, se han conseguido alcanzar todos los objetivos previstos adaptando las fechas de finalización de cada tarea.

Análisis del seguimiento de la planificación y la metodología

En general, no se ha seguido completamente la planificación puesto que ha habido tareas para las que se sobreestimaba la carga de trabajo, mientras que otras han requerido más dedicación. Este hecho ha permitido compensar los tiempos de elaboración del trabajo de unas tareas con mayor carga de la esperada con otras más sencillas. De esta manera, no se ha seguido la planificación inicial, pero se han cumplido las tareas y los objetivos previstos.

Se ha requerido realizar algunas pequeñas modificaciones en el calendario de trabajo debido a atrasos, pero han sido compensados después, por lo que se han conseguido cumplir todos los objetivos.

Líneas de trabajo futuro que no han podido explorarse

Debido al tiempo y los recursos limitados, hay algunos temas o ámbitos que no se han podido tratar en este trabajo final pero que pueden ser opciones muy interesantes para estudios futuros. El método SNF puede aplicarse también para realizar predicciones utilizando grupos de datos de entrenamiento y de prueba. De esta forma, podrían clasificarse pacientes de prueba en determinados metabotipos previamente conocidos. Esto es posible realizarlo gracias al mismo paquete SNFtool de R que se ha utilizado para el análisis, pero este estudio no llega a contemplar esta característica debido a la limitación temporal.

Otro aspecto interesante del método Similarity Network Fusion es que, por su versatilidad, puede servir no solo para el estudio de metabotipos, sino también para el estudio de enterotipos de forma aislada. Los enterotipos son una forma de clasificar según la composición de la microbiota

bacteriana intestinal, y puede ser útil para estudiar absorción de nutrientes, enfermedades gastrointestinales o intolerancias alimentarias (Cheng y Ning, 2019).

La interpretación de las variables, las interacciones entre ellas y el estudio de las características más importantes que definen a cada metabotipo ha sido una tarea que no se ha podido llevar a cabo en este trabajo dado su complejidad, pero sería interesante desarrollar un método o workflow para realizar esta tarea en futuros estudios.

6. Glosario

Clustering: agrupamiento de objetos similares (en este caso, pacientes) en función de los valores que toman para las variables de estudio, para crear grupos o clústers.

Fenotipo: rasgos observables físicos y metabólicos que se manifiestan a partir del genotipo y la interacción con los factores ambientales.

Heatmap: en este estudio, gráfica en dos dimensiones creada a partir de una matriz de similitud donde los valores de similitud entre dos muestras que son más altos aparecen marcados en cierto color o con más intensidad que los valores bajos.

Matriz de similitud o afinidad: matriz donde se presenta la semejanza entre muestras pareadas. Se crea a partir de matrices de distancias. Los valores funcionan al revés que la matriz de distancias, de forma que, las muestras similares en la matriz de similitud tendrán valores altos, mientras que las muestras mas diferentes tendrán valores cercanos a cero o negativos.

Metabotipo: la descripción probabilística multiparamétrica de un organismo dado su estado fisiológico basado en el análisis de sus células, biofluidos y tejidos (Gavaghan et al., 2000).

Nutrición personalizada: nutrición enfocada en dar consejo específico para el individuo, teniendo en cuenta sus características fisiológicas, genéticas y metabólicas.

PCA: Principal Component Analysis.

SNF: Similarity Network Fusion.

7. Bibliografía

Cheng M, Ning K. Stereotypes About Enterotype: The Old and New Ideas. *Genomics Proteomics Bioinformatics*. 2019 Feb;17(1):4-12.

Handakumbura PP, Stanfill B, Rivas-Ubach A, Fortin D, Vogel JP, Jansson C. Metabotyping as a Stopover in Genome-to-Phenome Mapping. *Sci Rep*. 2019 Feb 12;9(1):1858.

He, Y., Caporaso, J., Jiang, X., et al. Stability of operational taxonomic units: an important but neglected property for analyzing microbial diversity. *Microbiome*, 2015, 3 (34).

Hillesheim E, Brennan L. Metabotyping and its role in nutrition research. *Nutr Res Rev*. 2020 Jun;33(1): 33- 42.

Hillesheim E, Ryan MF, Gibney E, Roche HM, Brennan L. Optimization of a metabotype approach to deliver targeted dietary advice. *Nutr Metab (Lond)*. 2020 Sep 29;17:82.

Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. *Philos Trans A Math Phys Eng Sci*. 2016 Apr 13;374(2065):20150202.

Li CX, Wheelock CE, Sköld CM, Wheelock ÅM. Integration of multi-omics datasets enables molecular classification of COPD. *Eur Respir J*. 2018 May 10;51(5):1701930.

Maitre L, Bustamante M, Hernández-Ferrer C, Thiel D, Lau CE, Siskos AP, Vives-Usano M, Ruiz-Arenas C, Pelegrí-Sisó D, Robinson O, Mason D, Wright J, Cadiou S, Slama R, Heude B, Casas M, Sunyer J, Papadopoulou EZ, Gutzkow KB, Andrusaityte S, Grazuleviciene R, Vafeiadi M, Chatzi L, Sakhi AK, Thomsen C, Tamayo I, Nieuwenhuijsen M, Urquiza J,

Borràs E, Sabidó E, Quintela I, Carracedo Á, Estivill X, Coen M, González JR, Keun HC, Vrijheid M. Multi-omics signatures of the human early life exposome. *Nat Commun.* 2022 Nov 21;13(1):7024.

