



UNIVERSIDAD NACIONAL DE COLOMBIA

# Modelado computacional de astrocito humano usando datos de transcriptómica y proteómica.

Nicolás Mendoza Mejía

Universidad Nacional de Colombia  
Facultad de ingeniería, Departamento de Ingeniería de Sistemas e Industrial.  
Bogotá D.C., Colombia  
2022



# Modelado computacional de astrocito humano usando datos de transcriptómica y proteómica.

Nicolás Mendoza Mejía

Tesis o trabajo de grado presentada(o) como requisito parcial para optar al título de:  
**Magister en Bioinformática**

Director(a):

Ph.D. Andrés Mauricio Pinzón Velasco

Codirector (a):

Ph.D. Janet González Santos

Línea de Investigación:

Biología de sistemas.

Grupo de Investigación:

Grupo de Investigación en Bioinformática y Biología de Sistemas Computacional

Universidad Nacional de Colombia

Facultad de ingeniería, Departamento de Ingeniería de Sistemas e Industrial.

Bogotá D.C., Colombia

2022



# Declaración de obra original

Yo declaro lo siguiente:

He leído el Acuerdo 035 de 2003 del Consejo Académico de la Universidad Nacional. «Reglamento sobre propiedad intelectual» y la Normatividad Nacional relacionada al respeto de los derechos de autor. Esta disertación representa mi trabajo original, excepto donde he reconocido las ideas, las palabras, o materiales de otros autores.

Cuando se han presentado ideas o palabras de otros autores en esta disertación, he realizado su respectivo reconocimiento aplicando correctamente los esquemas de citas y referencias bibliográficas en el estilo requerido.

He obtenido el permiso del autor o editor para incluir cualquier material con derechos de autor (por ejemplo, tablas, figuras, instrumentos de encuesta o grandes porciones de texto).

Por último, he sometido esta disertación a la herramienta de integridad académica, definida por la universidad.



---

Nombre: Nicolás Mendoza Mejía

Fecha 22/02/2022

# Agradecimientos

Quiero extender un agradecimiento a todas aquellas personas que con su ayuda, consejos y aportes contribuyeron a la culminación de este proyecto. En especial gracias a mi director Andrés Pinzón y a mi codirectora Janet González por su guía y por permitirme participar en sus grupos de investigación, así como a mis padres, mi hermano y familia por su constante apoyo, al igual que a mi novia por su paciencia y soporte.

Para cerrar gracias a todos los que me han ayudado en mi búsqueda del conocimiento y en mi crecimiento tanto profesional como personal.

# Resumen

## **Modelado computacional de astrocito humano usando datos de transcriptómica y proteómica.**

Aunque la astrogliosis está relacionada con eventos neuroprotectores; su neurotoxicidad se ha correlacionado con enfermedades neurodegenerativas y otros desórdenes (Sofroniew & Vinters, 2010). Lo que ha aumentado la atención en el estudio de estas células. Sin embargo, los procesos de señalización y actividad metabólica relacionados con la neurotoxicidad aún son poco conocidos (González et al., 2020; Sofroniew, 2015), por lo que se han empleado modelos metabólicos a escala genómica (GEM) de astrocito para estudiar estas respuestas.

Por lo tanto, en este trabajo se hace la contextualización de un GEM de astrocito humano integrando datos multiómicos con una nueva aproximación en combinación con el algoritmo iMAT, permitiendo incluir información de diversos procesos biológicos en el modelo (Karahalil, 2016; Vivek-Ananth & Samal, 2016). En consecuencia, el GEM resultante presenta una mayor cobertura del metabolismo y una capacidad predictiva superior en los escenarios simulados coincidiendo con lo reportado en la literatura. Además, durante la reconstrucción de este modelo se generaron dos algoritmos, uno permite integrar el proteoma y transcriptoma, mientras el otro corrige los desbalances estequiométricos presentes en el modelo. Finalmente, este modelo tiene el potencial de acelerar el estudio de la astrogliosis, permitiendo descifrar la relación entre el metabolismo del astrocito y la aparición de enfermedades neurodegenerativas mediante la generación de hipótesis y la predicción del desempeño de fármacos.

**Palabras clave:** Modelo a escala genómica, astrocito, datos multi-ómicos, enfermedades neurodegenerativas, transcriptómica y proteómica..

# Abstract

## **Computational modeling of the human astrocyte using transcriptomic and proteomic data.**

Even though astrogliosis is related to neuroprotective events; its neurotoxicity has been correlated with neurodegenerative diseases and other disorders (Sofroniew & Vinters, 2010). Which have shifted the attention towards the study of these cells. However, the related signaling processes and metabolic activity related to the neurotoxicity are still poorly known (González et al., 2020; Sofroniew, 2015), thus genome-scale metabolic models (GEMs) of astrocytes have been used to study this response, as they allow modelling metabolic interactions (Martín-Jiménez et al., 2017; Osorio et al., 2020).

Therefore, in this work an astrocyte's GEM is contextualized by integrating multi-omic data with a new approach in combination with the algorithm iMAT, which allows including information from various biological processes in the model (Karahalil, 2016; Vivek-Ananth & Samal, 2016). Thus, the resulting GEM presets a greater coverage of the metabolic network and a higher predictive capability in the simulated scenarios, which is in line with the reported data in the literature. In addition, during the reconstruction of this model two algorithms were generated, one integrates the proteome and transcriptome together, meanwhile the other corrects the stoichiometric imbalances present in the model. Finally, this model has the potential to accelerate the study of astrogliosis allowing to decipher the relationship between astrocyte metabolism and the appearance of neurodegenerative diseases by generating hypotheses and predicting drug performance.

**Keywords:** Genome-scale metabolic model, astrocyte, lipotoxicity, neurodegenerative diseases, transcriptomics, proteomics.





Esta tesis de maestría se sustentó el 20 de mayo de 2022 a las 10:00 am,  
y fue evaluada por los siguientes jurados:

Ricardo Julián Cabezas Pérez Ph.D.  
Facultad de medicina  
Universidad Antonio Nariño

Juan David Garcia Arteaga Ph.D.  
Facultad de medicina  
Universidad Nacional de Colombia

# Contenido

<b>Declaración de obra original</b>	<b>v</b>
<b>Agradecimientos</b>	<b>vi</b>
<b>Resumen</b>	<b>vii</b>
<b>Lista de figuras</b>	<b>xii</b>
<b>Lista de tablas</b>	<b>xiii</b>
<b>1. Introducción</b>	<b>2</b>
1.0.1. Justificación . . . . .	3
1.0.2. Planteamiento del problema . . . . .	5
1.0.3. Objetivos . . . . .	8
<b>2. Capítulo 1: Marco teórico</b>	<b>9</b>
2.1. Procesos inflamatorios y los astrocitos . . . . .	9
2.2. Generalidades de los GEM . . . . .	11
2.2.1. Anotación del genoma . . . . .	12
2.2.2. Simulación de flujo . . . . .	13
2.2.3. Función objetivo . . . . .	14
2.2.4. Ambiente nutricional . . . . .	14
2.2.5. Curación de modelo . . . . .	15
2.3. Integración de datos ómicos en modelos metabólicos a escala genómica . . . . .	16
2.3.1. Reconstrucción de modelos específicos de tejido . . . . .	18
2.3.2. GEM contexto específicos . . . . .	30
<b>3. Capítulo 2: Métodos</b>	<b>34</b>
3.1. Obtención de datos de transcriptómica y proteómica . . . . .	36
3.1.1. Cultivo celular . . . . .	36
3.1.2. Tratamientos . . . . .	36
3.1.3. Datos de transcriptoma . . . . .	36
3.1.4. Datos del proteoma . . . . .	37
3.2. Estrategia de integración de datos ómicos . . . . .	39
3.2.1. Definición de la estrategia de contextualización . . . . .	39

---

3.2.2.	Mapeo de los datos ómicos al modelo . . . . .	40
3.2.3.	Reducción a una dimensión . . . . .	40
3.3.	Contextualización del modelo con datos de transcriptómica y proteómica. . .	41
3.3.1.	Curación del modelo . . . . .	42
3.3.2.	Llenado de huecos . . . . .	42
3.3.3.	Consistencia estequiométrica . . . . .	42
3.3.4.	Ambiente nutricional . . . . .	43
3.4.	Escenarios metabólicos . . . . .	44
<b>4.</b>	<b>Capítulo 3: Resultados</b>	<b>45</b>
4.1.	Integración del proteoma y transcriptoma . . . . .	45
4.1.1.	Modelo tejido específico de astrocito . . . . .	48
4.1.2.	Escenarios metabólicos . . . . .	51
<b>5.</b>	<b>Capítulo 4: Discusión</b>	<b>56</b>
<b>6.</b>	<b>Capítulo 5: Conclusiones</b>	<b>58</b>
<b>A.</b>	<b>Anexo: Composición DMEM</b>	<b>59</b>
	<b>Bibliografía</b>	<b>60</b>

# Lista de Figuras

1-1. Resumen de las capas celulares y las tecnologías ómicas para analizarlas. . .	6
1-2. Árbol del problema. . . . .	8
2-1. Expresión de GFAP en astrogliosis. . . . .	10
2-2. Espacio de solución en FBA. . . . .	13
2-3. Flujo de trabajo para el algoritmo GIMME. . . . .	19
2-4. El esquema del algoritmo PRIME. . . . .	21
2-5. Proceso de integración de datos ómicos. . . . .	28
2-6. Diagrama de Venn para los GEM de humano. . . . .	31
2-7. Línea de tiempo GEMs humano y astrocito . . . . .	33
3-1. Diagrama de las fases de la investigación para el modelado computacional del astrocito humano. . . . .	35
4-1. Reacciones con reglas en Recon3D . . . . .	46
4-2. Distribución de los datos mapeados . . . . .	47
4-3. Distribución de las enzimas incluidas en el modelo de astrocito . . . . .	48
4-4. Diagrama de venn de las interacciones entre los metabolitos y reacciones de los diferentes modelos. . . . .	51
4-5. Análisis de robustez para el escenario de isquemia. . . . .	53

# Lista de Tablas

<b>2-1.</b> Algoritmos para contextualización de modelos. Esta es una tabla comparativa entre los algoritmos revisados en este documento, mostrando el tipo de dato que soporta cada uno y cuando fue publicado. . . . .	18
<b>2-2.</b> Cuadro de decisión para el uso de algoritmos (E-Flux2 y SPOT) y plantillas propuestas (DC y AC) fuente: (M. K. Kim et al., 2016). . . . .	22
<b>2-3.</b> Tipos de métodos REMI. Los tres métodos que tiene REMI según los tipos datos ómicos que se vayan a usar. . . . .	29
<b>3-1.</b> Composición del buffer de lisis. Componentes utilizados para preparar el buffer para la extracción de las proteínas y sus respectivas concentraciones y cantidades. . . . .	38
<b>4-1.</b> Composición química del medio de entrada nutricional del modelo de acuerdo a (Tejera et al., 2020; Yang & Xiong, 2012; Branzoi et al., 2010). . . . .	50
<b>4-2.</b> Resultados de las funciones objetivo en estado sano, inflamado con palmitato y un pre-tratamiento con tibolona. Para cada una de las 6 funciones objetivo de los astrocitos planteadas por Osorio y colaboradores (2020). Donde, biomasa: indica el crecimiento, Glu2Gln: la conversión de glutamato a glutamina, Gly2SerD: la conversión de Glycina a D-serina, Glc2ATP: la producción de ATP a partir de glucosa y r0130: la obtención de cysteina para producir glutatión reducido. . . . .	54

# 1. Introducción

Los **astrocitos** son células gliales especializadas en el mantenimiento de la función, salud y unión del sistema nervioso central (SNC), siendo indispensables para la supervivencia neuronal (Phatnani & Maniatis, 2015). Hasta ahora, se sabe que estas células cumplen con funciones principales como: (i) la regulación de la homeóstasis en el SNC (Takuma et al., 2004), (ii) la reparación de tejidos ante lesiones, desórdenes metabólicos e infecciones (Osorio et al., 2020; Sofroniew, 2015) y (iii) la modulación de la actividad sináptica (Lange et al., 2012). Recientemente la función más estudiada de estas células es la astrogliosis o gliosis reactiva (Sofroniew & Vinters, 2010).

La **astrogliosis** es una respuesta a las lesiones en el SNC, compuesta de cambios que van desde hipertrofia hasta proliferación y mitigación (Phatnani & Maniatis, 2015), estos se han visto relacionados con eventos tanto neuroprotectores como neurotóxicos. Y aunque los procesos de señalización y actividad metabólica de la astrogliosis son poco conocidos (González et al., 2020; Sofroniew, 2015), los eventos neurotóxicos, presentan una correlación con enfermedades neurodegenerativas y otros desórdenes, por ejemplo, el aumento de ácido palmítico a demostrado producir una respuesta inflamatoria asociada generalmente con procesos neurodegenerativos, como: demencia, epilepsia, enfermedad de Parkinson y Alzheimer, etc. (Jha et al., 2016; Osorio et al., 2020; Phatnani & Maniatis, 2015; Takuma et al., 2004).

En los últimos años, ha habido un aumento en el uso de **técnicas ómicas** para el estudio de enfermedades neurodegenerativas, estos acercamientos permiten detectar las moléculas que conforman una célula, tejido u organismo, identificando relaciones estadísticas entre la composición de estos y la degeneración neuronal (Diaz-Ortiz and Chen-Plotkin, 2020). Además, los datos ómicos generan una imagen holística del organismo, algo que es muy usado en la biología de sistemas para el estudio de los complejos mecanismos biológicos (Horgan & Kenny, 2011). Los recientes avances de estas técnicas, sumado al aumento de la disponibilidad de los datos ómicos que generan, llevó a que la biología de sistemas se enfocara en métodos como la reconstrucción de **modelos metabólicos a escala genómica** (GEMs, por sus siglas en inglés) (Vivek-Ananth & Samal, 2016), que son modelos computacionales que permiten comprender los procesos de expresión de los fenotipos metabólicos, planteándose como una alternativa a las aproximaciones reduccionistas propias de las técnicas tradicionales (Karahalil, 2016).

---

Vale la pena aclarar que el objetivo del modelado computacional de estas redes metabólicas es predecir las respuestas que tendrá el organismo de interés ante uno o varios contextos para adaptarse al ambiente en el que se encuentra (Gu et al., 2019). En este sentido, un tema de vanguardia para mejorar dicha predictibilidad, así como para la mejor contextualización biológica de los modelos, es la integración de datos ómicos en los GEMs (Gu et al., 2019; Vivek-Ananth & Samal, 2016), pues al basarse en datos experimentales las predicciones generadas son más cercanas a la realidad, permitiendo modelar los resultados metabólicos que se producen a partir de la interacción de todos los niveles celulares (Karahalil, 2016; Vivek-Ananth & Samal, 2016). Facilitando el análisis e interpretación de datos como: bases de datos biológicas, datos ómicos de alto rendimiento y evidencia experimental (Gu et al., 2019).

Aunque nuestro grupo de investigación ha reconstruido los dos únicos GEMs de astrocito disponibles en la literatura, estos carecen de una integración de datos ómicos (Martín-Jiménez et al., 2017; Osorio et al., 2020), por lo tanto, en este documento se propone una integración de datos multiómicos en un GEM específico de astrocito, al aumentar la capacidad predictiva del modelo respecto a los existentes, promover el descubrimiento biológico y mejorar la generación de hipótesis (González et al., 2020; Gu et al., 2019; Vivek-Ananth & Samal, 2016).

### **1.0.1. Justificación**

Las enfermedades neurodegenerativas imponen una alta carga en el sistema de salud mundial (Checkoway et al., 2011), ocupando el segundo puesto en causa de muerte y el primero en causa de discapacidad, además, tienen una tendencia a la alta porque su prevalencia aumenta con la edad, al igual que la esperanza de vida a nivel mundial (World Health Organization, 2006). A la vez, las causas genéticas o ambientales que promueven estas enfermedades son, con algunas excepciones, poco conocidas (Feigin et al., 2019).

Así mismo, estas enfermedades se han visto asociadas con los astrocitos en procesos inflamatorios, también llamados astrogliosis; pero, la definición y comprensión de estas células en los procesos inflamatorios aún se está empezando a dilucidar (Soung & Klein, 2019) y no es claro si los efectos neurodegenerativos son propagados a través de la inflamación, o al contrario, la inflamación refleja una respuesta neuroprotectora para evitar la propagación del daño celular (Layé, 2010). Además la astrogliosis tiene efectos negativos inmediatos, provocando cambios en el comportamiento en animales y humanos, como: dificultades en la atención, concentración y memoria; siendo responsables del deterioro de la calidad de vida y bienestar de los individuos que padecen estas alteraciones, esto se debe a la secreción de citocinas por parte de los astrocitos durante estos procesos (Layé, 2010).

Además, los modelos animales genéticamente modificados para estudiar las enfermedades



neurodegenerativas, son incapaces de reproducir completamente los eventos fisiopatológicos que se observan en los humanos, en consecuencia, los modelos de ratón, tienen un bajo poder predictivo respecto a la eficacia de drogas (Dawson et al., 2018; Noble & Burns, 2010). Por otro lado, los modelos celulares, que son la alternativa a los modelos animales, carecen de características importantes como: complejos circuitos neuronales, un compartimento glial completo y otros componentes vasculares e inmunológicos. Sin mencionar que estos modelos parten de células anormales que tienen una mayor viabilidad *in-vitro* que las que se encuentran en un tejido sano. Consecuentemente, los tratamientos exitosos en modelos animales y celulares, en su gran mayoría, han fallado en pruebas clínicas (Dawson et al., 2018).

Se cree que parte del problema, es que las pruebas pre-clínicas no logran tener en cuenta el sinnúmero de factores que influyen en el desarrollo de estas enfermedades. Lo que lleva a pensar sobre el factor ético: si los modelos *in-vivo* e *in-vitro* han mostrado ser insuficientes para determinar si un tratamiento tiene posibilidades de éxito en los humanos ¿que medidas adicionales se pueden tomar para evitar riesgos innecesarios?, y es aquí donde surgen los GEM como una alternativa que cumple con dos (reemplazar y reducir) de las “tres R” propuestas como principios éticos para la experimentación *in-vivo* por Russell y Burch (1959). Siendo que los modelos funcionan como una alternativa que tiene una visión holística de escenarios multidimensionales, como los que caracterizan a estas enfermedades, los GEM pueden ser usados como una plataforma de prueba de hipótesis e incluso mediante análisis matemáticos avanzados, se pueden identificar mecanismos celulares, que funcionen como una base para el desarrollo de tratamientos más eficaces.

Adicionalmente, el análisis de las distintas capas biológicas, a través de datos multi-ómicos para producir información biológica relevante, es un reto de gran magnitud debido a su heterogeneidad (Haas et al., 2017), por lo que los GEM surgen como una plataforma de análisis que permite asociar fácilmente datos de diversas fuentes, logrando incluso contextualizar los datos experimentales con el conocimiento disponible en literatura.

Por esta razón, se han propuesto aproximaciones desde la biología de sistemas para suplir las falencias en el conocimiento mencionadas. De esta manera los GEMs se plantean como una herramienta indispensable para entender fenómenos complejos como las enfermedades neurodegenerativas y procesos inflamatorios (González et al., 2020; Osorio et al., 2020). Mientras que el uso de datos ómicos, ha permitido demostrar que los astrocitos son tan diferenciados como las neuronas, además de identificar algunas de sus funciones durante la astrogliosis (Soung & Klein, 2019). En este sentido, estas aproximaciones se presentan como una alternativa que permite adquirir una imagen holística de la astrogliosis ante diferentes lesiones en el SNC (Bleu Knight & Serrano, 2017; Karahalil, 2016; Zuo et al., 2017).

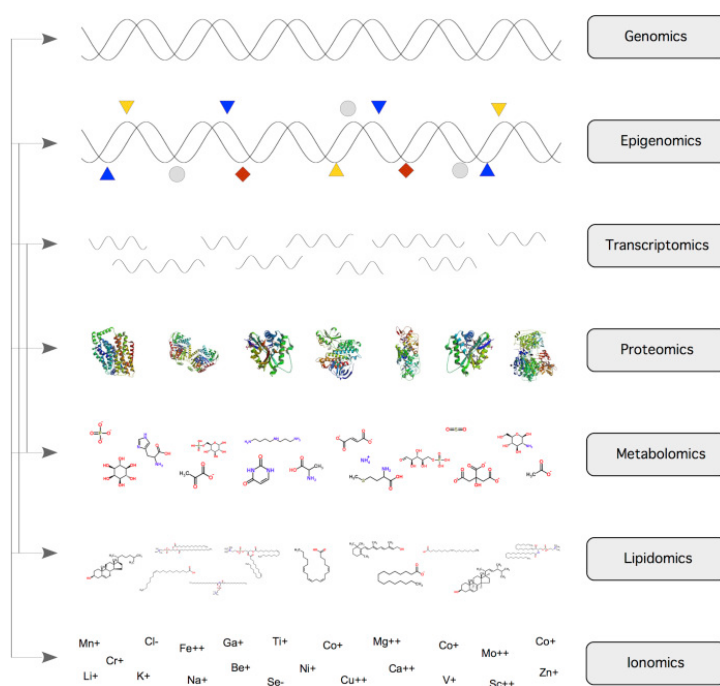
## 1.0.2. Planteamiento del problema

La integración de datos ómicos en los GEMs es un tema de vanguardia en la biología de sistemas (Vivek-Ananth & Samal, 2016), combinando lo mejor de las dos aproximaciones mencionadas anteriormente y obteniendo como resultado, una imagen holística del sistema contextualizada con datos ómicos reales, lo que permite realizar predicciones más exactas y concordantes con el conocimiento biológico actual (Gu et al., 2019; Karahalil, 2016; M. Kim et al., 2016). Lo anterior, es una herramienta clave para examinar perfiles transcripcionales, post-transcripcionales y epigenéticos (González et al., 2020), facilitando el estudio de las respuestas de los astrocitos ante lesiones en el SNC.

Por su parte, la baja disponibilidad de datos ómicos, contribuye al uso de aproximaciones reduccionistas propias de la biología clásica en la mayoría de los estudios en la literatura, dificultando la comprensión de la astrogliosis (Karahalil, 2016; Phatnani & Maniatis, 2015; Soung & Klein, 2019). Aun así, el aumento en la calidad y rendimiento de las técnicas ómicas en los últimos años, ha permitido avanzar en la comprensión de los mecanismos inflamatorios de los astrocitos mediante el uso de datos de transcriptómica y proteómica (Bleu Knight & Serrano, 2017; Karahalil, 2016; Oliveira et al., 2018; Soung & Klein, 2019; Zuo et al., 2017).

No obstante, estos datos ómicos no solo presentan un reto para el análisis por su complejidad y tamaño (Karahalil, 2016; Stephens et al., 2015), sino que, el uso de una sola fuente ómica (datos uni-ómicos) no provee la información suficiente para predecir el comportamiento de un organismo, ya que ignora las interacciones entre las distintas capas de resolución biológica (figura 1-1), porque en los mecanismos biológicos participa más de un tipo de biomolécula (Haas et al., 2017; Karahalil, 2016). Como alternativa, se han planteado uniones de varios datos uni-ómicos para generar un compendio que haga las veces de unos datos multi-ómicos, pero, los experimentos realizados por diferentes laboratorios no pueden ser combinados sin algunas restricciones, siendo entonces un reto complejo producir un resultado adecuado a partir de varias fuentes uni-ómicas (Haas et al., 2017).

Aparte de estos retos en la integración de datos uni-ómicos, en caso de tener los datos multi-ómicos desde el nivel experimental, el análisis de estos para producir información biológica nueva, es un reto de gran magnitud en sí mismo, porque los datos biológicos son heterogéneos, no sólo por motivos técnicos, sino también por razones biológicas, químicas y físicas (Haas et al., 2017).



**Figura 1-1.:** Resumen de las capas celulares y las tecnologías ómicas para analizarlas.

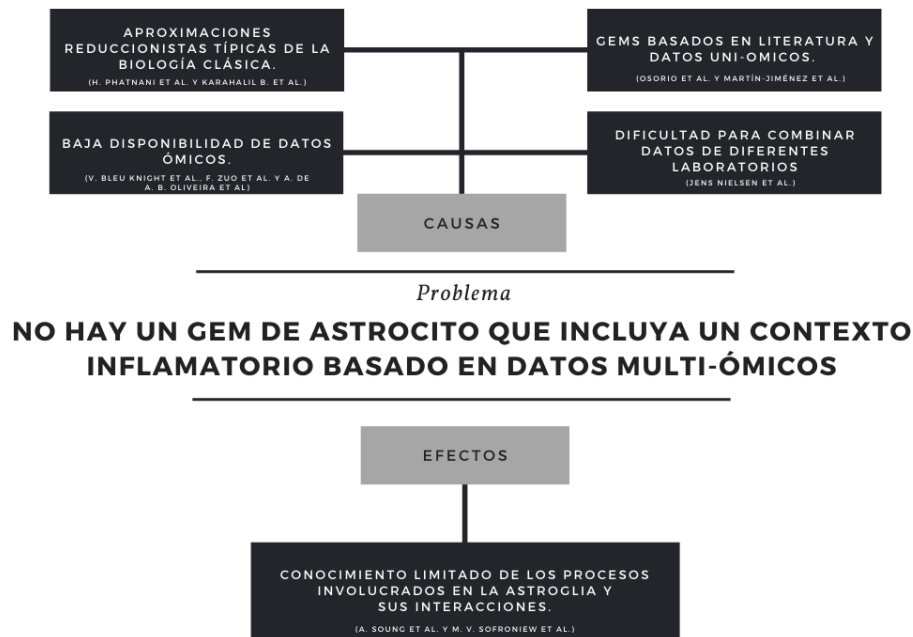
Las capas consisten de ADN (Genoma), ARN (transcriptoma), proteínas (proteoma), pequeñas moléculas (metaboloma y lipidoma) y composición elemental (medida como ionoma). Las fechas de la izquierda indican el grado de dependencia entre los diferentes niveles. (Haas et al., 2017).

Entonces, se ha propuesto el uso de GEM, con el propósito de representar las redes metabólicas en un modelo matemático, lo que permite predecir las respuestas del organismo modelado ante condiciones patológicas o estímulos específicos (González et al., 2020). Usando esta aproximación, se han desarrollado dos GEMs específicos de astrocito humano, basados en datos transcriptómicos, para estudiar el comportamiento de los astrocitos en condiciones de lesión y sana (normal) (González et al., 2020; Martín-Jiménez et al., 2017; Osorio et al., 2020). Sin embargo, al estar basados en datos uni-ómicos presentan las mismas carencias de información mencionadas anteriormente; además, no tienen un contexto inflamatorio basado en datos experimentales que asegure la presencia de las rutas metabólicas activadas en estos procesos, como por ejemplo, el uso de datos ómicos de astrocitos con una respuesta inflamatoria provocada ácido palmítico, que produce una respuesta inflamatoria en los astrocitos (Osorio et al., 2020).

En consecuencia, el problema es que no existe un GEM de astrocito que incluya un contexto inflamatorio basado en datos multi-ómicos, que al posicionarse como una nueva herramienta para los investigadores, permita una aproximación holística a la complejidad de los astrocitos y sus procesos inflamatorios (Soung & Klein, 2019), facilitando avanzar en el conocimiento biológico actual sobre la astrogliia, mediante predicciones más precisas debido a la integración multi-ómica, que promuevan el descubrimiento de nuevo conocimiento de los mecanismos de la astrogliia en procesos neuro inflamatorios, siendo útil en el desarrollo de terapias médicas (González et al., 2020).

### **Pregunta de investigación**

¿Qué tanto mejora el poder de predicción de respuestas inflamatorias en los astrocitos humanos en un modelo computacional de metabolismo astrocitario integrando datos transcriptómicos y proteómicos, con respecto a modelos basados en una sola fuente de información ómica?



**Figura 1-2.:** Árbol del problema.

fuente: elaboración propia.

### 1.0.3. Objetivos

#### Objetivo general

Desarrollar un modelo computacional funcional de astrocito humano contextualizado con los mecanismos inflamatorios de los astrocitos, basado en datos de transcriptómica y proteómica.

#### Objetivos específicos

- Definir una estrategia de desarrollo del modelo computacional para la integración de los datos de transcriptómica y proteómica, mediante una revisión bibliográfica.
- Asegurar un mejor poder estadístico y menor variabilidad de las muestras mediante la estandarización de los datos ómicos con herramientas bioinformáticas, para la integración de estos en un modelo.
- Integrar los datos de transcriptómica y proteómica en un modelo computacional con la estrategia definida en la revisión bibliográfica para la representación computacional de la red metabólica del astrocito.
- Comprobar la funcionalidad del modelo desarrollado con herramientas de optimización de flujos para la evaluación de la predictibilidad en un contexto biológico.

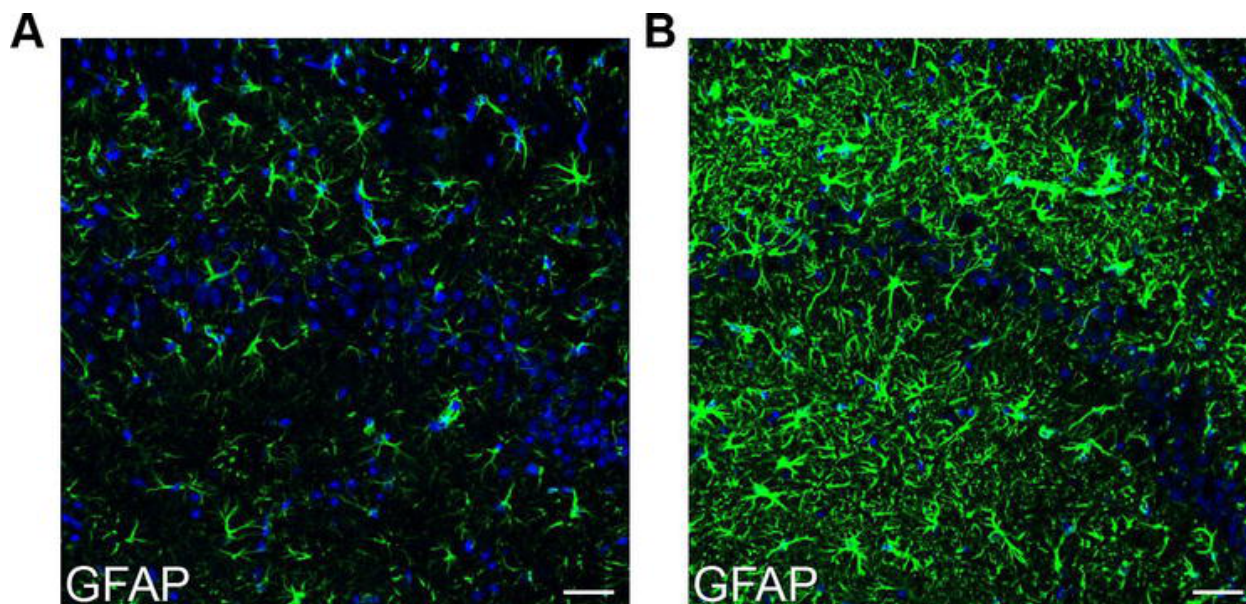
## 2. Capítulo 1: Marco teórico

### 2.1. Procesos inflamatorios y los astrocitos

Aunque por mucho tiempo se consideró al cerebro como un órgano inmunológicamente privilegiado, al estar supuestamente aislado del sistema inmune por la barrera hematoencefálica y por su presunta incapacidad para generar una respuesta inmune y procesar antígenos (Carson et al., 2006; Layé, 2010). El resurgimiento de análisis enfocados en las células gliales (Soung & Klein, 2019), han demostrado lo contrario, en realidad el cerebro tiene un sistema inmune innato altamente controlado, compuesto principalmente de células microgliales y astrocitos, que una vez activadas protegen las neuronas de lesiones, agentes infecciosos, entre otros ataques al SNC. (Layé, 2010; Soung & Klein, 2019).

La inflamación es una respuesta activa para defenderse ante lesiones, con el propósito de neutralizar agentes nocivos (Sofroniew, 2015; Soung & Klein, 2019), tanto la microglía como los astrocitos desempeñan un papel importante en esta respuesta inmunitaria (Bianchi et al., 2014; Pfefferkorn et al., 2016; Soung & Klein, 2019). Sin embargo, los astrocitos también muestran una función coordinadora y central en los mecanismos de inflamación, recibiendo y transmitiendo información de todos los tipos de células presentes en el SNC, siendo capaces de detectar señales de infecciones y lesiones en las neuronas, microglía, oligodendrocitos y células endoteliales, respondiendo a estos escenarios mediante la secreción de citocinas, quimiocinas y factores de crecimiento, que pueden a su vez activar defensas inmunitarias (Soung & Klein, 2019).

Estas citocinas son péptidos que están relacionados con la inflamación, activación del sistema inmune y diferenciación celular o muerte. Dentro de estos péptidos se incluyen las interleuquinas (IL), interferones (IFN), factores de necrosis tumoral (TFN), quimiocinas y factores de crecimiento (Hamby et al., 2012; Jensen et al., 2013); estas moléculas por su parte promueven la expresión de cientos de genes estimulados por interferón, como los que participan en la inflamación de células inflamatorias (Soung & Klein, 2019).



**Figura 2-1.:** Expresión de GFAP en astrogliosis.

La astrogliosis promueve la expresión de proteínas como la GFAP que se puede apreciar en la imagen con fluorescencia verde, (A) tejido normal y (B) astrogliosis, en el que se puede observar una sobre expresión de la proteína GFAP, fuente: (Soung & Klein, 2019).

Sin embargo, los astrocitos también expresan fenotipos heterogéneos y opuestos funcionalmente, algunos remueven moléculas tóxicas, restauran las funciones de la barrera hematoencefálica y promueven la matriz extracelular para soportar el crecimiento y reparación de axones, mientras otros inhiben la reparación y regeneración neuronal (Iglesias et al., 2017; Layé, 2010; Osorio et al., 2020; Phatnani & Maniatis, 2015; Sofroniew, 2015; Sofroniew & Vinters, 2010; Soung & Klein, 2019; Takuma et al., 2004).

En este sentido, es importante resaltar el papel de los astrocitos en la regulación de la respuesta inmune. Estas células pueden responder a lesiones en el SNC, mediante la alteración de la función o supervivencia neuronal, en donde, las lesiones neuronales resultan en la secreción de varias señales como: el aumento de la producción de glutamato, liberación de ATP y daño vascular (Soung & Klein, 2019). Un ejemplo, es la liberación de glutamato por parte de las neuronas como respuesta ante enfermedades como: un golpe, lesión traumática, epilepsia, infección viral y neurodegeneración (Goodrich et al., 2013). En estos casos, el glutamato liberado es dañino para las neuronas adyacentes, es aquí, donde se ha demostrado que los astrocitos toman el exceso de glutamato extracelular para amortiguar el efecto neurotóxico de estos neurotransmisores disminuyendo entonces la muerte neuronal (Soung & Klein, 2019).

Por otro lado, como ya se ha mencionado, los astrocitos también pueden llegar a tener efectos neurotóxicos ante un desbalance en el proceso inflamatorio (Lange et al., 2012; Osorio et

al., 2020; Phatnani & Maniatis, 2015; Soung & Klein, 2019; Takuma et al., 2004), en donde la inflamación mencionada anteriormente, puede descontrolar la toma del glutamato por parte de los astrocitos, lo que aumenta la toxicidad neuronal y produce una realimentación positiva para la neuroinflamación, esto se evidencia en estudios *in vivo* donde la microglia libera TNF para indicarle a los astrocitos que liberen glutamato (Bezzi et al., 2001; Soung & Klein, 2019), en casos de muerte neuronal, la concentración de ATP extracelular puede aumentar, provocando que los astrocitos liberen glutamato y que se active el inflamasoma (un mecanismo de inmunidad innata que produce citocinas proinflamatorias), contribuyendo en resumen a la toxicidad extracelular (Soung & Klein, 2019).

La expresión de estas citoquinas proinflamatorias tienen un rol fundamental en la patofisiología de desórdenes neurológicos inmunes y no-inmunes, depresión y enfermedades neurodegenerativas asociadas a la edad como el Alzheimer, Parkinson, Huntington y Esclerosis lateral amiotrófica, entre otras (Jensen et al., 2013; Layé, 2010; Phatnani & Maniatis, 2015; Jha et al., 2016).

## 2.2. Generalidades de los GEM

Los modelos metabólicos a escala genómica, también llamados modelos a escala genómica, cuyas siglas en inglés varían en la literatura entre GEM por *Genome-Scale Models* y GSMM por *Genome-Scale Metabolic Model*. Son modelos reconstruidos a partir del conocimiento curado y sistematizado disponible en bases de datos biológicas, que representan a nivel sistémico todas las funciones metabólicas de una célula, lo que les permite describir las relaciones entre el genotipo y el fenotipo (Bernstein et al., 2021).

Los GEM son una representación matemática de las complejas redes metabólicas celulares en una matriz estequiométrica, funcionando como un marco de referencia que permite asociar el conocimiento específico de un organismo y los datos ómicos complejos, con la red metabólica representada en el GEM, lo que permite un estudio más detallado de esta información con una comprensión holística de su papel en el organismo (Bernstein et al., 2021). Adicionalmente, al acoplarse con los métodos de reconstrucción y análisis basados en restricciones o COBRA por sus siglas en inglés (CONstraint-Based Reconstruction and Analysis), como por ejemplo el **análisis de balance de flujos** (FBA), se pueden hacer análisis matemáticos sofisticados del metabolismo en todos los niveles celulares, facilitando la transición de hipótesis en algoritmos que generan predicciones de fenotipos metabólicos multidimensionales a través de la optimización de una **función objetivo** (Bernstein et al., 2021; Fang et al., 2020).



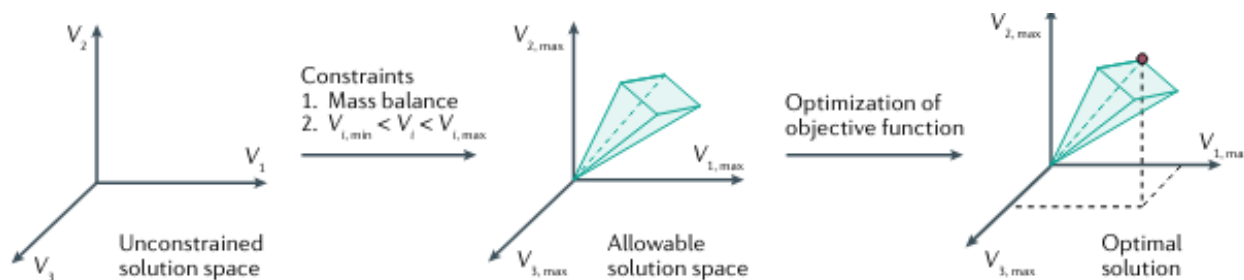
Por lo tanto, como estos modelos computacionales permiten interpretar y predecir estados fenotípicos y las consecuencias de perturbaciones genéticas y ambientales (Fang et al., 2020), los GEM tiene aplicaciones en el estudio de microorganismos, ingeniería metabólica, desarrollo de fármacos, predicción de funciones enzimáticas, en el estudio de las interacciones entre comunidades microbianas en ecología microbiana y para el estudio de enfermedades humanas (Bernstein et al., 2021; Fang et al., 2020).

### 2.2.1. Anotación del genoma

Como se ha mencionado, para la reconstrucción de un GEM se usa el conocimiento disponible en las bases de datos biológicas. Pero, específicamente cuando se reconstruye un GEM desde cero, se requieren dos fuentes: por un lado, el genoma anotado del organismo de interés, que permite identificar las enzimas que puede sintetizar el organismo. Y por el otro lado, la información disponible en bases de datos metabólicas curadas, como por ejemplo, Kyoto Encyclopedia of Genes and Genomes (KEGG), que contiene información de rutas metabólicas y las reacciones involucradas en estas. Al cruzar estas dos fuentes, asociando las enzimas codificadas en el genoma del organismo con las reacciones que estas catalizan, se obtiene una lista de asociaciones entre gen, proteína y reacción, conocidas como **reglas GPR**. Estas reglas buscan asociar los genes con las reacciones metabólicas. Estas reglas se valen de expresiones booleanas para codificar las asociaciones no lineales entre los genes y las reacciones, que ocurren en presencia de enzimas multiméricas, enzimas multifuncionales e isoenzimas (Bernstein et al., 2021).

Con esta información se logra reconstruir una red metabólica, que es representada matemáticamente en una matriz estequiométrica  $\mathbb{R}$  de  $m \times n$ , donde  $m$  representa los metabolitos y las reacciones que están presentes en el modelo, cada valor dentro de la matriz es el coeficiente del metabolito  $m$  en la reacción  $n$ .

No obstante, durante la reconstrucción de un GEM desde cero, se pueden introducir errores que requieran de una curación extensiva. Por ejemplo, debido a la baja precisión de los métodos de homología usados para la anotación funcional de las enzimas codificadas en el genoma. Así como, errores en las bases de datos metabólicas y el hecho de que varios genes sólo pueden ser anotados como secuencias hipotéticas con una función desconocida además de las enzimas huérfanas que no están asociadas a una secuencia genética. Por lo tanto, para reducir estas incertidumbres, se toman GEM previamente reconstruidos que ya pasaron por un proceso de curación como una plantilla de anotación. En donde asociar las reacciones presentes en el organismo de interés al organismo plantilla. Estos métodos obtienen un subgrupo de reacciones del modelo original para generar un modelo específico (Bernstein et al., 2021), como es el caso de este proyecto en el que se busca generar un GEM de astrocito a partir de uno de humano, los algoritmos para producir estos modelos con el uso



**Figura 2-2.:** Espacio de solución en FBA.

El gráfico muestra cómo se limita el modelo para obtener un espacio de solución consistente con las limitaciones termodinámicas. fuente: (Fang et al., 2020)

de datos ómicos se discute en el capítulo de **Integración de datos ómicos en modelos metabólicos a escala genómica**.

### 2.2.2. Simulación de flujo

Uno de los usos más importantes de los GEMs es la predicción de fenotipos metabólicos en un estado estacionario, que por lo general se calcula usando el *análisis de balance de flujos* o **FBA**, permitiendo caracterizar los GEMs mediante una simulación de flujos metabólicos en la red reconstruida. La matriz estequiométrica y la **función objetivo** definen un sistema de ecuaciones lineales con unas limitaciones impuestas que resulta en un espacio de solución (este es el espacio en el que todos los estados fenotípicos viables existen) FBA puede identificar una única distribución de flujo óptima o varias que optimicen la función objetivo y estén dentro de dicho espacio de soluciones (Fang et al., 2020).

Ya que la matriz estequiométrica es generalmente mayor que el número de reacciones involucradas en un flujo, el espacio de soluciones es demasiado grande, por lo que se incorporan una serie de restricciones que limitan el espacio de soluciones, para asegurar que la solución óptima identificada por FBA sea biológicamente relevante. Estas restricciones también aseguran que la predicción concuerda con: la topología de la red representada, la suposición de un estado estacionario, lo que quiere decir que los metabolitos deben ser producidos y consumidos de manera que se balancee el flujo, también se tiene en cuenta el medio nutricional, perfiles de expresión genética, así como capacidades y abundancias enzimáticas (Fang et al., 2020).

### 2.2.3. Función objetivo

Para el análisis de los GEM, generalmente se asume que el organismo (o célula) tiene un “objetivo” metabólico establecido, de acuerdo a su función biológica. A esta se le conoce como: “función objetivo”, y es comúnmente formulada en los GEMs como un inventario de todos los componentes esenciales para el crecimiento, supervivencia y mantenimiento de las funciones metabólicas de un organismo, representando las cantidades de cada metabolito en gramos de peso seco. Cuya definición es esencial para el cálculo de un flujo metabólico viable con herramientas como el FBA (Fang et al., 2020).

Por otro lado, la **formulación de biomasa** es la función objetivo más usada y está compuesta por los metabolitos esenciales para el crecimiento del organismo. Entonces, en estas funciones objetivo, se definen reacciones encargadas de transformar todos los componentes requeridos para las actividades metabólicas del organismo (Bernstein et al., 2021).

Sin embargo, vale la pena aclarar que si bien el crecimiento en Biomasa y/o el mantenimiento celular son las funciones objetivo más usadas en las simulaciones de flujo como FBA, es también factible utilizar otro tipo de objetivos metabólicos de acuerdo a la naturaleza biológica de la célula o del escenario metabólico bajo estudio.

### 2.2.4. Ambiente nutricional

Se ha demostrado que una representación precisa del medio nutricional es crucial para tener predicciones fisiológicas reproducibles y más precisas (Marinos et al., 2020), por lo que la transcripción de la información nutricional en una representación *in-silico* del ambiente químico, es importante para la predicción de fenotipos. El **ambiente nutricional** es representado en un GEM restringiendo los límites inferiores en las reacciones de intercambio, que, en consecuencia, disminuye el consumo máximo de los metabolitos transportados. En este sentido, se toma como valor de referencia la velocidad con la que se consumen los nutrientes, o en su defecto, la concentración de estos nutrientes en el medio, siendo este último el más fácil de medir y el más genérico, pues los flujos o velocidades de consumo pueden variar dependiendo del contexto (Marinos et al., 2020). En este sentido, se debe definir la composición química del ambiente en el que se encuentra el organismo y sus respectivas concentraciones.