Martínez S, Zegers Y, Stockins B, Bustos L, Sanhueza A, Rivera A, Soto L, Mackay A, Vega D, Rapimán P, Atton R, Alberti G. Evaluación de una maniobra nutricional tendiente a reducir los niveles de colesterol en pacientes portadores de enfermedad coronaria en el sistema público de salud chileno [Evaluation of a nutritional intervention to reduce cholesterol levels in patients with coronary artery disease]. *Rev Med Chil.* 2004 Dec;132(12):1457-65.

Moyne O, Castelli F, Bicout DJ, Boccard J, Camara B, Cournoyer B, Faudry E, Terrier S, Hannani D, Huot-Marchand S, Léger C, Maurin M, Ngo TD, Plazy C, Quinn RA, Attree I, Fenaille F, Toussaint B, Le Gouëllec A. Metabotypes of *Pseudomonas aeruginosa* Correlate with Antibiotic Resistance, Virulence and Clinical Outcome in Cystic Fibrosis Chronic Infections. *Metabolites.* 2021 Jan 21;11(2):63.

Noerman S, Kolehmainen M, Hanhineva K. Profiling of Endogenous and Gut Microbial Metabolites to Indicate Metabotype-Specific Dietary Responses: A Systematic Review. *Adv Nutr.* 2020 Sep 1;11(5):1237-1254.

Palmnäs M, Brunius C, Shi L, Rostgaard-Hansen A, Torres NE, González-Domínguez R, Zamora-Ros R, Ye YL, Halkjær J, Tjønneland A, Riccardi G, Giacco R, Costabile G, Vetrani C, Nielsen J, Andres-Lacueva C, Landberg R. Perspective: Metabotyping-A Potential Personalized Nutrition Strategy for Precision Prevention of Cardiometabolic Disease. *Adv Nutr.* 2020 May 1;11(3):524-532.

Petersen AK, Zeilinger S, Kastenmüller G, Römisch-Margl W, Brugger M, Peters A, Meisinger C, Strauch K, Hengstenberg C, Pagel P, Huber F,

Mohney RP, Grallert H, Illig T, Adamski J, Waldenberger M, Gieger C, Suhre K. Epigenetics meets metabolomics: an epigenome-wide association study with blood serum metabolic traits. *Hum Mol Genet.* 2014 Jan 15;23(2):534-45.

Rencher, & Christensen, W. F. (2012). *Methods of multivariate analysis* Alvin C. Rencher, William F. Christensen. (3rd ed.). Wiley.

Riedl A, Gieger C, Hauner H, Daniel H, Linseisen J. Metabotyping and its application in targeted nutrition: An overview. *Br J Nutr* 2017;117:1631–44

Riedl A, Wawro N, Gieger C, Meisinger C, Peters A, Roden M, Kronenberg F, Herder C, Rathmann W, Völzke H, et al. Identification of comprehensive metabolotypes associated with cardiometabolic diseases in the population-based KORA study. *Mol Nutr Food Res* 2018; 62:e1800117.

Romo-Vaquero M, Cortes-Martin A, Loria-Kohen V, Ramirez-de-Molina A, Garcia-Mantrana I, Collado MC, Espin JC, Selma MV. Deciphering the human gut microbiome of urolithin metabolotypes: Association with enterotypes and potential cardiometabolic health implications. *Mol Nutr Food Res* 2019; 63:e1800958

Ruiz Esparza-Garrido R, Velázquez-Flores MA, Arenas-Aranda DJ, Salamanca-Gómez F. La genómica en la medicina [Genomics in medicine]. *Rev Med Inst Mex Seguro Soc.* 2014 Sep-Oct;52(5):566-73.

Stephens ZD, Lee SY, Faghri F, Campbell RH, Zhai C, Efron MJ, et al. (2015) Big Data: Astronomical or Genomical? *PLoS Biol* 13(7): e1002195.

Urpi-Sarda M, Almanza-Aguilera E, Llorach R, Vázquez-Fresno R, Estruch R, Corella D, Sorli JV, Carmona F, Sanchez-Pla A, Salas-Salvadó J, Andres-Lacueva C. Non-targeted metabolomic biomarkers and metabolotypes of type 2 diabetes: A cross-sectional study of

PREDIMED trial participants. *Diabetes Metab.* 2019 Apr;45(2):167-174.

Wang B, Mezlini AM, Demir F, Fiume M, Tu Z, Brudno M, Haibe-Kains B, Goldenberg A. Similarity network fusion for aggregating data types on a genomic scale. *Nat Methods.* 2014 Mar;11(3):333-7.

Who.int. 2020. *Las 10 principales causas de defunción.* [online] Available at: <https://www.who.int/es/news-room/fact-sheets/detail/the-top-10-causes-of-death> [18 October 2022].

Zhang Z, Murtagh F, Van Poucke S, Lin S, Lan P. Hierarchical cluster analysis in clinical research with heterogeneous study population: highlighting its visualization with R. *Ann Transl Med.* 2017 Feb;5(4):75.

6. Anexos

El código utilizado para el análisis se encuentra en el siguiente repositorio de GitHub disponible para su consulta:

https://github.com/clauarce/TFM_Claudia_Arce