La dificultad de este proceso depende del tipo de medio de cultivo utilizado, por ejemplo, los **medios sintéticos**, que son ampliamente utilizados en la actualidad por su conveniencia y repetibilidad, son más sencillos de definir, pues al ser diseñados y preparados artificialmente, tienen una composición química estandarizada, entonces, basta con buscar esta información en la tabla de datos de proveedor para saber las concentraciones de cada metabolito (Bernstein et al., 2021). Por otro lado, están los **medios naturales** que se definen como fluidos de procedencia animal y aunque estos están siendo gradualmente reemplazados por los medios

sintéticos (Yang & Xiong, 2012), medios como el suero, siguen siendo usados en el cultivo de organismos para los que no se conoce completamente su requerimiento nutricional o expresan fenotipos no reproducibles con otros medios de cultivo (Bernstein et al., 2021). Este tipo de medios son ricos en nutrientes, hormonas y otros metabolitos, pero debido a su gran variabilidad entre lotes, su composición química no puede ser fácilmente listada y cuantificada, por lo que se dice que esta es desconocida.

En este sentido, el uso de medios naturales como ambiente nutricional de los GEMs introduce una incertidumbre en las predicciones del modelo, por lo que si se quiere obtener una predicción más precisa se recomienda utilizar técnicas experimentales como exo-metabólica en el medio de cultivo usado en las muestras *in-vitro*, midiendo los metabolitos extracelulares para inferir las tasas de consumo y secreción de cada metabolito (Bernstein et al., 2021).

### 2.2.5. Curación de modelo

En este trabajo se tienen en cuenta tres procesos de curación de GEMs: el llenado de huecos, balance de masas y la eliminación de los ciclos termodinámicamente inviables. En un GEM, cada metabolito debe estar presente como mínimo en dos reacciones, una en la que se consuma y otra en la que se produzca, de no ser así se considera un metabolito “dead-end”, lo que implica que hay una ruptura o “hueco” (gap) en la red metabólica. Estos huecos surgen en un GEM ya sea por que se hizo una reconstrucción desde cero y estos representan el conocimiento faltante de la red metabólica o se contextualizó un GEM y durante la contextualización se removieron reacciones importantes para mantener toda la red conectada. En este sentido, el **llenado de huecos** o “gap-fill”, consiste en tomar reacciones de bases de datos o del GEM de referencia para conectar los metabolitos dead-end con el resto de la red. Como resultado, se obtiene un modelo que genera un flujo, para los metabolitos requeridos en la función objetivo, a partir del ambiente nutricional definido (Bernstein et al., 2021).

Por otro lado, el **balance de masas y cargas**, permite eliminar las inconsistencias estequiométricas presentes en el GEM, causadas por la definición incorrecta de las estequiometrias en las reacciones, generando entonces conflictos con dos restricciones físicas fundamentales que deben ser cumplidas por un GEM válido, a saber: las masas moleculares de todos los metabolitos deben ser positivas y se debe respetar en principio termodinámico de conservación de masas en todas las reacciones (Gevorgyan et al., 2008). Aunque existen métodos ampliamente usados para asegurar los balances de masas y cargas en un GEM, como los implementados en COBRA Toolbox e incluso en SuBliMinimal Toolbox (Swainston et al., 2011) que fue usado en la reconstrucción de Recon2.2 (un predecesor de Recon3D) (Swainston et al., 2016), estas técnicas de balanceo automático requieren que todos los metabolitos presentes en el modelo tengan una formula química definida, requisito que no se cumple en Recon3D (modelo usado para la contextualización en este trabajo) con el propósito de tener

un mayor cubrimiento del metabolismo. Hasta donde sabemos, no hay un método definido en literatura que permita eliminar desbalances estequiométricos en GEM con metabolitos cuyas fórmulas químicas son indeterminadas.

Por último, los **ciclos termodinámicamente inviables** o rutas extremas de tipo III, son reacciones internas de la red metabólica que pueden generar flujo a pesar de la ausencia de nutrición al romper la “ley de bucle”. Esta regla es análoga a “la segunda ley eléctrica de Kirchhoff que dice que un estado estacionario no puede haber flujo neto alrededor de una red de ciclo cerrado” (Schellenberger et al., 2011). Tanto las inconsistencias estequiométricas como los ciclos termodinámicamente inviables pueden llevar a la aparición de: metabolitos de fuga, que se producen aun cuando no hay consumos y metabolitos sifón que se consumen aun cuando no hay producción. En términos prácticos, al presentarse estos ciclos en GEMs, el modelo es capaz de producir ATP de manera infinita, sin la necesidad de una fuente de carbono. Por esta razón deben ser siempre hallados y corregidos.

### 2.3. Integración de datos ómicos en modelos metabólicos a escala genómica

Los avances recientes en las tecnologías ómicas han llevado a una abundancia sin precedentes de datos correspondientes a genomas, expresión genética global y metabolitos. Aunque estos avances en la recolección de datos de alto rendimiento (high-throughput data) ofrecen una excelente oportunidad para un entendimiento más profundo de los sistemas biológicos, han causado una brecha considerable entre la generación de datos y la integración de estos datos (Pandey et al., 2019).

Como fue mencionado en la introducción, los modelos metabólicos a escala genómica, son una aproximación que permite representar los fenotipos de los organismos como un compendio de reacciones bioquímicas, sirviendo como una representación *in silico* de dichos organismos, lo que permite realizar predicciones de flujos metabólicos ante diferentes escenarios, permitiendo predecir los comportamientos fenotípicos de un organismo (Machado & Herrgård, 2014). Sin embargo, la representación de estas reacciones bioquímicas en los GEMs, no tiene una conexión con otros procesos biológicos importantes, como por ejemplo lo es la regulación genética, el splicing alternativo e incluso las tasas de cinética metabólica, entre muchos otros (Karahalil, 2016; Machado & Herrgård, 2014).

### 2.3 Integración de datos ómicos en modelos metabólicos a escala genómica

Entonces, se han propuesto diversos algoritmos que permiten la integración de datos ómicos en un solo GEM, lo que facilita la comprensión de los diferentes fenotipos de un genotipo, además de permitir dilucidar las interacciones entre diferentes tipos de datos ómicos (M. Kim et al., 2016). La implementación de estos algoritmos ha evidenciado restricciones metabólicas biológicamente más relevantes que permiten aumentar la precisión en las predicciones de flujo y por lo tanto deducir la fisiología observada (Pandey et al., 2019).

Aunque el análisis del comportamiento diferencial de los datos ómicos permite identificar las diferencias entre dos escenarios experimentales, esto raramente provee las explicaciones de mecánicas que producen las diferencias fenotípicas observadas (Sajitz-Hermstein et al., 2016). La complejidad de los hallazgos del análisis diferencial es exacerbado por la estrecha conexión entre los diferentes niveles de organización celular y sus mecanismos de regulación, en ese orden de ideas, la integración de estos y otros datos ómicos a los GEMs permite obtener una herramienta, para interpretar biológicamente los comportamientos estudiados (Sajitz-Hermstein et al., 2016).

De esta manera, los datos ómicos son una fuente de gran información, que en términos generales cumplen dos propósitos en el contexto de los GEM (Machado & Herrgård, 2014). El primero de ellos, es que los datos ómicos (por ejemplo, datos transcripcionales, proteómicos e incluso metabolómicos), permiten **generar modelos tejidos específicos**. Es decir que es posible tomar modelos globales del metabolismo, como HMR o Recon, y con base en los datos ómicos, identificar el conjunto de reacciones propias de un tejido específico y de esta manera generar un modelo metabólico (GEM) de una línea celular dada. Y el segundo propósito que cumplen los datos ómicos, es que permiten **contextualizar** los modelos metabólicos ya creados, recreando ambientes biológicos específicos.

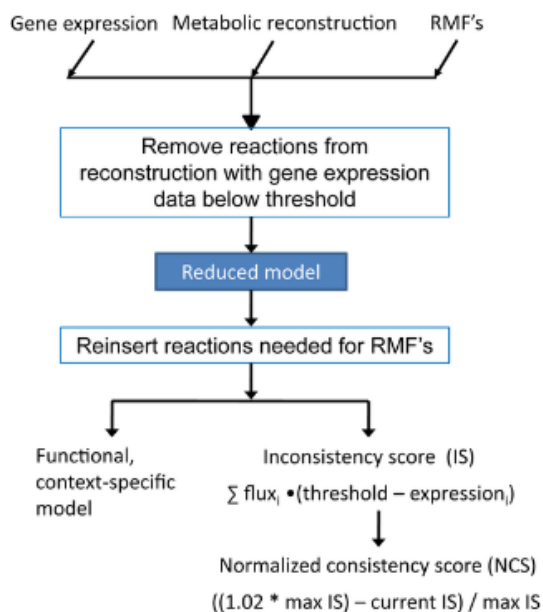
**Tabla 2-1.:** Algoritmos para contextualización de modelos. Esta es una tabla comparativa entre los algoritmos revisados en este documento, mostrando el tipo de dato que soporta cada uno y cuando fue publicado.

Método	Metabólica	Transcriptómica	Proteómica	Otro
GIMME (2008)	No	Si	Si	No
E-flux (2009)	No	Si	No	No
PRIME (2014)	No	Si	No	Datos de un fenotipo específico
E-Flux 2 y SPOT (2016)	No	Si	No	No
iMAT (2010)	No	Si	Si	No
GIMMEp (2012)	No	Si	Si	No
INIT (2012)	No	Si	No	base de datos HPA
tINIT (2014)	No	Si	No	base de datos HPA
Método de Xin Fang y compañía (2012)	No	Si	No	No
GX-FBA (2012)	No	Si	No	No
MBA (2010)	Si	No	No	No
GIM3E (2013)	Si	Si	No	No
ReMet-Flux (2016)	Si	No	No	No
MOMA (2016)	Si	Si	Si	No
PROM (2010)	No	Si	No	Red de regulación genética
Advanced data integration method (2016)	No	Si	Si	Fluxómica y perfil ribosomal
REMI (2019)	Relativa	Expresión relativa	No	Termodinámica

### 2.3.1. Reconstrucción de modelos específicos de tejido

#### Expresión genética y abundancia proteica

Las aproximaciones que integran los datos de expresión genética en los GEMs, pueden ser clasificados en dos clases: (i) los que se basan en la integración de los datos absolutos de la expresión genética, que presentan una serie de desventajas, como la necesidad de establecer umbrales de expresión definidos por el usuario y la sensibilidad de sus resultados a dichos umbrales (Pandey et al., 2019). Estas limitaciones motivaron el desarrollo de una segunda generación de métodos, (ii) los algoritmos que integran los datos relativos de expresión genética mientras intentan maximizar la correlación entre los cambios diferenciales en la expresión y los flujos de reacciones, asumiendo que los cambios relativos en la expresión entre dos condiciones se correlaciona con los perfiles de flujo diferencial resultantes (Fang et al., 2012; Navid & Almaas, 2012).



**Figura 2-3.:** Flujo de trabajo para el algoritmo GIMME.

Fuente: (Becker & Palsson, 2008).

### Clase 1: Basados en datos absolutos de expresión.

**GIMME (Becker & Palsson, 2009)** Propone un método llamado “*interactividad genética moderada por metabolismo y expresión*” (GIMME, por sus siglas en inglés), esta es una aproximación de extracción de redes que integra datos transcriptómicos para producir reconstrucciones de contexto específico más consistentes con los datos disponibles, a pesar del ruido inherente a técnicas como los microarreglos (Becker & Palsson, 2008), aunque también permite el uso de datos de RNA-seq o incluso de proteoma (Machado & Herrgård, 2014). Mediante el uso de un “puntaje de inconsistencia”, se evalúa qué tan consistente es un grupo de datos de expresión genética respecto a un objetivo metabólico particular, produciendo un modelo de contexto específico con un puntaje de penalidad minimizado (Schmidt et al., 2013; Töpfer et al., 2015). Este método requiere tres entradas, (i) datos de expresión genética, (ii) un GEM y (iii) una o más funcionalidades metabólicas asumidas (RMF por sus siglas en inglés). En general el flujo de reconstrucción consiste en remover las reacciones del GEM que no concuerdan con los datos transcriptómicos, para posteriormente reactivar las que sean necesarias para el RMF si es que las hay.

**E-flux (Colijn et al., 2009)** E-flux usa expresión genética, extendiendo la técnica de FBA mediante el modelado de los límites del máximo flujo como una función de la expresión genética medida. Contrario a otros acercamientos que interpretan metabólicamente



el transcriptoma, E-Flux utiliza un modelo de una red metabólica subyacente para predecir cambios en la capacidad del flujo metabólico. Este algoritmo no asume que el flujo de las reacciones sea determinado por los valores de expresión de mRNA, sin embargo, si se permite una acumulación mínima de enzimas, el nivel de expresión puede emplearse como un indicador del límite máximo; por lo tanto, sería erróneo retirar una reacción si esta no está expresada, ya que aún puede tener un flujo mínimo, entonces en este algoritmo las restricciones de cada reacción son variadas proporcionalmente al nivel de expresión del gen correspondiente (Colijn et al., 2009).

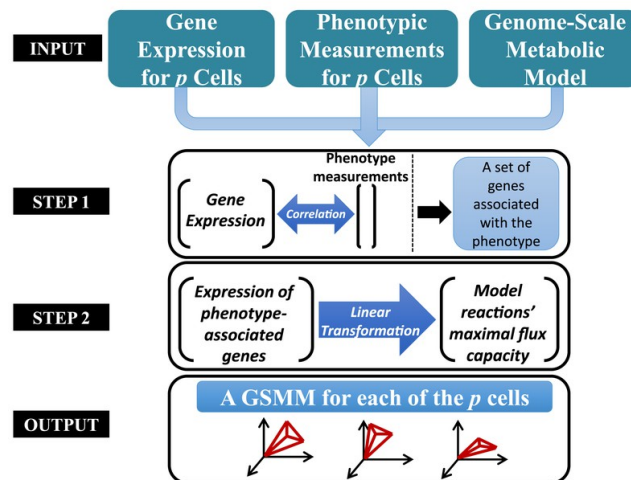
**PRIME (Yizhak et al., 2014)** El algoritmo PRIME, que por sus siglas en inglés significa reconstrucción de modelos metabólicos personalizados, utiliza información molecular y fenotípica de las células, para reconstruir modelos personalizados específicos. Este configura la capacidad de flujo máxima en un subgrupo seleccionado del modelo genérico de acuerdo a los niveles de expresión asociada y los datos fenotípicos (Yizhak et al., 2014).

Aunque PRIME está basado en E-Flux, que es usado para reconstruir modelos con datos medidos en dos condiciones diferentes, agregando dichos niveles de expresión genética entre sí para obtener un modelo. PRIME por otra parte, reconstruye un modelo metabólico basado en solo una muestra para cada célula e implementa tres mejoras que permiten generar modelos más precisos: primero, (i) busca un el grupo de genes metabólicos que afectan el fenotipo de interés significativamente, y solo modifica los límites superiores de esas reacciones, ya que estas limitaciones se consideran restricciones fuertes y es mejor evitar sobre restringir el modelo basado en evidencia ruidosa o irrelevante. Además, (ii) emplea la información fenotípica para determinar la dirección (signo) de la relación entre la expresión genética y las tasas de flujo. Y por último, (iii) PRIME modifica los límites de las reacciones dentro de un rango predefinido donde se sabe que las modificaciones tendrán el mayor impacto en el fenotipo dado (Yizhak et al., 2014).

Este algoritmo presenta algunas limitaciones como por ejemplo, que intenta maximizar la proliferación de las células por lo que no puede ser utilizado en organismos que no tengan este comportamiento. Además, asume que las células a modelar comparten el mismo grupo de enzimas y difieren solo en la abundancia celular, sin embargo, pueden presentarse diferencias por las variantes genéticas. Y por último, depende de la medida directa de un fenotipo específico, pero, esto no está siempre disponible.

**E-Flux2 y SPOT (M. K. Kim et al., 2016)** Estos dos métodos utilizan como fuente experimental datos de transcriptómica y se complementan entre sí, por lo que pueden ser elegidos o combinados dependiendo de si se conocen o no las fuentes de carbono y la función objetivo, ambos están disponibles en el paquete open-source MOST(<http://most.ccib.rutgers.edu/>). A continuación se presenta la tabla de decisión que permite determinar qué método utilizar.

## 2.3 Integración de datos ómicos en modelos metabólicos a escala genómica



**Figura 2-4.:** El esquema del algoritmo PRIME.

Como entrada las medidas de expresión genética y fenotípica (ejem. Tasa de proliferación) de varias células es usada, (paso 1): un grupo de genes cuya expresión está asociada significativamente con el fenotipo identificado. (paso 2) una transformación lineal de la expresión de los genes asociados al fenotipo, la aplicación de los límites máximos en el GEM. Por último este algoritmo entrega un GEM para cada célula de entrada, de manera que cada célula genera un espacio de solución diferente. Fuente: (Yizhak et al., 2014).

**Tabla 2-2.:** Cuadro de decisión para el uso de algoritmos (E-Flux2 y SPOT) y plantillas propuestas (DC y AC) fuente: (M. K. Kim et al., 2016).

	<b>Fuentes de C conocidas</b>	<b>Fuentes de C desconocidas</b>
Función objetivo conocida	DC + E-Flux2	AC + E-Flux2
Función objetivo desconocida	DC + SPOT	AC + SPOT

Respecto a su funcionamiento interno, el primer paso es común para ambos, convirtiendo el GEM de entrada en una plantilla DC cuando las fuentes de carbono se conocen o AC si son indeterminadas, cambiando los límites de las reacciones para resolver los problemas de escalado entre los flujos del GEM y los datos de transcriptómica (M. K. Kim et al., 2016). Luego se emplea E-Flux2 solo cuando se conoce la composición de biomasa y el objetivo es maximizarla, como en los microorganismos que buscan acumular masa para proliferar. Sin embargo, ese no es el caso para la mayoría de las células en los organismos multicelulares, por lo que para estos casos se recomienda el uso de SPOT, que por sus siglas en inglés traduce correlación de Pearson simplificada con datos transcriptómicos (M. K. Kim et al., 2016).

El problema de E-Flux es que la solución obtenida no es única, haciendo difícil identificar las respuestas metabólicas predichas, por lo que en E-Flux2 se añade la normalización Euclidiana, que obtiene la distribución de flujo en la que se logra conseguir la máxima tasa de crecimiento de la forma más eficiente. Mientras que SPOT maximiza la correlación entre el vector de flujo y los datos de expresión correspondientes, ya que la existencia de un transcrito es necesaria para la presencia o actividad de la enzima correspondiente. En resumen, los datos de transcriptómica son usados para restringir los flujos en el modelo en E-Flux2 y para definir la función objetivo en SPOT (M. K. Kim et al., 2016).

**iMAT (Zur et al., 2010)** Por sus siglas en inglés es una herramienta de análisis de integración metabólica, que permite utilizar datos de expresión de proteínas y/o genes para un determinado organismo, para generar un mapa del estado metabólico de este organismo, mostrando los flujos más probables predichos dentro de la red de reacciones (Zur et al., 2010), permitiendo generar un GEM específico de tejido, como también integrándolo a uno existente, para aumentar su predicción (Machado & Herrgård, 2014).

Dentro de los datos de entrada, se tienen valores de expresión baja, media y alta en la condición, lo que le permite al algoritmo iMAT determinar los estados de flujo para cada reacción en el modelo metabólico, reflejando la presencia o ausencia de los flujos asociados. Además, se pueden determinar la actividad de flujo como activa o inactiva para algunas reacciones, ya que otras no pueden ser completamente determinadas debido a la presencia de isomerasas o rutas alternativas; por lo tanto, para este último caso, si el flujo se desvía

del estado de expresión, se considera que la enzima que cataliza la reacción es sobre o sub regulada después de la transcripción (Zur et al., 2010).

**MADE (P. A. Jensen & Papin, 2011)** El algoritmo de ajuste metabólico por expresión diferencial, o MADE por sus siglas en inglés, que trata de capturar las dinámicas de expresión genética en un modelo metabólico, identificando los cambios en el transcriptoma que son más significativos estadísticamente, y mediante la optimización de dichos cambios obtiene una secuencia de estados binarios de expresión, esta última debe producir un modelo funcional, ya que la optimización tiene impuesto un valor mínimo en la función objetivo del sistema. Este algoritmo asume que la función objetivo del organismo es maximizar el flujo a través de un subgrupo de rutas, que en organismos unicelulares corresponde a la producción de biomasa, como una aproximación para describir el crecimiento (P. A. Jensen & Papin, 2011).

Principalmente, propone un método para determinar los umbrales de activación de cada reacción de manera automática y basada en los datos de transcriptómica, pues aunque métodos como GIMME puede generar modelos funcionales, estos requieren de un umbral de expresión determinado a priori y arbitrario, omitiendo detalles cinéticos para la asociación metabólica entre los genes y las reacciones.

**GIMMEp (Bordbar et al., 2012)** El algoritmo llamado interactividad genética moderada por metabolismo y expresión por proteoma o GIMMEp por sus siglas en inglés, es una versión basada en el algoritmo GIMME, que en resumen, construye un modelo contexto-específico, usando datos de expresión para activar la menor cantidad de reacciones necesarias para satisfacer un una reacción objetivo, definida por el usuario; teniendo en cuenta lo anterior, GIMMEp usa datos de proteómica como una guía de funciones objetivo, evaluadas en el algoritmo GIMME tradicional, produciendo subredes individuales para cada función objetivo asociada al proteoma, posteriormente combinadas para crear un único modelo (Bordbar et al., 2012).

### Clase 2: Basados en datos relativos de expresión

**INIT (Agren et al., 2012)** La interfaz integrativa de red para tejidos o INIT por sus siglas en inglés, es un algoritmo que permite la reconstrucción de redes metabólicas específicas de tejido, usando como fuente de información principal el Atlas de proteínas humanas (HPA, por sus siglas en inglés) y como fuente secundaria usa transcriptómica específica de tejido. Su principal limitación es que solo puede reconstruir modelos de células o tejidos que estén incluidos en el HPA, la cual utiliza para asignar pesos a cada reacción dependiendo de la de la abundancia reportada en esta base de datos, que posteriormente son cruzados con el transcriptoma para determinar qué rutas metabólicas se encuentran activas en el tejido

analizado.

Este algoritmo es formulado como un problema lineal de enteros mixtos (MILP), además se plantea que todas las reacciones sean capaces de llevar flujo, obteniendo un vector de la acumulación o consumo para cada metabolito en vez de asumir un estado estacionario para todos los metabolitos internos, esto se hace basado en la premisa de que las células humanas secretan una variedad de metabolitos más amplia que los microorganismos.

**tINIT (Agren et al., 2014)** El algoritmo de reconstrucción impulsado por tareas o tINIT, por sus siglas en inglés, se encuentra disponible en el RAVEN Toolbox, y está basado en el algoritmo INIT desarrollado por Agren et al. (2012), el cual ayuda a automatizar la generación de GEMs específicos de células. tINIT permite definir una tarea metabólica que el modelo resultante debe ser capaz de realizar. En el primer paso revisa las reacciones que al ser excluidas hacen fallar una o más tareas, por lo que deben estar presentes en el modelo final. Para el siguiente paso, se usa el algoritmo INIT original, pero como requerimiento adicional, las reacciones encontradas en el paso anterior deben estar incluidas en la solución. Ya en este punto, el modelo es probablemente capaz de realizar las tareas planteadas, sin embargo como último paso se prueba secuencialmente cada tarea, y si no es capaz de realizar alguna se aplica el algoritmo de llenado de huecos de RAVEN Toolbox (Agren et al., 2014).

Este algoritmo presenta dos ventajas adicionales respecto a INIT, (i) limita la solución para que las reacciones reversibles no tengan un flujo hacia ambas direcciones al mismo tiempo, lo que evita que se formen algunos bucles independientes, que no están conectados a la red principal. (ii) Permite elegir si se permite la producción de todos los metabolitos (Agren et al., 2014).

**Método de Xin Fang y colaboradores (Fang et al., 2012)** Se plantea una aproximación para la integración de datos de expresión genética en GEMs, que contrario a otros métodos, como GIME e iMAT, no requiere de un umbral arbitrario para determinar qué genes son relevantes. Este método utiliza los niveles de expresión para definir las restricciones de cada reacción, utilizando un método de reducción continua y no discreta, pues en estos casos se pierde el sentido de regulación de las reacciones metabólicas mediante los niveles de expresión. Además, se asume que un cambio significativo en la expresión de los genes debe significar un cambio en el flujo de la reacción relacionada, aunque ese cambio no es requerido, también se permite que la composición de la masa cambie de acuerdo a la condición (Fang et al., 2012).

**GX-FBA (Navid & Almaas, 2012)** Es una aproximación que permite combinar datos de expresión genética con FBA (GX-FBA), y permite el modelado metabólico de diferentes respuestas en un amplio rango de perturbaciones ambientales. Usa un acercamiento de múltiples pasos, restringiendo el uso de los datos transcriptómicos, provenientes de microarreglos a

los genes que ya se encuentran incluidos en el modelo, y que además presentan cambios de al menos 50%. Aunque se ha demostrado que el umbral de cambio (fold-change) no afecta significativamente los resultados de este algoritmo (Navid & Almaas, 2012), es importante tener en cuenta que si este parámetro es muy alto, se integrarán pocas restricciones en el modelo.

El procedimiento consiste en correr un FBA para el modelo existente, obteniendo un campo de flujos necesarios para cumplir con la función objetivo y un FVA que es usado para determinar los flujos máximos y mínimos para cada reacción del modelo. Luego, se identifican los datos de mRNA que pueden ser mapeados dentro del modelo, y aquellos valores de flujo demasiado altos son eliminados, para evitar problemas numéricos. Finalmente, se le asigna una nueva restricción dependiendo de si el gen se vio sobreexpresado o subexpresado (Navid & Almaas, 2012).

### Metabólicos

Si bien los datos de expresión han sido ampliamente usados en el contexto de los GEM, más recientemente, gracias al aumento en la disponibilidad y calidad de estos, se ha promovido el desarrollo de métodos que también usan datos metabólicos. El metaboloma puede ser integrado en los GEMs para refinar la reconstrucción del modelo y reducir el espacio de soluciones, permitiendo hacer una predicción más acertada del estado fisiológico del sistema (Pandey et al., 2019).

**MBA (Jerby et al., 2010)** El algoritmo para construcción de modelos (MBA por sus siglas en inglés) hace uso de metabolitos que son detectados en un órgano o tejido específico (Jerby et al., 2010; Töpfer et al., 2015). La metabolómica es usada en conjunto con otra información específica del tejido, como conocimiento de literatura, transcriptómica, proteómica y datos fenotípicos, para definir dos grupos de reacciones: (i) uno del alta probabilidad y (ii) otro con una probabilidad media, en donde las reacciones que tienen una alta probabilidad son aquellas que pertenecen a una ruta metabólica específica curada. Mientras que una reacción que es considerada con una probabilidad media, solo si la reacción es necesaria para la inclusión de un metabólico que apareció en los datos metabólicos o es soportada por al menos dos de las fuentes de información usadas (Töpfer et al., 2015).

El siguiente paso de optimización usa una búsqueda heurística para llegar a el modelo más parsimonioso y consistente que garantice la inclusión de todas las reacciones de alta probabilidad y maximice el número de reacciones de probabilidad media, además de un grupo adicional de reacciones para rellenar los huecos en la red (Töpfer et al., 2015).

**GIM3E (Schmidt et al., 2013, p.3)** La inactivación de genes moderados por metabolismo, metabolómicas y expresión (GIM3E por sus siglas en inglés) es una extensión del algoritmo GIMME (Becker & Palsson, 2008), con datos metabólicos añadidos a los datos de expresión genética, mediante recambio de metabolitos y su respectiva reacción de sifón para el modelo genérico, por lo tanto, un flujo a través de dicha reacción que se obtiene y detecta experimentalmente en los metabolitos, puede ser integrada forzando un flujo mínimo para el recambio del metabolito correspondiente (Töpfer et al., 2015).

**iReMet-flux (Sajitz-Hermstein et al., 2016)** Para integrar datos metabolómicos relativos en un GEM y predecir flujos a una escala genómica (Sajitz-Hermstein et al., 2016), se propone la aproximación iReMet-Flux (integration of relative metabolite levels for flux prediction, por sus siglas en inglés). Como otras aproximaciones basadas en restricciones, iReMet-Flux usa una matriz estequiométrica  $S$  de la red metabólica considerada.

Asumiendo que el flujo de las reacciones que es modelado cumple con la ley de acción de masas (Sajitz-Hermstein et al., 2016), se usa una transformación que evita el requerimiento de conocer la tasa constante de cada enzima, pues este dato es desconocido, permitiendo realizar un análisis diferencial solo con los datos de concentración relativa de los metabolitos y encontrar el flujo de un caso basado en el flujo del otro y los cambios de las concentraciones de metabolitos entre estos (Sajitz-Hermstein et al., 2016). Adicionalmente, debido a la variabilidad técnica y biológica se obtienen las restricciones superiores e inferiores de cada reacción. Con estos datos, se busca el flujo más cercano entre los dos casos estudiados, dicha suposición también es usada por otras aproximaciones basadas en restricciones como MOMA (Segrè et al., 2002), que minimiza la distancia euclidiana de la distribución de flujos entre dos escenarios, sin embargo, esta aproximación no depende de un análisis de distribución de flujos de referencia (como es el caso de MOMA) y encuentra un flujo con el mínimo de distancia euclidiana entre los dos casos mediante una optimización simultánea de ambos modelos.

$$v_i^A = [q_i \prod_{j=1}^n (r_j)^{s_{ij}}] v_i^B \quad (2-1)$$

Correlación entre los flujos de las reacciones  $v$ , de la reacción  $i$ , para los casos A y B, en donde se tienen en cuenta el nivel relativo del metabolito  $j(r_j)$ , el nivel relativo de la enzima catalizadora de la reacción  $i$  ( $q_i$ ) y el coeficiente estequiométrico del metabolito  $S_{ij}$  (Sajitz-Hermstein et al., 2016).

Se ha demostrado que se puede usar para perfilar el comportamiento diferencial del flujo en una escala genómica, con el compromiso de hacer algunas suposiciones de simplificación (Sajitz-Hermstein et al., 2016). Sin embargo, esta aproximación requiere una evaluación de los cambios diferenciales de todos los metabolitos existentes en un GEM, esto limita su aplicación,

ya que la mayoría de las técnicas disponibles no permiten obtener los datos metabólicos de todo el sistema, sino sólo obtienen los datos para unos pocos metabolitos (Pandey et al., 2019).

### Integración de datos ómicos

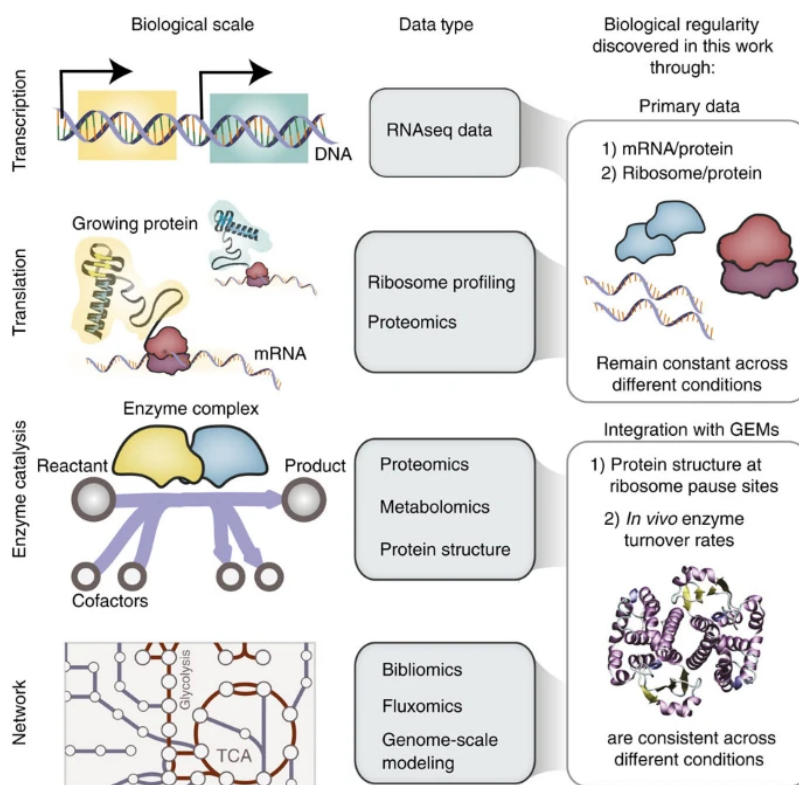
Ya expuestas las aproximaciones que mediante el uso de datos ómicos entrenan modelos metabólicos para mejorar su precisión y fidelidad respecto a la fisiología real de los organismos estudiados. Se procede a mostrar una familia diferente de aproximaciones, que al contrario de las anteriormente expuestas, no se limita a una sola fuente de información experimental e integra múltiples fuentes ómicas, para mejorar GEMs existentes con ayuda de dichas fuentes.

**PROM (Chandrasekaran & Price, 2010)** El algoritmo de modelado probabilístico para la integración de redes metabólicas y reguladoras a escala-genómica o PROM, por sus siglas en inglés, integra redes de regulación genética (GRN) a GEMs con la ayuda de datos de transcriptómica. PROM introduce probabilidades para representar los estados genéticos y factores de interacción entre gen y proteína, en donde, para la regulación genética, se usan las probabilidades y para el modelado metabólico se usa FBA (Chandrasekaran & Price, 2010).

El uso de este algoritmo requiere cuatro cosas: (i) un GEM reconstruido, (ii) la GRN de organismo, (iii) interacciones adicionales que involucre regulación de enzimas por proteínas y metabolismo y por último (iv) datos de expresión génica abundantes y de diferentes perturbaciones genéticas y ambientales. El tamaño de transcriptómica afecta directamente la robustez de la probabilidad obtenida, ya que son usados para determinar las interacciones y limitar el flujo metabólico máximo de cada reacción en factor de dicha probabilidad (Chandrasekaran & Price, 2010).

**Advanced data integration method (Ebrahim et al., 2016)** Es un método para integración de datos ómicos que permite el análisis de múltiples niveles utilizando datos genómicos, transcriptómicos, perfilado ribosomal, proteómicos y fluxómicos. Mediante el uso de GEMs reconstruidos a partir de información genómica y de bibliografía, se puede realizar la sincronización de varios tipos de datos discordantes, permitiendo una interpretación y predicción coherente de los mecanismos subyacentes en el sistema estudiado (Ebrahim et al., 2016).





**Figura 2-5.:** Proceso de integración de datos ómicos.

Se muestra el dogma central de la biológica (a la izquierda), y cómo estas etapas pueden ser medidas mediante técnicas experimentales (columna central) para ser integrados en un GEM (columna derecha) (Ebrahim et al., 2016).

**MOMA (M. Kim et al., 2016)** Tradicionalmente, la integración de datos específicas a un hospedero ha sido limitada a dos capas, principalmente debido a la falta de datos a través de múltiples capas para las mismas condiciones experimentales, a medida que se acumulan más datos de diferentes capas ómicas, se hace más útil la implementación de este tipo de integraciones horizontales y verticales.

La plataforma MOMA (Multi-Omics Model and Analytics plataforma), es un modelo integrado que aprende de Ecomcs y otras redes de datos disponibles para predecir la expresión y crecimiento en todo el genoma. En esta aproximación se plantea un método semi-supervisado para la normalización y caracterización experimental de múltiples datos ómicos (crecimiento, transcriptómica y proteómica), que es usado como base para el entrenamiento de un modelo multi escala de *Escherichia coli*, que integra cuatro capas ómicas, evidenciando una mejora en el desempeño predictivo de concentraciones y dinámicas de crecimiento en todo el genoma (M. Kim et al., 2016).

**REMI (Pandey et al., 2019)** Asumir que los datos de expresión genética pueden ser directamente correlacionados con los valores de flujo es cuestionable y puede no ser verdad para todos los genes (Pandey et al., 2019); además los otros métodos requieren de umbrales definidos por el usuario para identificar y categorizar los niveles de expresión de los genes metabólicos (expresión alta, baja y moderada) y los resultados son sensibles a los umbrales, por lo tanto se propone el método REMI (Relative Expression and Metabolomic Integrations por sus siglas en inglés).

Este método permite integrar la expresión genética y abundancia de metabolitos relativos en modelos GEM curados con datos termodinámicos para maximizar la consistencia entre los datos, lo que ha demostrado reducir el espacio de soluciones factible respecto al análisis de balance de flujos (FBA por sus siglas en inglés) (Pandey et al., 2019). Adicionalmente permite traducir los datos de expresión diferencial y abundancia de metabolitos en flujos diferenciales para analizar la fisiología alterada para cualquier condición dada por perturbaciones genéticas o ambientales. Basado en el tipo de datos integrados hay tres métodos REMI:

**Tabla 2-3.:** Tipos de métodos REMI. Los tres métodos que tiene REMI según los tipos datos ómicos que se vayan a usar.

Método	Datos integrados en un modelo FBA
REMI-TGex	Datos termodinámicos y de expresión genética.
REMI-TM	Datos termodinámicos y metabólicos.
REMI-TGexM	Datos termodinámicos, de expresión genética y metabólicos.

Muestra los tipos de métodos REMI dependiendo de los datos integrados, aunque solo se muestran tres, se puede utilizar un modelo FBA sin optimización termodinámica obteniendo dos tipos adicionales REMI-Gex y REMI-T que integran datos de expresión genética y metabólicos en un modelo FBA respectivamente.

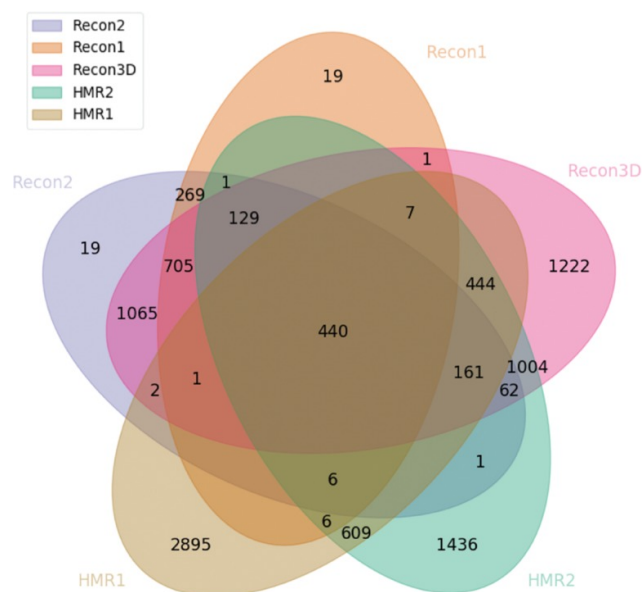
### 2.3.2. GEM contexto específicos

En sus inicios, los GEM estuvieron siempre enfocados en la representación del metabolismo general de un organismo completo. Si bien esta es una aproximación válida para formas de vida más simples, como bacterias, esta consideración no es enteramente válida para organismos multicelulares, que tienen una alta especialización celular, en la que distintos tejidos tienen particularidades metabólicas específicas y en ocasiones completamente distintas, aunque se encuentren dentro del mismo organismo. Un claro ejemplo de esta situación es el metabolismo humano y su alta especialización tejido específica (como por ejemplo el metabolismo de fructosa, específico de los hepatocitos), por esta razón cada vez se desarrollan más modelos tejido específicos, que nos permitan representar de manera más acertada la realidad biológica.

#### GEMs modelo para *Homo sapiens*:

En el 2007 se publicaron los dos primeros GEM específicos de humano, bajo el nombre de Recon1 y EHMN (Edinburgh human metabolic network, por sus siglas en inglés), que se reconstruyeron en paralelo a partir de la secuenciación del genoma humano (Masid et al., 2020; Vieira et al., 2019). Como los dos modelos buscaban cubrir todo el metabolismo humano, esfuerzos posteriores por integrar ambos modelos dieron origen: primero a un GEM de hepatocito en 2010 llamado *HepatoNet1*, y luego, en 2013 a la primera versión de otra familia de modelos humanos llamado *HMR*, que por sus siglas en inglés significa *reacciones metabólicas humanas*. En este mismo año también se presentó una nueva versión de Recon (*Recon2*) que abarca las partes del metabolismo incluidas en *HepatoNet1* y en *EHMN* (Vieira et al., 2019).

Al igual que Recon2, HMR contiene información de localización subcelular y expresión genética tejido específica, pero, en HMR el metabolismo de ácidos grasos es más completo (Gu et al., 2019), en consecuencia, se reconstruyeron varios GEM para contextos en los que esta característica era importante, dando lugar a iAdipocytes1809 para estudiar obesidad (Mardinoglu et al., 2013), iHepatocytes2322 para NAFLD (Enfermedad del hígado graso no alcohólico) (Mardinoglu et al., 2014) e iMyocyte2419 para el estudio de la diabetes (Väremo et al., 2015).



**Figura 2-6.:** Diagrama de Venn para los GEM de humano.

En esta gráfica se comparan las reacciones presentes en cada modelo, ilustrando la cantidad de reacciones en común entre los diferentes GEM de humano reconstruidos hasta la fecha, fuente: (Vieira et al., 2019).

Esfuerzos posteriores para desarrollar un GEM humano de alta calidad, incluyeron correcciones e información biológica adicional en Recon para generar las versiones subsiguientes de esta serie, que incluyeron a Recon 2, Recon 2.2, Recon 2 M.2 y por último Recon 3D (Gu et al., 2019; Masid et al., 2020).

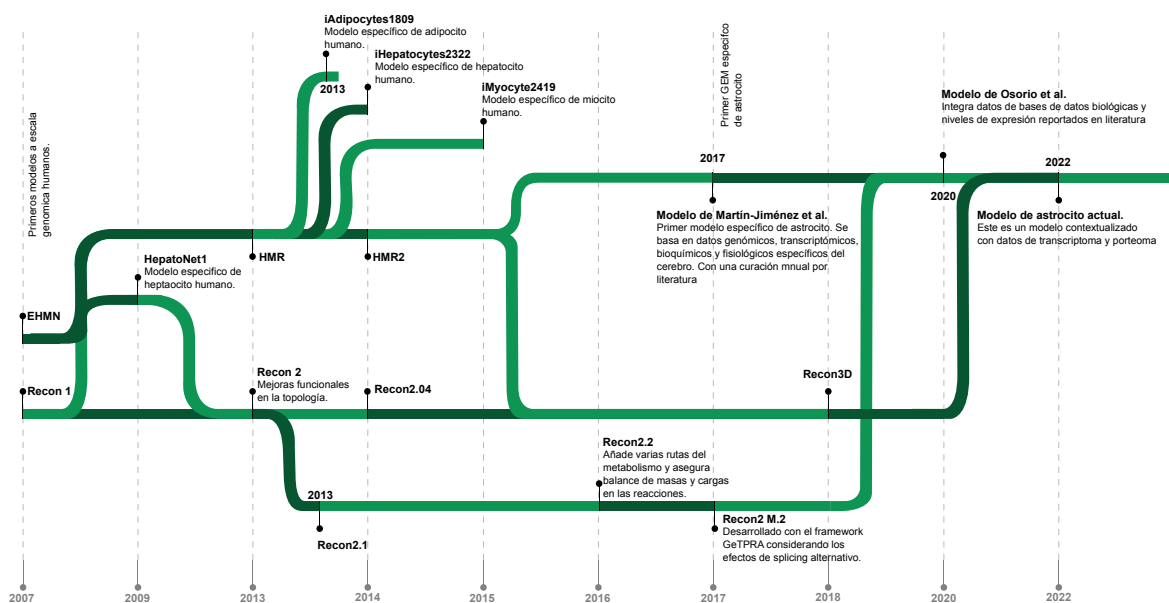
En la versión Recon 2 M.2, se empiezan a evidenciar los esfuerzos de la comunidad científica por utilizar a los GEM como medios idóneos para la integración y contextualización de datos biológicos, por ejemplo en esta versión de Recon se desarrolla el framework GeTPRA para generar reacciones metabólicas considerando los efectos de splicing alternativo para los genes metabólicos que fue publicada en 2017 (Gu et al., 2019). Posterior a esto, se publica Recon 3D en el 2018, que contiene la asociación más extensiva de GPR e información estructural de metabolitos y enzimas, este puede ser usado como una herramienta para varias aplicaciones biomédicas, incluyendo la caracterización de enfermedades asociadas a mutaciones y respuestas a medicamentos (Gu et al., 2019; Masid et al., 2020).

**GEMs específicos de astrocito:**

Por otra parte, en la actualidad existen dos modelos metabólicos específicos para el astrocito, siendo el primero, el GEM de astrocito reconstruido por Martín-Jiménez et al. (2017) en el laboratorio de bioquímica computacional en la Pontificia Universidad Javeriana, basándose en datos genómicos, transcriptómicos, bioquímicos y fisiológicos específicos del cerebro, y posteriormente pasando por una curación manual basada en literatura para generar un modelo de alta calidad. Dicho modelo incluye en su mayoría metabolismo energético, que consiste en metabolismo de carbohidratos, aminoácidos, nucleótidos, cofactores y vitaminas, además de la biodegradación de xenobióticos (Martín-Jiménez et al., 2017).

Más recientemente, nuestro grupo de investigación Gibbs en conjunto con el laboratorio de bioquímica computacional en la Pontificia Universidad Javeriana, realizó la reconstrucción del segundo GEM de astrocito (Osorio et al., 2020). Mediante una revisión bibliográfica, integrando datos provenientes de bases de datos biológicas y datos de expresión genética reportados en muestras experimentales de dominio público. A diferencia del modelo de Martín-Jimenez que fue basado en HMR, el modelo de Osorio tomó como referencia el modelo global a RECON 2.04.. Este GEM es usado para la evaluación de tres estados metabólicos: (i) un estado de astrocito sano o normal, (ii) un estado de respuesta a un alto ingreso de ácido palmítico y (iii) un pre-tratamiento con tibolona después del estado ii; permitiendo analizar los mecanismos inflamatorios y el rol neuroprotector que cumple la tibolona ante dichos estados.

## 2.3 Integración de datos ómicos en modelos metabólicos a escala genómica



**Figura 2-7.:** Línea de tiempo GEMs humano y astrocito

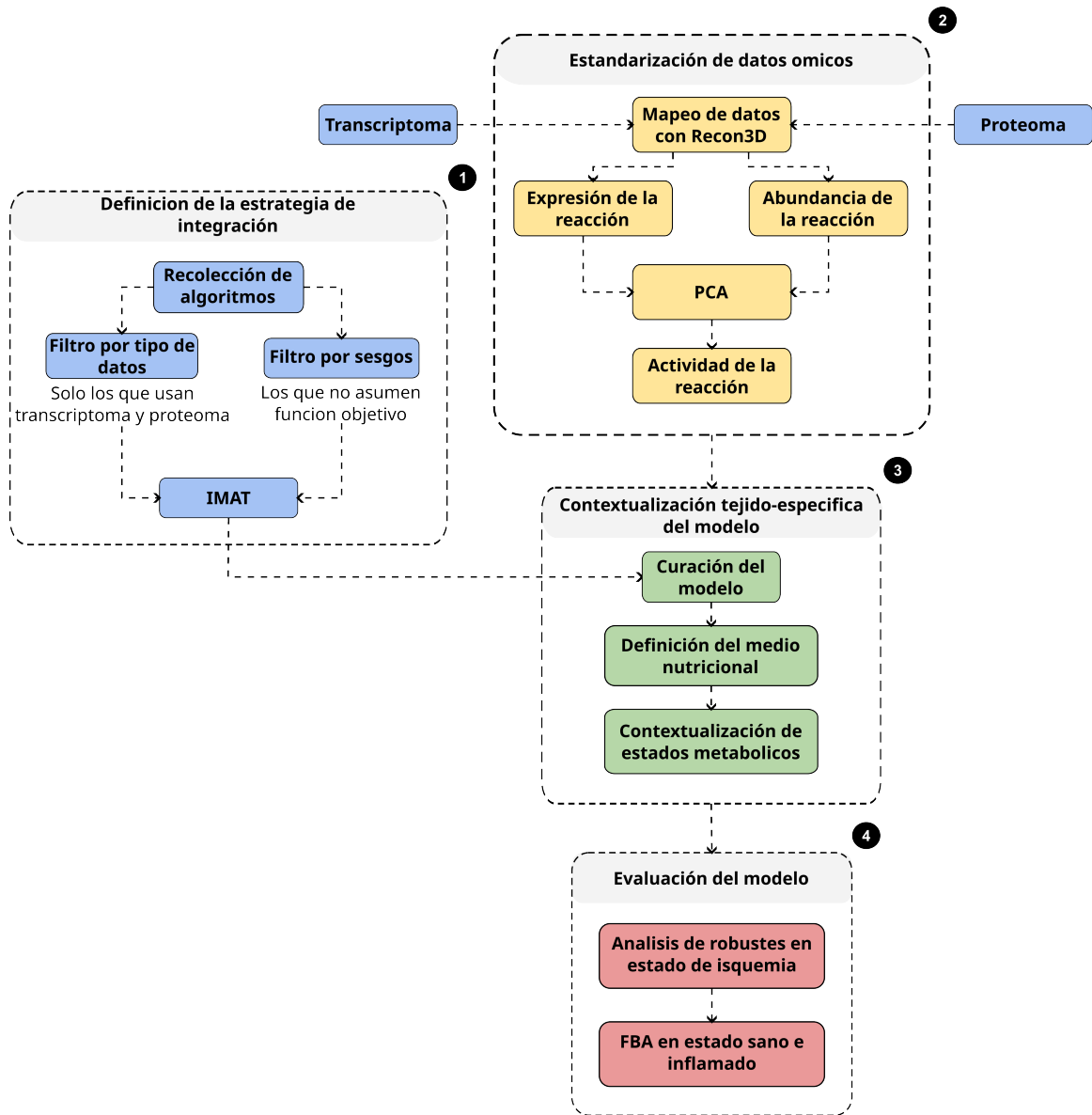
Línea de tiempo de los modelos metabólicos a escala genómica específicos de humano Recon y HMR. Y cómo estos han sido usados para la reconstrucción de GEM específicos de astrocito, fuente: elaboración propia basado en (Masid et al., 2020; Vieira et al., 2019).

## 3. Capítulo 2:Métodos

A continuación se presenta el proceso realizado para generar un GEM tejido específico de astrocito humano, cuya construcción se basó en datos transcriptómicos y proteómicos para reducir la red metabólica de Recon3D, el GEM general de humano más completo hasta el momento.

Los datos de transcriptómica y proteómica usados en este proyecto fueron proveídos por el laboratorio de bioquímica computacional de la Pontificia Universidad Javeriana. Para la obtención de estos, se usó la línea celular *Normal Human Astrocytes* (NHA) de Lonza, con la que se reprodujeron los tres estados metabólicos planteados por Osorio et al. (2020) para tomar las medidas, a saber: (i) uno de astrocito normal (sano), (ii) otro de astrocito con la respuesta inflamatoria inducida mediante con ácido palmítico y por último, (iii) uno con ácido palmítico, pre-tratado con tibolona.

Estos datos ómicos se unificaron en una sola dimensión usando el análisis de componentes principales como técnica de reducción de dimensionalidad. Para luego ser usados en la contextualización de Recon3D con iMAT (Zur et al., 2010), el modelo resultante fue curado combinando las herramientas disponibles en COBRA Toolbox con un conjunto de algoritmos desarrollados en este trabajo, uno permite integrar datos multi-ómicos y el otro se usa para eliminar desbalances en el modelo. Y por último, se comprobó la capacidad predictiva del modelo en cuatro escenarios metabólicos establecidos en la construcción de los modelos de astrocito anteriores.



**Figura 3-1.:** Diagrama de las fases de la investigación para el modelado computacional del astrocito humano.

fuelle: elaboración propia.



## 3.1. Obtención de datos de transcriptómica y proteómica

### 3.1.1. Cultivo celular

Como se menciona anteriormente, este proceso fue realizado en el laboratorio de bioquímica computacional de la Pontificia Universidad Javeriana, donde se cultivaron tres lotes (#0000612736, #0000565612, #0000514417) de células primarias de *Normal Human Astrocyte* o NHA, extraídas de tres donantes diferentes, que fueron cultivadas usando ABM o *Astrocyte Basal Medium* (Lonza, Basel, Switzerland) con Suplementos *SingleQuots* (Lonza, Basel, Switzerland) en una incubadora húmeda a 37°C y 5% CO<sub>2</sub>. Luego, se aplicaron tres tratamientos para obtener un estado sano, tratado con palmitato y pre-tratado con tibolona + palmitato, generando 3 réplicas biológicas y dos técnicas por lo que se obtuvieron 18 muestras.

### 3.1.2. Tratamientos

Adicionalmente, a un subgrupo se les aplicó un **pre-tratamiento de tibolona** antes de añadir el palmitato. Se tomó como solución madre la tibolona disuelta en DMSO al 100% a 40mM. A esta se le aplicaron diluciones adicionales con DMEM libre de suero hasta llegar a una concentración final de 0.000025%. Las alícuotas se almacenaron a -20°C y cada una fue usada 3 veces o menos. El pre-tratamiento de tibolona a 10nM se aplicó por 24 horas para preservar la viabilidad celular ante el tratamiento con ácido palmítico (Martin-Jiménez et al., 2020).

Antes de aplicar el tratamiento se lavaron las células NHA en una placa microtituladora de 48 pozos a 5.000 células/cm<sup>2</sup> por 12 días, después se lavaron con 10X PBS y privados de comida en DMEM libre de suero y sin L-Glutamina, fenol rojo y suplementos (Lonza) por 6 horas. Las células fueron tratadas con: DMEM libre de suero, **ácido palmítico** a 2 mM por 24 horas (Sigma, St Louis, MO, USA), 1,35% de albúmina de suero bovino como proteína de transporte (SigmaSigma, St Louis, MO, USA, lot A2153) y carnitina a 2 mM (Sigma, St Louis, MO, USA) como un transportador dentro de la matriz mitocondrial.

### 3.1.3. Datos de transcriptoma

El ARN total se extrajo con el mini kit RNeasy (Qiagen, EE. UU.), siguiendo las instrucciones del fabricante. Se usó *ADNasa I* sin ARNasa para evitar la contaminación con ADN genómico. Las muestras se enviaron a -80 °C en una solución tampón libre de nucleasas a la instalación de secuenciación de *GeneWiz Plainfield* (NJ, EE. UU.). Después del envío, se evaluó la integridad del ARN con un *Agilent TapeStation Bioanalyzer 2100* (RIN > 7) y la concentración con un fluorómetro *Qubit 2.0*. Luego, se realizó una preparación de la biblioteca de ARN específica

de cadena con selección de poli(A) utilizando el kit de preparación de biblioteca de ARNm de cadena TruSeq siguiendo el protocolo del fabricante (Illumina, San Diego, CA, EE. UU., n.º de catálogo RS-122-2101). Posteriormente, el **ARN se secuenció** en una máquina Illumina HiSeq 2000 con una configuración de extremo emparejado de 2 x 150 pb, lo que produjo ~75 millones de lecturas por muestra. Estos datos se encuentran disponibles en el repositorio Gene Expression Omnibus o GEO por sus siglas en inglés bajo el identificador [GSE166500](#).

El **control de calidad de RNA-seq** se evaluó a través de *QUARS* (QUALity control for RNA-Seq; [github.com/tluquez/QUARS](https://github.com/tluquez/QUARS)), un flujo de trabajo que integra varias herramientas incluidas en *Nextflow* (v18.10.1). Las herramientas que QUARS integra son: *FastQC* (v0.11.7), para evaluar la calidad de la secuenciación y *Fastp* (v0.20.0) para el recorte automático de adaptadores y el filtrado de baja calidad de base. Luego, se usó el pipeline *rnaseq* de *nf-core* que utiliza *Nextflow* otro conjunto de herramientas que permitieron mapear las lecturas a Gencode (versión 29) como referencia, además de cuantificar el nivel de los genes mapeados utilizando *Salmon* (v0.13.1). La matriz de recuento en cuanto a genes de salmón se derivó mediante *tximport* (v1.12), y con *NOISeq* (v2.28) se evaluó la calidad de la secuenciación a través de gráficos de saturación. Siguiendo las mejores prácticas (Conesa et al., 2016), los genes se descartaron si, después de agregar los recuentos de lectura en las réplicas biológicas y técnicas de 3 x 2 para cada tratamiento, el gen tenía menos de 10 lecturas totales para cualquiera de los 3 tratamientos.

### 3.1.4. Datos del proteoma

Para la **extracción de proteínas** se usó 1 mL de buffer de lisis con la composición que se muestra en la tabla 4. Primero se quitó el medio y se limpió el recipiente con 1 mL de 1X PBS frío, para después ser extraído con una pipeta. Luego, se añadieron 72 mL del cotel RIPA más inhibidores de proteasas en frascos de 25 cm<sup>2</sup> y se dejaron por 10 minutos en un congelador a -4°C, después se raspó y recolectó toda la superficie en un eppendorf, que se dejó en hielo por 30 minutos, mientras se agitaba por 10 a 15 segundos cada 10 minutos. Una vez cumplidos los 30 minutos, se centrifugó a 15200 rpm y -4°C por 13 minutos, el sobrenadante se pasó a un nuevo tubo y fue congelado a -80°C.

**Tabla 3-1.:** Composición del buffer de lisis. Componentes utilizados para preparar el buffer para la extracción de las proteínas y sus respectivas concentraciones y cantidades.

Nombre	Concentración	cantidad
RIPA buffer, Lonza	-	720 mL
sodium fluoride	10mM	10 mL
proteases inhibitors Halt cocktail Thermo	-	10 mL
pyrophosphate	2,5mM	250 mL
orthovanadate	1m	10 mL

Los pellets de proteína fueron solubilizados en 200 $\mu$ L de urea 6M y sometidos al centro de genómica *UC Davis Proteomics Core*. Para la **digestión de las proteínas** se añadió ditiotritol a 200 mM hasta una concentración final de 15mM, estas muestras fueron incubadas por 30 minutos a 37°C. Luego se añadió iodoacetamida a 20 mM hasta una concentración final de 15 mM y fue incubado por 30 minutos a temperatura ambiente, seguido de otros 20  $\mu$ L de ditiotritol para aplicar iodoacetamida. Posteriormente, se añadió Lys-c a la muestra, que fue incubada durante 2 horas a 30°C. Se continuó con una dilución de las muestras a más de 1 M de urea mediante la adición de AMBIC a 50mM y se agregó tripsina, para dejar las muestras durante la noche a 37°C. Al día siguiente, las muestras se desalinizan usando la columna de Macro Spin (Nest Group).

Los péptidos digeridos fueron **analizados por LC-MS/MS** en una espectrómetro de masas *Thermo scientific Q-Exactive Orbitrap* junto con *Easy-nLC II HPLC* de Proxeon (Thermo Scientific) y una fuente *nanospray* de Proxeon. Las muestras se desalinizaron antes de ser separadas empleando una columna de fase inversa de 75  $\mu$ m x 150 mm (Magic C18 200Å 3U), y se pasaron a una trampa de fase inversa de 100- $\mu$ m x 25 mm (Magic C18 100Å 3U). Los péptidos se eluyeron utilizando un gradiente de 180 minutos con un caudal de 300 nL/min, obteniendo un espectro entre m/z 300-1600, se obtuvo el espectro MS/MS utilizando un método de los top 15, donde los 15 iones principales del espectro se sometieron a un HCD (High Energy Collisional Dissociation). Al seleccionar los iones precursores, se utilizó una ventana de masa de aislamiento de 2,0 m/z, y la fragmentación se hizo con un 27% de la energía de colisión normalizada. Para la exclusión dinámica se usó una duración de 22 segundos.

En el **procesamiento de los archivos** resultantes se usaron los siguientes parámetros: máximo 2 clivajes faltantes, mínimo 50% en la identificación de precursores de iones y búsqueda de péptidos razer y únicos. Las proteínas válidas se obtuvieron de la base de datos humana de SwissProt. Usando el programa *proteome discoverer 2.3*, se cuantifican sin etiqueta en el motor de búsqueda *Sequest* y *AMANDA*. Además, para identificar las proteínas válidas se usó *MaxQuant* y *Perseus*.

Por último, para obtener el **conteo relativo**, se importaron las intensidades de proteínas no normalizadas en R versión 4.0.1 (R Core Team, 2019). Primero, los datos fueron transformados con  $\log_2$  para obtener una distribución más simétrica antes del análisis y se mantuvieron todas las proteínas con un 70 % de valores válidos por grupo, es decir 6 réplicas por grupo (Karpievitch et al., 2012). Luego, se normalizaron mediante la estabilización de varianza VSN, con el objetivo de mantener la varianza constante en todo el rango de datos y se aproxima al logaritmo de valores grandes para eliminar la heterocedasticidad empleando el seno hiperbólico inverso (Huber et al., 2002). Finalmente, la imputación de los valores faltantes se realizó mediante *K-Nearest Neighbor Imputation* (KNN) que identifica  $k$  características muy similares a las proteínas con valores faltantes, donde la similitud se estima mediante la medida de distancia euclidiana, y los valores faltantes se imputan con los valores del promedio ponderado de estas proteínas vecinas (Chai et al., 2014; Välikangas et al., 2017), para esto se usó un valor KNN de 10 como lo sugiere Välikangas y colaboradores (2017).

## 3.2. Estrategia de integración de datos ómicos

### 3.2.1. Definición de la estrategia de contextualización

Se realizó una revisión bibliográfica, en la que se buscó recopilar los algoritmos de contextualización de modelos metabólicos a escala genómica por medio de datos multi-ómicos. Donde se identificaron las fortalezas y debilidades de cada método. Además, esta búsqueda se limitó solo a los algoritmos que usan datos transcriptómicos y proteómicos pues el objetivo del estudio es la integración de estos.

Para la recopilación de la literatura, se utilizó el método propuesto por Kitchenham B. y colaboradores (2009), cuyo propósito es integrar toda la literatura disponible de un tema y hacer un filtrado sistemático de esta. Asegurando así la repetibilidad del proceso y una buena cobertura del área de interés. Entonces, se plantearon las siguientes tres preguntas orientadoras para guiar la búsqueda: ¿Qué estrategias existen para contextualizar GEMs? ¿Qué datos soportan estas estrategias? ¿Cuáles son sus características principales?.

Con el fin de resolver estas preguntas, se buscaron recursos bibliográficos disponibles en bases de datos estructuradas y no estructuradas. En principio se obtienen 68 recursos bibliográficos, que fueron filtrados mediante lecturas de primer y segundo nivel, lo que permitió descartar los artículos poco relevantes. Como resultado se obtuvieron 27 recursos bibliográficos que fueron usados para la construcción del apartado [sec:integracion-de-datos-omicos-en-modelos-metabolicos-a-escala-genomica]**Integración de datos ómicos en modelos metabólicos a escala genómica.**

### 3.2.2. Mapeo de los datos ómicos al modelo

Las reglas GPR están presentes dentro de Recon3D y representan las asociaciones entre genes, proteínas y reacciones. Estas usan expresiones booleanas para codificar relaciones no lineales entre estos tres niveles biológicos (Bernstein D. et al., 2021), siendo útiles para cruzar datos experimentales con el modelo. Entonces, con el objetivo de asociar los datos del proteoma y transcriptoma con el modelo, se usaron estas reglas para mapear los datos ómicos a las reacciones. En este sentido, cuando hay más de un valor asociado con una reacción, se toma el valor menor cuando la relación es de tipo AND y el mayor cuando es de tipo OR.

Como resultado de este proceso se obtienen dos vectores  $\mathbb{R}^n$  para cada réplica, donde  $n$  es el número de reacciones. Los datos del proteoma mapeados a las reacciones, representan la abundancia de las enzimas que catalizan la reacción. Mientras que el transcriptoma mapeado a las reacciones representa el nivel de expresión de los genes necesarios para sintetizar dichas enzimas.

### 3.2.3. Reducción a una dimensión

Según el dogma central de la biología, la abundancia de transcritos y proteínas son equivalentes, ya que las proteínas se producen a partir de los transcritos y, por lo tanto, dependen de la concentración de estos. Varios estudios han mostrado una correlación lineal a nivel experimental entre estos dos datos (Ghazalpour A. et al., 2011; Stare T. et al., 2017; Våremo L. et al., 2015). Donde la variabilidad entre estos dos datos relaciona mayormente a los procesos postranscripcionales y en una menor medida a la variabilidad técnica y biológica (Ludvigsen M. et al., 2018). En este sentido, el razonamiento detrás de la reducción dimensional es obtener un valor que incluya la información que explica la relación lineal entre ambas medidas, disminuyendo las falencias del proteoma por sus sesgos y baja sensibilidad, y del transcriptoma por su distancia con el fenómeno que realmente se quiere medir, es decir la tasa a la que la reacción se está produciendo.

Al combinar el transcriptoma y proteoma se usó una el análisis de componentes principales (PCA) como un acercamiento de reducción de dimensionalidad, con el propósito de extraer la información relevante de los datos, se proyectaron en un espacio de menor dimensión (Guebel & Torres, 2013). Entonces, cómo se combinaron variables de unidades diferentes, se normalizaron los datos, ya que el PCA requiere que las variables tengan una media de 0 y una varianza igual a 1. De no ser así, se puede afectar la construcción de la matriz de covarianza usada para construir las componentes principales (Jolliffe & Cadima, 2016). Esto es una característica no deseada, por lo que para dicha normalización se usó la fórmula propuesta por Jolliffe & Cadima (2016), donde se le resta representa la media  $media(x)$  a cada valor  $x_i$  y se divide entre la desviación estándar  $\sqrt{Var(x)}$ , que es definida como la raíz de la varianza.

$$z_i = \frac{x_i - \text{media}(x)}{\sqrt{\text{Var}(x)}} \quad (3-1)$$

Luego, se aplicó la función *pca* que viene por defecto en MATLAB, que calcula la matriz de covarianza multiplicando la matriz de varianzas y covarianzas así:  $Z^T Z$ , lo que permite observar cómo se relacionan las variables entre sí, luego calcula los valores y vectores propios para crear los componentes principales definidos como  $Sa = \lambda a$  donde  $S$  es la matriz de covarianza, mientras  $\lambda$  es el vector propio y  $a$  el valor propio. Los vectores propios se construyen de manera que no se correlacionen entre sí permitiendo que la información entre dos variables que tiene una correlación linear sean representadas en un solo componente. Por lo tanto, al tomar el primer componente, se obtiene la variable que representa la mayor parte de la varianza presente en los datos iniciales.

### 3.3. Contextualización del modelo con datos de transcriptómica y proteómica.

Para determinar el algoritmo de contextualización, como primer criterio se tuvieron en cuenta sólo las aproximaciones que aceptan datos transcriptómicos y proteómicos, lo que reduce la lista a cinco algoritmos. Donde MOMA (M. Kim et al., 2016) es específico para *E. Coli* y el método avanzado de integración propuesto Ebrahim y colaboradores (2016), requiere de un perfilado ribosomal y datos fluxomicos; por lo que estos dos quedaron descartados. Mientras que GIMME (Becker & Palsson, 2008) y GIMMEp (Bordbar et al., 2012) requieren de una única función metabólica asumida, por lo que también se descartaron, ya que el astrocito cumple más de una sola función y usar estos algoritmos hubiese segado el proceso de contextualización. Entonces, el método seleccionado para la contextualización es iMAT (Zur et al., 2010) que mediante dos umbrales determina estados de expresión alto, medio y bajo, usados para identificar las reacciones que deben estar presentes en el modelo.

Como el método empleado para la contextualización del modelo usa umbrales, primero se aplicó una transformación logarítmica, para asegurar que los datos resultantes de la combinación por PCA tengan una distribución normal. Luego, para obtener el modelo tejido específico de astrocito humano se realizó una contextualización usando como base el modelo Recon3D (Brunk E. et al., 2018). A este modelo, se le aplicaron los datos resultantes de la transformación logarítmica con la ayuda del algoritmo iMAT (Zur et al., 2010) implementado en COBRA Toolbox V3.0 (Heirendt et al., 2018) a través de la función *createTissueSpecificModel*. Este algoritmo busca un balance entre la inclusión de reacciones con una expresión alta y la eliminación de las que tienen una expresión baja (Zur et al., 2010). Respecto al umbral bajo que determina qué reacciones están inactivas se definió como -500 y un umbral

alto para las reacciones activas que se definió como 500.

Por último, se utilizó *exp2flux* (Osorio et al., 2016) para producir tres modelos específicos de astrocito sano, inflamado y tratado con tibolona, utilizando los datos generados a partir de una reducción dimensional con los datos específicos de cada escenario con el método descrito en el numeral anterior por PCA.

### 3.3.1. Curación del modelo

#### Objetivo metabólico

Los objetivos metabólicos definidos para este modelo fueron determinados con base en las funciones metabólicas definidas en la literatura para los astrocitos. A saber: (i) La reacción genérica de biomasa humana de Recon3D (Brunk E. et al., 2018), usada para determinar la tasa de crecimiento. Junto con otras funciones específicas de los astrocitos, como lo es la liberación de gliotransmisores como (ii) ATP de las rutas energéticas, (iii) D-serina a partir de la glicina, (iv) glutamato generado de la glutamina obtenida del espacio extracelular y (v) el lactato sintetizado a partir de glucosa (González et al., 2020; Osorio et al., 2020; Martin-Jimenez et al., 2017). (vi) Además de la captura de cisteína para producir glutatión reducido (Osorio et al., 2020).

### 3.3.2. Llenado de huecos

Ya que la generación de un modelo tejido específico, por definición elimina reacciones que no hacen parte del tejido, este proceso puede dejar metabolitos que se producen, pero no se consumen, o al revés, metabolitos que se consumen, pero no son producidos. A esto se le denomina huecos o gaps (en inglés) en la red metabólica. Por consiguiente, para eliminar estos huecos se usó el algoritmo *fastGapFill* (Thiele et al., 2014) disponible en COBRA Toolbox V3.0 (Heirendt et al., 2018), este usa bases de datos de reacciones bioquímicas universales, para llenar los huecos con reacciones estequiométricamente consistentes.

### 3.3.3. Consistencia estequiométrica

Por otro lado, para comprobar la consistencia del modelo, se buscaron metabolitos de fuga y sifón con la ayuda de la función *findMassLeaksAndSiphons* disponible en COBRA Toolbox V3.0 (Heirendt et al., 2018). Cabe aclarar que los metabolitos de fuga son aquellos que se producen cuando todas las entradas a la célula están cerradas, es decir que se producen “de la nada”. Mientras los de sifón son aquellos que se consumen cuando todas las entradas están cerradas, por consiguiente se consumen de “la nada”. Esto por supuesto implica una violación a la ley de conservación de masa (Heirendt et al., 2018).

Ya que estos metabolitos aparecen en el modelo por inconsistencias estequiométricas, ciclos de tipo III o ambos, se aplicó un proceso en cuatro pasos para eliminarlos:

1. Primero se busco el grupo mínimo de reacciones responsables de cada metabolito de sifón o de fuga usando la función *findMinimalLeakageModeRxn* implementada en el COBRA Toolbox V3.0 (Heirendt et al., 2018).
2. Luego se revisó si alguna de las reacciones encontradas en este subgrupo presenta un desbalance estequiométrico entre los productos y los reactivos.
3. Y por último, se implementó el toolbox *Stoichiometric tools* de MATLAB (Jeffrey K.,2021) que busca el espacio nulo en la matriz de elementos de cada reacción desbalanceada para determinar el balance estequiométrico correcto.

Una vez se aseguró el balance de todas las reacciones con metabolitos conocidos, algunos pares de reacciones que sintetizan-degradan un metabolito pueden generar fugas o sifones. Esto se debe a que estos metabolitos tienen una composición química desconocida en el modelo, como aquellos que tienen radicales no definidos, que están representados por una R o RFULL, incluso algunos tienen una composición química representada por una sola X. Para estos casos, es posible encontrar un balance entre el par de reacciones de síntesis-degradación. Por lo tanto, en el cuarto paso se realizó un proceso semi-automático en el que se busca el grupo mínimo de reacciones que generan una fuga o sifón, como en el paso uno y se busca la presencia de pares de reacciones síntesis-degradación en este subgrupo para balancearlos.

#### 3.3.4. Ambiente nutricional

Las restricciones de las reacciones de intercambio del modelo se establecieron teniendo en cuenta como medio de cultivo un medio DMEM enriquecido con FBS al 10 %. Donde, los metabolitos que componen DMEM y sus respectivas contracciones, fueron tomadas de lo reportado por HiMedia, que es la misma información en la que se basó Osorio et al. (2020) para su modelo, la tabla con esta información está disponible en el anexo 1. Por otro lado, ya que FBS es un medio natural, los metabolitos y sus concentraciones pueden variar, por lo que se hizo una recopilación de varias fuentes para generar la tabla de composición de este medio que está disponible en el Anexo 2. Una vez determinada esta información, se definieron los límites inferiores de las reacciones de intercambio como la concentración de los correspondientes.



### 3.4. Escenarios metabólicos

Finalmente, con el propósito de determinar la funcionalidad del modelo para comprobar que las predicciones sean consistentes con los datos experimentales y las predicciones de los modelos anteriores, se simularon 4 escenarios metabólicos: (i) un escenario sano en el que se simulan las condiciones normales del metabolismo en los astrocitos, con la entrada nutricional definida previamente en el apartado 3.3.4. El otro, (ii) es una condición de isquemia como la propuesta por Martín-Jimenez y colaboradores (2017), en la que se reduce progresivamente el consumo de oxígeno y glucosa, iniciando en una entrada de  $2.5 \text{ mM gDW}^{-1}\text{h}^{-1}$  hasta llegar a una secreción de  $0.5 \text{ mM gDW}^{-1}\text{h}^{-1}$  para ambos metabolitos. La secreción también se simuló, con el objetivo de ver el comportamiento cuando se exige una producción de estos metabolitos, ya que algunas funciones metabólicas se mantienen estables en ambientes anaeróbicos con glucosa.

Junto con otros dos escenarios metabólicos propuestos por Osorio y colaboradores (2020), en los que se simula (iii) la inflamación inducida con  $0.208 \text{ mM/g}$  de ácido palmítico. Y (iv) un pre-tratamiento a este escenario de inflamación con la adición de  $70\mu\text{M/g}$  de tibolona como lo reportó González-Giraldo y colaboradores (2019). Finalmente, los modelos bajo estos escenarios fueron optimizados usando la implementación FBA disponible en COBRA ToolBox con loopless COBRA (Schellenberger et al., 2011), para reducir la formación de bucles termodinámicamente inviables que puedan afectar la predicción.

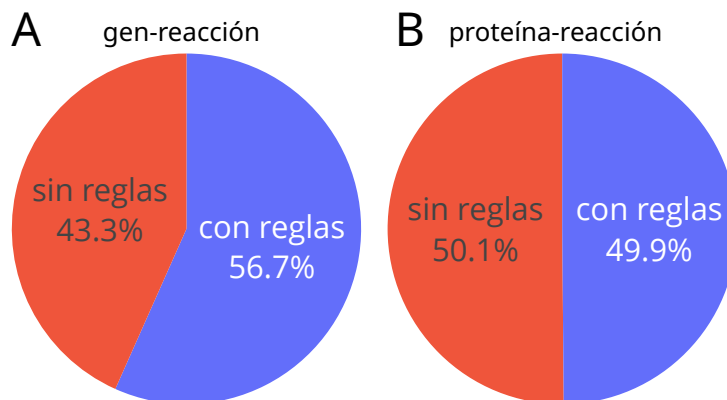
## 4. Capítulo 3: Resultados

El código y los datos usados en este trabajo están disponibles en el repositorio de github (<https://github.com/nmendozam/masterThesis>), los modelos resultantes de este proyecto, así como los otros modelos generados anteriormente por el grupo de investigación GIBBS están disponibles en el repositorio de github (<https://github.com/gibbslab/GSMs>), en este repositorio, los archivos que contienen los modelos desarrollados en este trabajo siguen la convención `Astrocyte_*_Mendoza2022.xml.gz` donde `*` representa el estado con el que se contextualizó el modelo.

### 4.1. Integración del proteoma y transcriptoma

Como COBRA ToolBox no soporta el mapeo de datos de abundancia proteica a un GEM, se desarrolló un paquete para añadir esta funcionalidad. Este consiste de 5 funciones implementadas en MATLAB y se encuentra disponible en GitHub para su libre uso y descarga (<https://github.com/gibbslab/GEM-multiomic-integrator>). Las funciones y una breve descripción de estas se encuentran a continuación:

1. *getProteinCodes*: Convierte los identificadores UniProt de proteínas a números de la comisión de enzimas (EC)
2. *makePrules*: Construye las reglas proteína-reacción utilizando los números EC asociados a las reacciones presentes en el modelo.
3. *PRparser*: Es un analizador sintáctico de las reglas proteína reacción para convertirlas en un formato que pueda ser leído por los métodos de contextualización implementados en COBRA ToolBox.
4. *findUsedLevels*: Con los identificadores del gen o proteína encuentra el nivel de expresión o abundancia para cada gen presente en el modelo. Esta es una extensión de la función con el mismo nombre implementada en COBRA ToolBox, para integrar las abundancias proteicas.
5. *mapAbundanceToReactions*: Mapea los valores de abundancia proteica obtenidos experimentalmente en las reacciones del modelo general. Usando la función *PRparser* para interpretar las reglas proteína-reacción y *findUsedLevels* para buscar los niveles de abundancia en los datos de entrada.



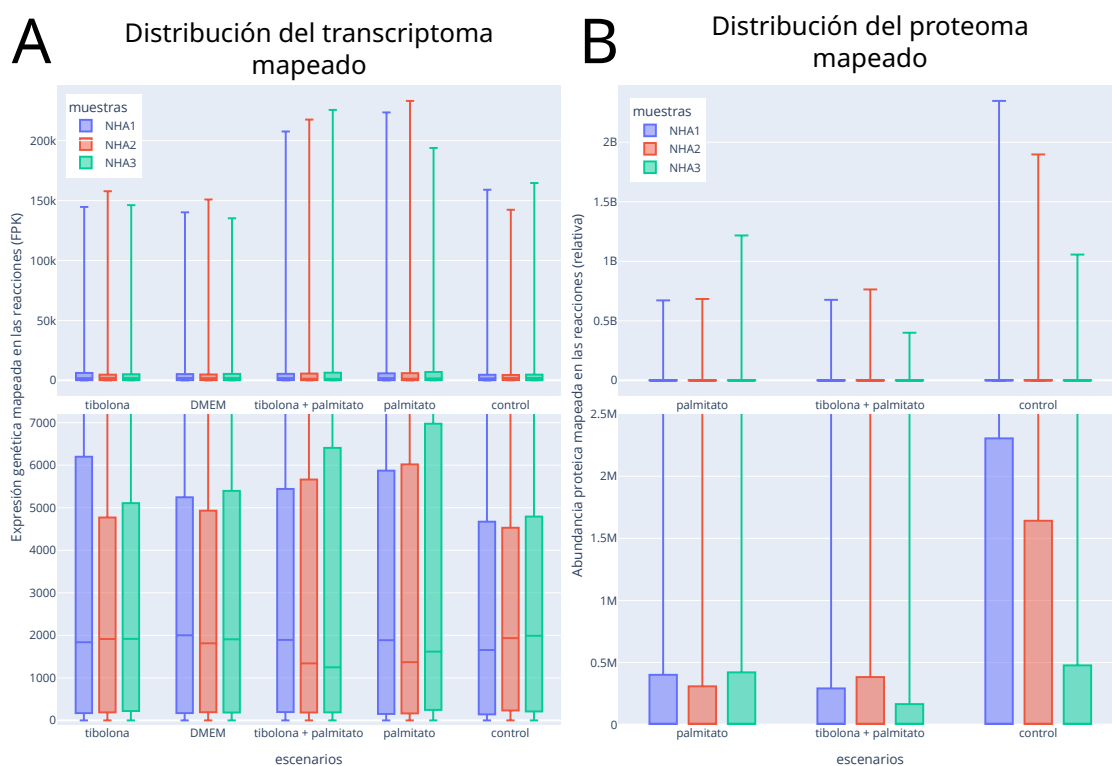
**Figura 4-1.:** Reacciones con reglas en Recon3D

Gráfico de torta mostrando la cantidad de reacciones que tienen una regla de asociación en Recon3D. A la derecha se muestra la gráfica para las reglas que relacionan los genes con las reacciones y a la izquierda las de las proteínas con las reacciones. Respecto al color, en naranja se ve el porcentaje de reacciones que no tienen una regla y en azul las que sí la tienen.

Entonces, con este paquete de funciones, más las implementadas en COBRA ToolBox, se mapean los datos de transcriptoma y proteoma a las reacciones del modelo, valiéndose de las reglas GPR presentes en Recon3D. Cabe destacar que solo el 56.7% de las reacciones las tienen por lo menos un **gen** asociado (Figura 4-1a), mientras que el 49.9% tienen por lo menos una **proteína** asociada (Figura 4-1b). Esto quiere decir que se pueden asociar datos experimentales con aproximadamente la mitad de las reacciones en el modelo.

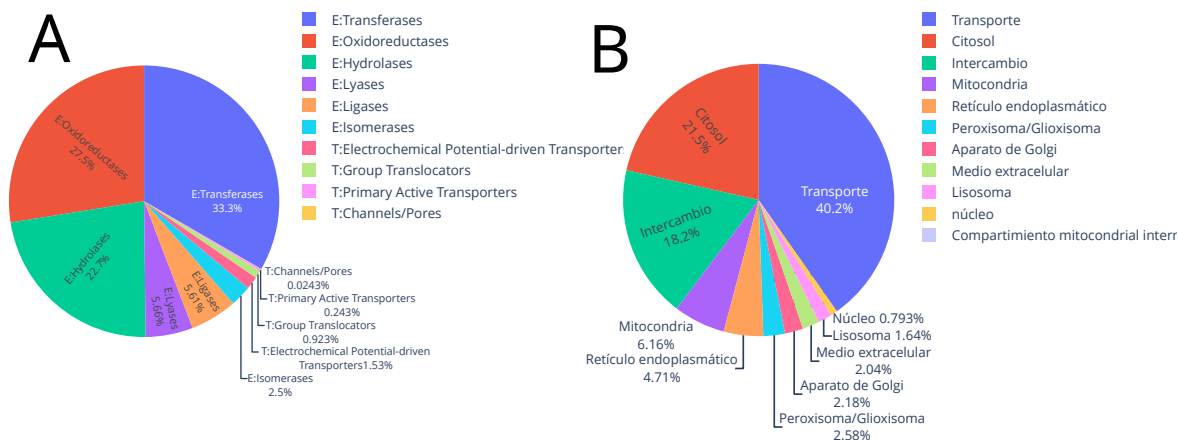
Luego, con la ayuda de las funciones *mapExpressionToReactions* (de COBRA Toolbox) y *mapAbundanceToReactions* se interpretan estas reglas para mapear los datos de transcriptoma y proteoma (respectivamente) a las reacciones correspondientes. Logrando cubrir un 97.76% de las reacciones que tienen reglas GPR con los datos del transcriptoma y un 99.81% de las que tienen reglas proteína-reacción con los datos del proteoma.

Una vez mapeados los datos se procedió a aplicar un análisis de componentes principales (PCA, por sus siglas en inglés) como técnica de reducción dimensional. De este proceso se obtuvo un vector tan largo como la cantidad de reacciones en COBRA. Como se puede ver en la figura 4-2, este dato presenta un sesgo a la izquierda y como se va a usar un algoritmo de contextualización de GEM basado en un umbral, se aplicó una transformación logarítmica, de manera que los datos presentaran una distribución uniforme.



**Figura 4-2.:** Distribución de los datos mapeados

En este diagrama de caja se muestra la distribución de los datos de transcriptómica (A) y proteómica (B) al ser mapeados en las reacciones del modelo. Las dos gráficas superiores muestran la distribución de los datos, mientras que las inferiores muestran la distribución de los tres primeros cuartiles.



**Figura 4-3.:** Distribución de las enzimas incluidas en el modelo de astrocito Clasificadas por (A) tipo de proteína aquellas que empiezan por “E:” son enzimas y por “T:” son proteínas de transporte. Y (B) las que están clasificadas por compartimento celular.

#### 4.1.1. Modelo tejido específico de astrocito

El modelo tejido específico de astrocito generado con el algoritmo iMAT a partir de los datos ómicos incluye 6.520 metabolitos y 10.586 reacciones. Donde 5.297 reacciones son reversibles y 5.289 irreversibles. Todo lo anterior distribuido en 9 compartimentos que incluye citosol [c], medio extracelular [e], aparato de Golgi [g], compartimiento mitocondrial interno [i], lisosoma [l], mitocondria [m], núcleo [n], retículo endoplasmático [r] y peroxisoma/glioxisoma [x].

En la figura 4-3a se puede ver la distribución de funciones enzimática en las reacciones del modelo que se obtuvieron usando el esquema de clasificación de los números EC para las enzimas y TCDB para las proteínas de transporte. De estas últimas sólo están presentes los primeros 4 grupos por lo que el modelo no tiene proteínas de transporte de tipo transportador de electrones transmembranal, ni factores de acceso relacionados con el transporte. Entonces, las funciones enzimáticas se distribuyen así: un canal o poro, 63 transportadores impulsados por potencial electroquímico, 10 transportadores activos primarios y 38 translocadores grupales por el lado de las proteínas de transporte, mientras las enzimas se componen de 1132 oxidoreductasas, 1371 transferasas, 933 hidrolasas, 233 liasas, 103 isomerases y 231 ligasas.

Por otro lado, 5880 reacciones no están en compartimentos (figura 4-3b), de las cuales 1834 son de intercambio y 4056 son de transporte. Adicionalmente, las 10084 reacciones que si están en compartimentos se dividen de la siguiente manera: 2167 están en el citosol, 621 en la mitocondria, 475 en el retículo endoplasmático, 269 en el peroxisoma / glioxisoma, 220 en el aparato de Golgi, 206 en el medio extracelular, 165 en el lisosoma y 80 en el núcleo. Respecto al compartimiento mitocondrial interno, este no tiene ninguna reacción, a pesar de tener varios metabolitos, que son transportados entre este compartimento y la mitocondria.

Una vez contextualizado el modelo se procedió a la curación, en este proceso se encontraron 2205 metabolitos de fuga y 2312 metabolitos de sifón. Para corregir los desbalances estequiométricos de una sola reacción y de pares de reacciones síntesis-degradación se creó una función llamada *fixLeaks*. Después de aplicar esta función y corregir todas las reacciones con un desbalance aparente, quedaron 0 metabolitos de fuga y 2312 de sifón. Lo que significa que la aproximación usada para curar el modelo es muy efectiva con las fugas de metabolitos, pero no tuvo efecto alguno en los metabolitos de sifón. Con el objetivo de reducir los problemas generados por esta falencia, se usó *loopless COBRA* (Schellenberger et al., 2011), para eliminar los ciclos termodinámicamente inviables que puedan estar generando la aparición de estos metabolitos sifón.

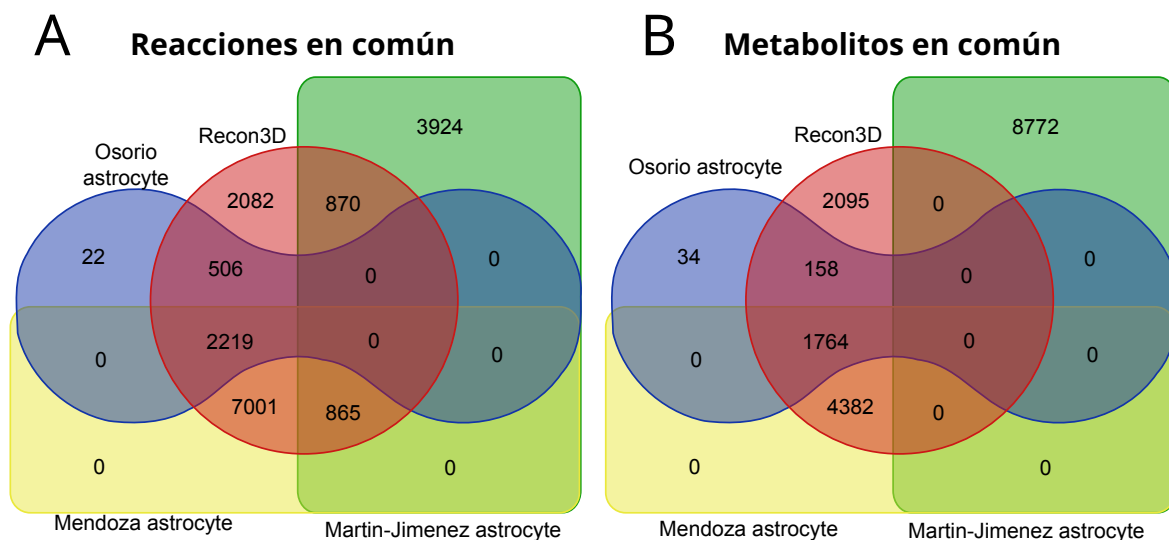
La composición química del escenario nutricional para el crecimiento del modelo y las concentraciones de cada metabolito para DMEM se definieron de acuerdo a lo reportado por Tejera y colaboradores (2020). Mientras que FBS, al ser un medio natural, no solo varían las concentraciones, sino también los metabolitos, incluso entre lotes de un mismo proveedor. Por lo tanto, se tomaron las concentraciones reportadas por Branzoi y colaboradores (2010) y se enriqueció con lo reportado por Yang y Xiong (2012) con el propósito de cubrir la mayor cantidad de metabolitos que puedan estar presentes en el medio cuando se usa FBS. El compendio de todos los metabolitos y sus respectivas concentraciones usados para como medio nutricional para este modelo están en la Tabla 5.

**Tabla 4-1.:** Composición química del medio de entrada nutricional del modelo de acuerdo a (Tejera et al., 2020; Yang & Xiong, 2012; Branzoi et al., 2010).

Metabolito	Concentración	Metabolito	Concentración
Glycine	-0.03 mM	L-Arginine	-846.7 $\mu$ M
L-Glutamine	-584 mM	L-Valine	-451.3 $\mu$ M
L-Isoleucine	-105 mM	L-Isoleucine	-415.2 $\mu$ M
L-Leucine	-105 mM	L-Leucine	-450.1 $\mu$ M
L-Methionine	-0.03 mM	L-Lysine	-624.1 $\mu$ M
L-Phenylalanine	-66 mM	L-Proline	-149.9 $\mu$ M
L-Serine	-42 mM	L-Threonine	-448.8 $\mu$ M
L-Threonine	-95 mM	L-Cystine	-130.2 $\mu$ M
L-Valine	-94 mM	L-Glutamine	-2.5 mM
7,8-Dihydrofolate	-4 mM	Hypoxanthine	-15.4 $\mu$ M
L-Alanine	-50 $\mu$ M	Lipoate	-508.9 $\mu$ M
Myo-Inositol	-72 mM	Triglyceride	-1.26 mM
D-Glucose	-4.5 mM	Phosphatidylserine	-0.72 mM
Myo-Inositol	-69.9 $\mu$ M	Cholesterol	-1.1 mM
Thymidine	-1.5 $\mu$ M	Oxygen	-10 mM
Riboflavin	-581.9 nM	Nitric Oxide	-10 mM
Pyridoxal	-150.7 nM	Cytidine	-1.3 $\mu$ M
L-Cysteine	-100.2 $\mu$ M	Uridine	-5.1 $\mu$ M
L-Asparagine	-56.8 $\mu$ M	Urate	-130 $\mu$ M
L-Tryptophan	-44.2 $\mu$ M	Hypoxanthine	-74.7 $\mu$ M
L-Serine	-249.8 $\mu$ M	Homocitrulline	-30.1 $\mu$ M
Nicotinamide Adenine Dinucleotide	-4 mM		

Con el propósito de comparar el modelo de astrocito generado en este trabajo con los existentes generados por Osorio et al. (2020) y Martín-Jimenez et al. (2017). Se compararon los identificadores de reacciones y metabolitos entre estos tres modelos y Recon 3D (Figura 4-4).

Para facilitar la comprensión, se nombraron los modelos como el apellido del autor principal de cada uno. Entonces, respecto a las reacciones (Figura 4-4a), el modelo Osorio (azul) y Martín-Jimenez (verde) no comparten ninguna debido a incompatibilidades entre los identificadores de las reacciones. Por otro lado, 22 reacciones se incluyen en Osorio pero no están en los otros modelos, también hay 3924 en Martín-Jimenez con esta misma característica. Por otro lado, Recon3D (rojo) comparte reacciones con todos los modelos con 2082 reacciones que no están presentes en ninguno de los modelos de astrocito, mientras otras 7001 reacciones que no se contemplaban en los otros modelos anteriores, se incluyeron en Mendoza (amarillo, el modelo en este trabajo), además de las 2219 que si comparte con Osorio y las otras 865 que



**Figura 4-4.:** Diagrama de venn de las interacciones entre los metabolitos y reacciones de los diferentes modelos.

Usando los identificadores de las reacciones y los metabolitos en los modelos de Recon3D (rojo), Mendoza (amarillo), Osorio (azul) y Martín-Jimenez (verde); se generó un diagrama de venn para mostrar las entidades que compartían estos modelos. A la izquierda (A) está la comparación de los identificadores de reacciones y a la derecha (B) la de los metabolitos.

comparte con Martín-Jimenez. Esto quiere decir que según los datos ómicos que se integraron, 870 de las reacciones que comparten Martín-Jimenez y Recon3D no estaban realmente activas en el astrocito, más otras 506 que están en Osorio y Recon3D que tampoco se incluyeron en Mendoza teniendo en cuenta los datos experimentales.

Respecto a la comparación entre los metabolitos de los modelos (Figura 4-4b), Martín-Jimenez no comparte ningún metabolito con los otros modelos a pesar de compartir reacciones, esto se debe a las notaciones diferentes de los metabolitos en este modelo. Por otro lado 2095 metabolitos de Recon3D y 34 de Osorio son exclusivos de cada modelo, y estos dos modelos comparten 158 metabolitos que no están en los demás. Mientras que Mendoza comparte 4382 solo con Recon3D y otros 1764 con este Recon3D y Osorio.

#### 4.1.2. Escenarios metabólicos

Luego de limitar los metabolitos de entrada según el ambiente nutricional previamente definido, se optimizó el modelo con FBA. Obteniendo una predicción de crecimiento  $0.33 \text{ mM gDW}^{-1}\text{h}^{-1}$  (tabla 6) para la reacción de biomasa de Recon3D. Dicha predicción es similar a los  $0.37 \text{ mM gDW}^{-1}\text{h}^{-1}$  obtenidos en el modelo anterior de astrocito (Osorio et al., 2020) y es más cercana a la medida experimental de  $0.32 \text{ mM gDW}^{-1}\text{h}^{-1}$  reportada por Prah y colaboradores (2019), la medida original fue convertida usando el método de conversión de conteo celular a

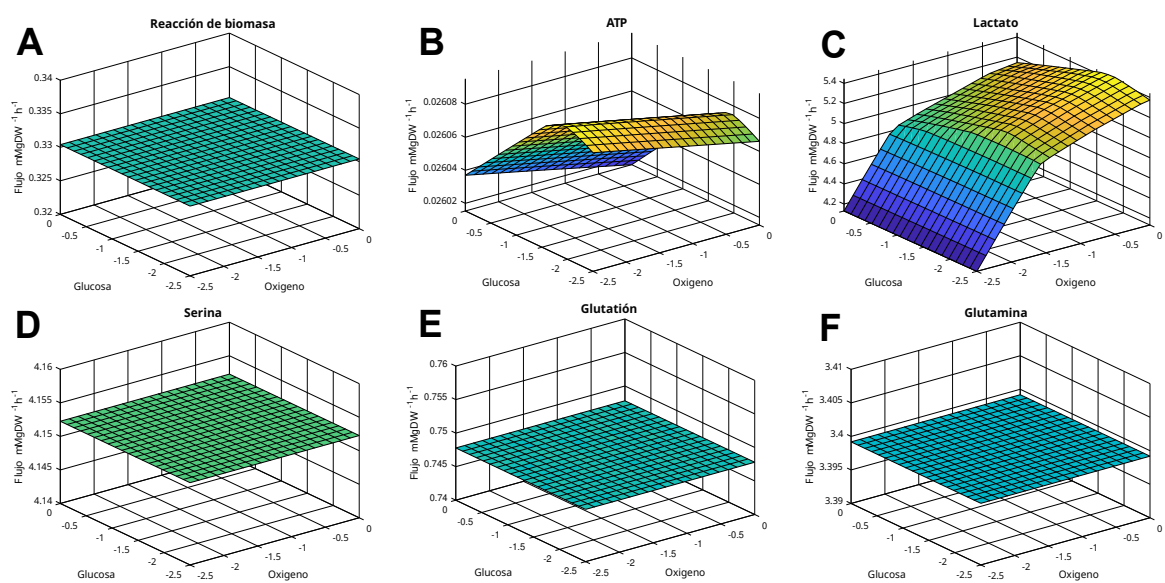


milimoles por peso seco para la evaluación de GEMs (Shinfuku et al., 2009).

### Escenario metabólico de Isquemia

Por otro lado, se simuló el estado de isquemia con un análisis de robustez al reducir el consumo de oxígeno y glucosa, para las seis funciones objetivo definidas. Como resultado se obtuvieron las seis superficies 3D que se muestran en la figura 4-5. En este sentido, la reacción de la biomasa no muestra una diferencia importante en las diferentes concentraciones de oxígeno y glucosa (Figura 4-5a), ya que los astrocitos tienen acceso a sustratos alternativos de glucólisis y respiración, a saber, glucógeno, lactato y ácidos grasos (Rossi et al., 2007). Sin embargo, la restricción de la captación de glucosa afecta la síntesis de ATP (Figura 4-5b), pero solo en 71,9 nM gDW<sup>-1</sup>h<sup>-1</sup>. Aunque se inhiba la respiración, la persistencia de la glucólisis genera lactato (Rossi et al., 2007), que aumenta con la reducción del consumo de oxígeno (Figura 4-5c). Además, se sabe que la D-serina contribuye al daño neuronal en condiciones isquémicas, pero no hay pruebas que muestren un aumento de la producción en estas condiciones (Rossi et al., 2007; Shen et al., 2021). Esto está en línea con la producción no influenciada de D-serina por la simulación de isquemia (Figura 4-5d). Además, como la ingesta de glutamato depende del ATP (Shen et al., 2021), la producción de glutamina a partir de glutamato tampoco se ve afectada (Figura 4-5f). Mientras el glutatión, un antioxidante que regula las ROS, tampoco se ve afectado por las restricciones de glucosa y oxígeno (Bragin et al., 2010).

Por el contrario, la secreción de glutamato (Figura 4-5e) es inversamente proporcional a la cantidad de oxígeno consumido por el astrocito, y no es afectada por las cantidades de glucosa. Mientras que la secreción de serina-L (Figura 4-5c) permanece entre 2.35 y 2.25 mM gDW<sup>-1</sup>h<sup>-1</sup> siempre y cuando los niveles de oxígeno o glucógeno se mantengan encima de 1.2 mM gDW<sup>-1</sup>h<sup>-1</sup>, esta condición se cumple hasta que se empieza a producir oxígeno lo que produce la pequeña inclinación que se ve a la derecha de la figura 4-5c. La producción de lactato por otro lado (Figura 4-5e), pasa de 2.8 mM gDW<sup>-1</sup>h<sup>-1</sup> en la condición de más alto consumo y llegando hasta 2.2 mM gDW<sup>-1</sup>h<sup>-1</sup> en ausencia de oxígeno y glucosa extracelular. Por su parte, la producción de glutatión, que es una molécula conocida por sus propiedades antioxidantes y además se ha reportado en literatura que aumenta en condiciones de isquemia (Bragin et al., 2010), muestra un aumento inversamente proporcional al consumo de oxígeno, empezando con un consumo de 0.6 mM gDW<sup>-1</sup>h<sup>-1</sup> en condiciones normales e independiente de la glucosa, para estabilizarse en una producción de 0.5 mM gDW<sup>-1</sup>h<sup>-1</sup> cuando los niveles de oxígeno están por debajo de 0.1 mM gDW<sup>-1</sup>h<sup>-1</sup>.



**Figura 4-5.:** Análisis de robustez para el escenario de isquemia.

En esta gráfica se muestra como varían las seis funciones objetivo en un escenario de isquemia, donde: (A) es la reacción de biomasa definida en Recon3D, la secreción de (B) ATP, (C) lactato, (D) serina, (E) glutatión y (F) glutamina. Las unidades de todos los ejes son milimoles de peso seco por hora ( $\text{mM gDW}^{-1} \text{h}^{-1}$ ), donde el eje vertical representa el flujo de la reacción objetivo respectiva, el horizontal derecho el de la reacción de intercambio de glucosa y el izquierdo representa el intercambio de oxígeno.

### Escenario de inflamación por Palmitato + Tibolona

Posteriormente se simuló el escenario de inflamación por palmitato y el de pre-tratamiento con tibolona, contrastando los resultados con los datos reportados por Osorio y colaboradores (2020) en las simulaciones con el modelo anterior. En la tabla 6 se pueden ver las predicciones para cada una de las funciones objetivo definidas para los escenarios simulados. Y como ya se había mencionado, las predicciones de crecimiento en un astrocito sano son similares para ambos modelos, al igual que predicciones de crecimiento en un escenario inflamatorio, pero, las similitudes acaban ahí.

**Tabla 4-2.:** Resultados de las funciones objetivo en estado sano, inflamado con palmitato y un pre-tratamiento con tibolona. Para cada una de las 6 funciones objetivo de los astrocitos planteadas por Osorio y colaboradores (2020). Donde, biomasa: indica el crecimiento, Glu2Gln: la conversión de glutamato a glutamina, Gly2SerD: la conversión de Glycina a D-serina, Glc2ATP: la producción de ATP a partir de glucosa y r0130: la obtención de cysteina para producir glutatión reducido.

Reacción	Flujo en mM gDW-1h-1				
	Sano	Palmitato	Tibolona	Sano (Osorio)	Palmitato (Osorio)
Biomasa	0.3304	0.2261	0.2072	0.3770	0.1838
Glu2Gln	5.8992	5.4353	5.4	3.1175	2.1303
Gly2SerL	4.1523	2.2992	2.2009	47.3709	45.0716
Glc2Lac	6.1128	3.629	3.5586	1.7908	0.9659
Glc2ATP	0.0261	0.0142	0.0141	1.8104	0.9697
r0130	0.7478	0.2839	0.3495	84.8271	84.4418

Por ejemplo, en la producción de glutamina el modelo de Osorio predice un flujo de 3.11 mM gDW<sup>-1</sup>h<sup>-1</sup> para un astrocito sano y 2.13 mM gDW<sup>-1</sup>h<sup>-1</sup> cuando está inflamado. Mientras que el modelo desarrollado en este trabajo predice un flujo de 5.89, 5.43 y 5.4 mM gDW<sup>-1</sup>h<sup>-1</sup> para sano, inflamado y tratado con tibolona respectivamente. Lo que implica una diferencia de 2.78 mM gDW<sup>-1</sup>h<sup>-1</sup> en el mejor de los casos, no obstante, un estudio computacional anterior indica que los astrocitos producen glutamina dependiendo de la fase sináptica, donde producen 1.5 mM pre-sinapsis, 5.0 mM durante la sinapsis y 10.0 mM post-sinapsis (Flanagan et al., 2018), lo que soporta las predicciones del actual modelo reconstruido. Además, Osorio reporta un flujo de 1.8104 y 0.9697 mM gDW<sup>-1</sup>h<sup>-1</sup> para sano e inflamado, lo que implica una diferencia de casi 70 veces el flujo generado en la simulación que es de 0.0261, 0.0142 y 0.0141 mM gDW<sup>-1</sup>h<sup>-1</sup> para sano, inflamado y tratado con tibolona. Sin embargo, se ha demostrado que las concentraciones intracelulares de ATP superiores a 0.5 mM son neurotóxicas (Amadio et al., 2005), entonces, teniendo en cuenta que la proporción de ATP respecto al flujo reportado es de 1 a 36, las predicciones de Osorio se vuelven inviables, lo

que muestra la efectividad de la aproximación multi ómica. Otros objetivo metabólico con una gran diferencia entre las predicciones de ambos modelos es la producción de Serina-L con 47.3709 y 45.0716 mM gDW<sup>-1</sup>h<sup>-1</sup> de flujo para sano e inflamado en Osorio y 4.1523, 2.2992 y 2.2009 mM gDW<sup>-1</sup>h<sup>-1</sup> para sano, inflamado y tratado con tibolona en el modelo reconstruido.

También está la secreción de glutatión a través de la reacción r0130 para la que se predijo un flujo de 0.7478, 0.2839 y 0.3495 mM gDW<sup>-1</sup>h<sup>-1</sup> para el estado sano, inflamado y tratado con tibolona respectivamente. Aunque la tasa de secreción de este metabolito no se ha medida experimentalmente hasta donde sabemos, el nivel extracelular de este metabolito ronda los 2.8 μM y en el citosol de astrocitos primarios se encuentra entre 8 y 10 mM (McBean, 2017), siendo la concentración más alta reportada para células en mamíferos, por lo que, aunque la evidencia no soporta los resultados obtenidos por el modelo, sin duda mejora las estimaciones de Osorio que rondan los valores de 84.8271 y 84.4418 mM gDW<sup>-1</sup>h<sup>-1</sup> para sano e inflamado. Por último, la producción de lactato se predijo con un flujo de 6.1128, 3.629 y 3.5586 mM gDW<sup>-1</sup>h<sup>-1</sup> para el estado sano, inflamado y tratado con tibolona respectivamente, mientras que Osorio reporta un flujo de 1.7908 y 0.9659 mM gDW<sup>-1</sup>h<sup>-1</sup> para sano e inflamado.

## 5. Capítulo 4: Discusión

En este trabajo se presenta una estrategia para la integración de datos de proteómica y transcriptómica con el propósito de ser usado en la reconstrucción y contextualización de GEMs. Dicha estrategia permitió generar el GEM de astrocito más completo hasta el momento en términos de cantidad de genes, proteínas y reacciones además de contar con la enorme ventaja de estar basado en datos multiómicos. Comparado con los modelos anteriores, este modelo abarca una gran parte del metabolismo considerado en dichos trabajos (Figura 4-4). Siendo una herramienta *in silico* útil para el estudio de las funciones metabólicas del astrocito y la generación de nuevas hipótesis que permitan dilucidar los procesos normales y patológicos de estas células.

En el modelo reconstruido en este trabajo, se lograron incluir las rutas metabólicas de trolona, ya que los metabolitos como la progesterona que interactúan estas rutas están presentes en el modelo. También, se tomaron en cuenta seis funciones objetivo que abarcan las principales funciones metabólicas que cumplen los astrocitos (Harada et al., 2016), que se incluyen en los modelos anteriores (Osorio et al., 2020; Martín-Jimenez et al., 2017), pero, se planearon al contrario, planteándose desde el punto de vista de la neurona, en donde las funciones objetivo consumen los metabolitos producidos por el astrocito, en vez de suplantar la red metabólica interna en el proceso de producción de estos metabolitos. Por ejemplo, para la conversión de glutamato a glutamina, la función objetivo se plantea como glutamina  $\rightarrow$  glutamato, donde la glutamina es producida por el astrocito y usada en el metabolismo de la neurona, que genera glutamato como residuo.

Por otro lado, el balance de las reacciones en los GEM demuestra ser un paso crucial en la reconstrucción de un modelo, sin embargo, no hay herramientas disponibles en la literatura que faciliten este proceso, siendo generalmente un proceso que se realiza de manera manual. Por lo cual, se desarrolló un algoritmo semiautomático para corregir los desequilibrios en las reacciones, que fue capaz de eliminar completamente los metabolitos de fuga. No obstante, dicho algoritmo es ineficaz para eliminar los metabolitos de sifón, por lo que se recomienda usar optimizadores que eviten los ciclos termodinámicamente inviables con el propósito de reducir los errores generados por estos metabolitos de sifón que aún están presentes en el modelo, por ejemplo, durante las pruebas de los escenarios metabólicos se usó la implementación de *loopless COBRA* (Schellenberger et al., 2011). Con todo, este tipo de optimizadores requieren cálculos más complejos y pueden limitar las aplicaciones del modelo, por lo que, en futuros trabajos se puede desarrollar un algoritmo de balanceo más complejo que sea eficaz con los

metabolitos de sifón.

En este sentido, al probar el funcionamiento del GEM de astrocito en varios ambientes metabólicos, éste muestra una capacidad predictiva mejorada respecto a los modelos anteriores, realizando predicciones más cercanas a lo reportado en la literatura, a pesar de los desbalances estequiométricos que no se lograron corregir. Demostrando entonces un resultado favorable para las dos estrategias propuestas, a saber: (i) la integración de datos de transcriptómica y proteómica usando PCA es una alternativa viable para generar modelos contexto específico basados en datos multiómicos. Y (ii) el algoritmo de balanceo de estequiométrico a pesar de no eliminar los metabolitos de sifón genera un modelo lo suficientemente resiliente como para generar predicciones viables.

Una fuente de error que hay que tener en cuenta cuando se use el modelo es que el ambiente nutricional contiene un medio natural (FBS), ya que su composición tiende a variar (Yang & Xiong, 2012), por lo que se puede esperar una diferencia entre las predicciones y los datos experimentales. Por lo tanto, en caso de que se quiera evaluar el modelo, se recomienda medir las concentraciones del medio nutricional usado y hacer los cambios respectivos en el modelo, o usar un medio de cultivo sintético y cambiar el medio nutricional de este.

## 6. Capítulo 5: Conclusiones

Finalmente, se obtiene un modelo metabólico a escala genómica de astrocito humano contextualizado con datos experimentales transcriptómicos y proteómicos. Que comparado con los modelos anteriores abarca una porción más amplia de la red metabólica del astrocito, cuyas predicciones son más cercanas a la realidad. Además de beneficiarse de las ventajas que trae Recon3D como los datos estructurales de metabolitos y proteínas. Por lo que el modelo presentado en este trabajo puede acelerar el estudio de la relación entre los astrocitos y enfermedades neurodegenerativas, mediante la generación de hipótesis, predicción de respuestas metabólicas y el uso de estrategia de análisis de redes para dilucidar procesos metabólicos. Adicionalmente, se presentaron dos metodologías nuevas, una para la integración de datos transcriptómicos y proteómicos en un GEM y la otra más orientada hacia la curación de un modelo ya construido que busca eliminar las inconsistencias termodinámicas y estequiométricas que se presentan en este tipo de modelos. Estos dos métodos pueden ser los cimientos de futuros trabajos que generen estrategias para reconstrucción, contextualización y curación de GEMs más robustos.

## A. Anexo: Composición DMEM

Nombre	Concentración	Nombre	Concentración
<i>SALES INORGÁNICAS</i>			
CaCl <sub>2</sub>	793.1 $\mu$ M	NaH <sub>2</sub> PO <sub>4</sub>	452.9 $\mu$ M
Fe(NO <sub>3</sub> ) <sub>3</sub> ·9H <sub>2</sub> O	123.8 nM	FeSO <sub>4</sub> ·7H <sub>2</sub> O	1.5 $\mu$ M
MgSO <sub>4</sub>	405.8 $\mu$ M	CuSO <sub>4</sub> ·5H <sub>2</sub> O	5.2 nM
KCl	4.2 mM	MgCl <sub>2</sub>	301.0 $\mu$ M
NaHCO <sub>3</sub>	14.3 mM	Na <sub>2</sub> H <sub>2</sub> PO <sub>4</sub>	400 $\mu$ M
NaCl	119.7 mM	ZnSO <sub>4</sub> ·7H <sub>2</sub> O	1.5 $\mu$ M
<i>VITAMINAS</i>			
Niacinamide	16.6 $\mu$ M	Folic Acid	6 $\mu$ M
Biotin	14.3 nM	D-Calcium Pantothenate	4.7 $\mu$ M
Thiamine Hcl	6.4 $\mu$ M	Thymidine	1.5 $\mu$ M
Myo-inositol	69.9 $\mu$ M	Riboflavin	581.9 nM
Pyridoxal	9.8 $\mu$ M	Vit B12	501.7 nM
Choline chloride	64.3 $\mu$ M	Pyridoxine Hcl	150.7 nM
<i>AMINOÁCIDOS</i>			
L-Cysteine	100.2 $\mu$ M	L-Arginine	846.7 $\mu$ M
L-Methionine	115.5 $\mu$ M	L-Valine	451.3 $\mu$ M
L-Glycine	249.8 $\mu$ M	L-Glutamic acid	50.0 $\mu$ M
L-Asparagine	56.8 $\mu$ M	L-Isoleucine	415.2 $\mu$ M
L-Tryptophan	44.2 $\mu$ M	L-Leucine	450.1 $\mu$ M
L-Histidine	150.2 $\mu$ M	L-Lysine	624.1 $\mu$ M
L-Phenylalanine	214.8 $\mu$ M	L-Proline	149.9 $\mu$ M
L-Tyrosine	213.6 $\mu$ M	L-Threonine	448.8 $\mu$ M
L-Aspartic acid	50.0 $\mu$ M	Cystine	130.2 $\mu$ M
L-Serine	249.8 $\mu$ M	Glutamine	2.5 mM
L-Alanine	50.0 $\mu$ M		
<i>OTROS</i>			
Na Pyruvate	0.5 mM	Putrescine	918.8 $\mu$ M
Hypoxanthine	15.4 $\mu$ M	DL-68-Lipoic acid	508.9 nM
Linoleic acid	149.7 nM		



# Bibliografía

A, S., W, J., Di, W., Dp, J., & X, D. (2018, mayo 1). ADAP-GC 3.2: Graphical Software Tool for Efficient Spectral Deconvolution of Gas Chromatography-High-Resolution Mass Spectrometry Metabolomics Data. *Journal of Proteome Research; J Proteome Res.* <https://doi.org/10.1021/acs.jproteome.7b00633>

Alonso, A., Marsal, S., & Julià, A. (2015). Analytical Methods in Untargeted Metabolomics: State of the Art in 2015. *Frontiers in Bioengineering and Biotechnology*, 3. <https://doi.org/10.3389/fbioe.2015.00023>

Amadio, S., D'Ambrosi, N., Trincavelli, M. L., Tuscano, D., Sancesario, G., Bernardi, G., Martini, C., & Volonté, C. (2005). Differences in the neurotoxicity profile induced by ATP and ATP  $\gamma$ S in cultured cerebellar granule neurons. *Neurochemistry International*, 47(5), 334–342. <https://doi.org/10.1016/j.neuint.2005.05.008>

Becker, S. A., & Palsson, B. O. (2008). Context-Specific Metabolic Networks Are Consistent with Experiments. *PLOS Computational Biology*, 4(5), e1000082. <https://doi.org/10.1371/journal.pcbi.1000082>

Beleites, C., & Sergo, V. (2020). hyperSpec: A package to handle hyperspectral data sets in R. <https://cran.r-project.org/web/packages/hyperSpec/hyperSpec.pdf>

Bernstein, D. B., Sulheim, S., Almaas, E., & Segrè, D. (2021). Addressing uncertainty in genome-scale metabolic model reconstruction and analysis. *Genome Biology*, 22(1), 64. <https://doi.org/10.1186/s13059-021-02289-z>

Bezzi, P., Domercq, M., Brambilla, L., Galli, R., Schols, D., De Clercq, E., Vescovi, A., Bagetta, G., Kollias, G., Meldolesi, J., & Volterra, A. (2001). CXCR4-activated astrocyte glutamate release via TNF $\alpha$ : Amplification by microglia triggers neurotoxicity. *Nature Neuroscience*, 4(7), 702-710. <https://doi.org/10.1038/89490>

---

Bianchi, M. G., Bardelli, D., Chiu, M., & Bussolati, O. (2014). Changes in the expression of the glutamate transporter EAAT3/EAAC1 in health and disease. *Cellular and Molecular Life Sciences: CMLS*, 71(11), 2001-2015. <https://doi.org/10.1007/s00018-013-1484-0>

Bleu Knight, V., & Serrano, E. E. (2017). Hydrogel scaffolds promote neural gene expression and structural reorganization in human astrocyte cultures. *PeerJ*, 2017(1). Scopus. <https://doi.org/10.7717/peerj.2829>

Böcker, S., Letzel, M. C., Lipták, Z., & Pervukhin, A. (2009). SIRIUS: Decomposing isotope patterns for metabolite identification. *Bioinformatics*, 25(2), 218. <https://doi.org/10.1093/bioinformatics/btn603>

Bond, N. J., Shliaha, P. V., Lilley, K. S., & Gatto, L. (2013). Improving Qualitative and Quantitative Performance for MSE-based Label-free Proteomics. *Journal of Proteome Research*, 12(6), 2340-2353. <https://doi.org/10.1021/pr300776t>

Bragin, D. E., Zhou, B., Ramamoorthy, P., Müller, W. S., Connor, J. A., & Shi, H. (2010). Differential changes of glutathione levels in astrocytes and neurons in ischemic brains by two-photon imaging. *Journal of Cerebral Blood Flow and Metabolism: Official Journal of the International Society of Cerebral Blood Flow and Metabolism*, 30(4), 734–738. <https://doi.org/10.1038/jcbfm.2010>

Branzoi, I. V., Iordoc, M., Branzoi, F., Vasilescu-Mirea, R., & Sbarcea, G. (2010). Influence of diamond-like carbon coating on the corrosion resistance of the NITINOL shape memory alloy. *Surface and Interface Analysis*, 42(6–7), 502–509. <https://doi.org/10.1002/sia.3473>

Brunk, E., Sahoo, S., Zielinski, D. C., Altunkaya, A., Dräger, A., Mih, N., Gatto, F., Nilsson, A., Preciat Gonzalez, G. A., Aurich, M. K., Prlić, A., Sastry, A., Danielsdottir, A. D., Heinken, A., Noronha, A., Rose, P. W., Burley, S. K., Fleming, R. M. T., Nielsen, J., . . . Palsson, B. O. (2018). Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nature Biotechnology*, 36(3), 272–281. <https://doi.org/10.1038/nbt.4072>

Ca, S., G, O., Ej, W., C, Q., Sa, T., Tr, B., De, C., R, A., & G, S. (2005, diciembre). METLIN: A metabolite mass spectral database. *Therapeutic Drug Monitoring; Ther Drug Monit.* <https://doi.org/10.1097/01.ftd.0000179845.53213.39>

Carroll, A. J., Badger, M. R., & Harvey Millar, A. (2010). The Metabolome-Express Project: Enabling web-based processing, analysis and transparent dissemination of GC/MS metabolomics datasets. *BMC Bioinformatics*, 11(1), 376. <https://doi.org/10.1186/1471-2105-11-376>

Carson, M. J., Doose, J. M., Melchior, B., Schmid, C. D., & Ploix, C. C. (2006). CNS immune privilege: Hiding in plain sight. *Immunological Reviews*, 213, 48. <https://doi.org/10.1111/j.1600-065X.2006.00441.x>

Chai, L. E., Law, C. K., Mohamad, M. S., Chong, C. K., Choon, Y. W., Deris, S., & Illias, R. M. (2014). Investigating the Effects of Imputation Methods for Modelling Gene Networks Using a Dynamic Bayesian Network from Gene Expression Data. *The Malaysian Journal of Medical Sciences : MJMS*, 21(2), 20.

Chambers, M. C., Maclean, B., Burke, R., Amodei, D., Ruderman, D. L., Neumann, S., Gatto, L., Fischer, B., Pratt, B., Egertson, J., Hoff, K., Kessner, D., Tasman, N., Shulman, N., Frewen, B., Baker, T. A., Brusniak, M.-Y., Paulse, C., Creasy, D., ... Mallick, P. (2012). A cross-platform toolkit for mass spectrometry and proteomics. *Nature Biotechnology*, 30(10), 918-920. <https://doi.org/10.1038/nbt.2377>

Chong, J., Soufan, O., Li, C., Caraus, I., Li, S., Bourque, G., Wishart, D. S., & Xia, J. (2018). MetaboAnalyst 4.0: Towards more transparent and integrative metabolomics analysis. *Nucleic Acids Research*, 46(W1), W486-W494. <https://doi.org/10.1093/nar/gky310>

Conesa, A. (2014). The STATegra project: New statistical tools for analysis and integration of diverse omics data. *EMBnet.Journal*, 20(A), 768. <https://doi.org/10.14806/ej.20.A.768>

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M. W., Gaffney, D. J., Elo, L. L., Zhang, X., & Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biology*, 17, 13. <https://doi.org/10.1186/s13059-016-0881-8>

Costa, C., Maraschin, M., & Rocha, M. (2016). An R package for the integrated analysis of metabolomics and spectral data. *Computer Methods and Programs in Biomedicine*, 129, 117-124. <https://doi.org/10.1016/j.cmpb.2016.01.008>

---

D, S., I, S., Ne, I., A, H., & Cp, P. (2014, febrero). Sequencing depth and coverage: Key considerations in genomic analyses. *Nature Reviews. Genetics*; *Nat Rev Genet.* <https://doi.org/10.1038/nrg3642>

Deutsch, E. W., Albar, J. P., Binz, P.-A., Eisenacher, M., Jones, A. R., Mayer, G., Omenn, G. S., Orchard, S., Vizcaíno, J. A., & Hermjakob, H. (2015). Development of data representation standards by the human proteome organization proteomics standards initiative. *Journal of the American Medical Informatics Association : JAMIA*, 22(3), 495-506. <https://doi.org/10.1093/jamia/ocv001>

Diaz-Ortiz, M. E., & Chen-Plotkin, A. S. (2020). Omics in neurodegenerative disease: Hope or hype? *Trends in Genetics: TIG*, 36(3), 152. <https://doi.org/10.1016/j.tig.2019.12.002>

Ds, W., Yd, F., A, M., Ac, G., K, L., R, V.-F., T, S., D, J., C, L., N, K., Z, S., E, L., N, A., M, B., S, S., D, A., Y, L., H, B., J, G., ... A, S. (2018, abril 1). HMDB 4.0: The human metabolome database for 2018. *Nucleic Acids Research; Nucleic Acids Res.* <https://doi.org/10.1093/nar/gkx1089>

Ebrahim, A., Brunk, E., Tan, J., O'Brien, E. J., Kim, D., Szubin, R., Lerman, J. A., Lechner, A., Sastry, A., Bordbar, A., Feist, A. M., & Palsson, B. O. (2016). Multi-omic data integration enables discovery of hidden biological regularities. *Nature Communications*, 7(1), 13091. <https://doi.org/10.1038/ncomms13091>

F, N., Y, G., Ea, C., S, M., L, L., & M, H.-M. (2012, enero 9). MSeasy: Unsupervised and untargeted GC-MS data processing. *Bioinformatics (Oxford, England); Bioinformatics.* <https://doi.org/10.1093/bioinformatics/bts427>

Fang, X., Wallqvist, A., & Reifman, J. (2012). Modeling Phenotypic Metabolic Adaptations of *Mycobacterium tuberculosis* H37Rv under Hypoxia. *PLOS Computational Biology*, 8(9), e1002688. <https://doi.org/10.1371/journal.pcbi.1002688>

Fang, X., Lloyd, C. J., & Palsson, B. O. (2020). Reconstructing organisms in silico: Genome-scale models and their emerging applications. *Nature Reviews Microbiology*, 18(12), 731–743. <https://doi.org/10.1038/s41579-020-00440-4>

Feigin, V. L., Nichols, E., Alam, T., Bannick, M. S., Beghi, E., Blake, N., Culpepper, W. J., Dorsey, E. R., Elbaz, A., Ellenbogen, R. G., Fisher, J. L., Fitzmaurice, C., Giussani, G., Glennie, L., James, S. L., Johnson, C. O., Kassebaum, N. J., Logroscino, G., Marin, B., ... Vos, T. (2019). Global, regional, and national burden of neurological disorders, 1990–2016: A systematic analysis for the Global Burden of Disease Study 2016. *The Lancet Neurology*, 18(5), 459-480. [https://doi.org/10.1016/S1474-4422\(18\)30499-X](https://doi.org/10.1016/S1474-4422(18)30499-X)

Fiehn, O., Robertson, D., Griffin, J., van der Werf, M., Nikolau, B., Morrison, N., Sumner, L. W., Goodacre, R., Hardy, N. W., Taylor, C., Fostel, J., Kristal, B., Kaddurah-Daouk, R., Mendes, P., van Ommen, B., Lindon, J. C., & Sansone, S.-A. (2007). The metabolomics standards initiative (MSI). *Metabolomics*, 3(3), 175-178. <https://doi.org/10.1007/s11306-007-0070-6>

Finkel, L., Millán Arroyo, Crespo, C., & Garcés, M. (2016). Estudio sobre las enfermedades neurodegenerativas en España y su impacto económico y social. <https://doi.org/10.13140/RG.2.1.3269.8009>

Flanagan, B., McDaid, L., Wade, J., Wong-Lin, K., & Harkin, J. (2018). A computational study of astrocytic glutamate influence on post-synaptic neuronal excitability. *PLOS Computational Biology*, 14(4), e1006040. <https://doi.org/10.1371/journal.pcbi.1006040>

Fondi, M., & Liò, P. (2015). Multi -omics and metabolic modelling pipelines: Challenges and tools for systems microbiology. *Microbiological Research*, 171, 52-64. <https://doi.org/10.1016/j.micres.2015.01.003>

Gatto, L. (2020, abril 27). Human Protein Atlas in R. Bioconductor. <https://www.bioconductor.org/packages/release/bioc/vignettes/hpar/inst/doc/hpar.html>

Gatto, L., & Christoforou, A. (2014). Using R and Bioconductor for proteomics data analysis. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1844(1, Part A), 42-51. <https://doi.org/10.1016/j.bbapap.2013.04.032>

Gevorgyan, A., Poolman, M. G., & Fell, D. A. (2008). Detection of stoichiometric inconsistencies in biomolecular models. *Bioinformatics*, 24(19), 2245–2251. <https://doi.org/10.1093/bioinformatics/btn425>

---

Ghazalpour, A., Bennett, B., Petyuk, V. A., Orozco, L., Hagopian, R., Mungrue, I. N., Farber, C. R., Sinsheimer, J., Kang, H. M., Furlotte, N., Park, C. C., Wen, P.-Z., Brewer, H., Weitz, K., Camp, D. G., Pan, C., Yordanova, R., Neuhaus, I., Tilford, C., ... Lusic, A. J. (2011). Comparative analysis of proteome and transcriptome variation in mouse. *PLoS Genetics*, 7(6), e1001393. <https://doi.org/10.1371/journal.pgen.1001393>

González-Giraldo, Y., Forero, D. A., Echeverria, V., Garcia-Segura, L. M., & Barreto, G. E. (2019). Tibolone attenuates inflammatory response by palmitic acid and preserves mitochondrial membrane potential in astrocytic cells through estrogen receptor beta. *Molecular and Cellular Endocrinology*, 486, 65–78. <https://doi.org/10.1016/j.mce.2019.02.017>

González, J., Pinzón, A., Angarita-Rodríguez, A., Aristizabal, A. F., Barreto, G. E., & Martín-Jiménez, C. (2020). Advances in Astrocyte Computational Models: From Metabolic Reconstructions to Multi-omic Approaches. *Frontiers in Neuroinformatics*, 14. Scopus. <https://doi.org/10.3389/fninf.2020.00035>

Goodrich, G. S., Kabakov, A. Y., Hameed, M. Q., Dhamne, S. C., Rosenberg, P. A., & Rotenberg, A. (2013). Ceftriaxone treatment after traumatic brain injury restores expression of the glutamate transporter, GLT-1, reduces regional gliosis, and reduces post-traumatic seizures in the rat. *Journal of Neurotrauma*, 30(16), 1434-1441. <https://doi.org/10.1089/neu.2012.2712>

Gu, C., Kim, G. B., Kim, W. J., Kim, H. U., & Lee, S. Y. (2019). Current status and applications of genome-scale metabolic models. *Genome Biology*, 20. <https://doi.org/10.1186/s13059-019-1730-3>

Guebel, D. V., & Torres, N. V. (2013). Principal Component Analysis (PCA). In W. Dubitzky, O. Wolkenhauer, K.-H. Cho, & H. Yokota (Eds.), *Encyclopedia of Systems Biology* (pp. 1739–1743). Springer. [https://doi.org/10.1007/978-1-4419-9863-7\\_1276](https://doi.org/10.1007/978-1-4419-9863-7_1276)

Haas, R., Zelezniak, A., Iacovacci, J., Kamrad, S., Townsend, S., & Ralser, M. (2017). Designing and interpreting ‘multi-omic’ experiments that may change our understanding of biology. *Current Opinion in Systems Biology*, 6, 37-45. <https://doi.org/10.1016/j.coisb.2017.08.009>

Hamby, M. E., Coppola, G., Ao, Y., Geschwind, D. H., Khakh, B. S., & Sofroniew, M. V. (2012). Inflammatory mediators alter the astrocyte transcriptome and calcium signaling elicited by multiple G-protein-coupled receptors. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 32(42), 14489-14510. <https://doi.org/10.1523/JNEUROSCI.1256-12.2012>

Harada, K., Kamiya, T., & Tsuboi, T. (2016). Gliotransmitter release from astrocytes: Functional, developmental and pathological implications in the brain. *Frontiers in Neuroscience*, 9. <https://www.frontiersin.org/article/10.3389/fnins.2015.00499>

Heirendt, L., Arreckx, S., Pfau, T., Mendoza, S. N., Richelle, A., Heinken, A., Haraldsdóttir, H. S., Wachowiak, J., Keating, S. M., Vlasov, V., Magnusdóttir, S., Ng, C. Y., Preciat, G., Žagare, A., Chan, S. H. J., Aurich, M. K., Clancy, C. M., Modamio, J., Sauls, J. T., ... Fleming, R. M. T. (2019). Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nature Protocols*, 14(3), 639–702. <https://doi.org/10.1038/s41596-018-0098-2>

HmdbQuery. (s. f.). Bioconductor. Recuperado 16 de septiembre de 2020, de <http://bioconductor.org/packages/hmdbQuery/>

Huber, W., von Heydebreck, A., Sültmann, H., Poustka, A., & Vingron, M. (2002). Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics*, 18(suppl\_1), S96–S104. [https://doi.org/10.1093/bioinformatics/18.suppl\\_1.S96](https://doi.org/10.1093/bioinformatics/18.suppl_1.S96)

Iglesias, J., Morales, L., & Barreto, G. E. (2017). Metabolic and Inflammatory Adaptation of Reactive Astrocytes: Role of PPARs. *Molecular Neurobiology*, 54(4), 2518-2538. Scopus. <https://doi.org/10.1007/s12035-016-9833-2>

IMAT: an integrative metabolic analysis tool | Bioinformatics | Oxford Academic. (s. f.). Recuperado 20 de agosto de 2020, de <https://academic.oup.com/bioinformatics/article/26/24/3140/290045>

Jeffrey Kantor (2021). *Stoichiometry Tools* (<https://www.mathworks.com/matlabcentral/fileexchange/290045>) MATLAB Central File Exchange. Retrieved December 29, 2021.

---

Jensen, C. J., Massie, A., & De Keyser, J. (2013). Immune players in the CNS: The astrocyte. *Journal of Neuroimmune Pharmacology: The Official Journal of the Society on NeuroImmune Pharmacology*, 8(4), 824-839. <https://doi.org/10.1007/s11481-013-9480-6>

Jerby, L., Shlomi, T., & Ruppin, E. (2010). Computational reconstruction of tissue-specific metabolic models: Application to human liver metabolism. *Molecular Systems Biology*, 6(1), 401. <https://doi.org/10.1038/msb.2010.56>

Jha, M. K., Park, D. H., Kook, H., Lee, I.-K., Lee, W.-H., & Suk, K. (2016). Metabolic Control of Glia-Mediated Neuroinflammation. *Current Alzheimer Research*, 13(4), 387-402. <https://doi.org/10.2174/1567205013666151116124755>

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), 20150202. <https://doi.org/10.1098/rsta.2015.0202>

Jr, C. E. D. (2014). Optimal Algorithm for Metabolomics Classification and Feature Selection varies by Dataset. *International Journal of Biology*, 7(1), p100. <https://doi.org/10.5539/ijb.v7n1p100>

Karahalil, B. (2016, octubre 31). Overview of Systems Biology and Omics Technologies. *Current Medicinal Chemistry*. <https://www.eurekaselect.com/145809/article>

Karpievitch, Y. V., Dabney, A. R., & Smith, R. D. (2012). Normalization and missing value imputation for label-free LC-MS analysis. *BMC Bioinformatics*, 13(16), S5. <https://doi.org/10.1186/1471-2105-13-S16-S5>

Kim, M., Rai, N., Zorraquino, V., & Tagkopoulos, I. (2016). Multi-omics integration accurately predicts cellular state in unexplored conditions for *Escherichia coli*. *Nature Communications*, 7(1), 13090. <https://doi.org/10.1038/ncomms13090>

Kitchenham, B., Pearl Brereton, O., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2009). Systematic literature reviews in software engineering – A systematic literature review. *Information and Software Technology*, 51(1), 7-15. <https://doi.org/10.1016/j.infsof.2008.09.009>



Lange, S. C., Bak, L. K., Waagepetersen, H. S., Schousboe, A., & Norenberg, M. D. (2012). Primary cultures of astrocytes: Their value in understanding astrocytes in health and disease. *Neurochemical research*, 37(11), 2569-2588. <https://doi.org/10.1007/s11064-012-0868-0>

Layé, S. (2010). Polyunsaturated fatty acids, neuroinflammation and well being. *Prostaglandins Leukotrienes and Essential Fatty Acids*, 82(4-6), 295-303. Scopus. <https://doi.org/10.1016/j.plefa.2010.02.006>

Li, B., & Dewey, C. N. (2011). RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, 12(1), 323. <https://doi.org/10.1186/1471-2105-12-323>

Ludvigsen, M., & Honoré, B. (2018). Transcriptomics and Proteomics: Integration? In *ELS* (pp. 1–7). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9780470015902.a0006188.pub2>

Machado, D., & Herrgård, M. (2014). Systematic Evaluation of Methods for Integration of Transcriptomic Data into Constraint-Based Models of Metabolism. *PLOS Computational Biology*, 10(4), e1003580. <https://doi.org/10.1371/journal.pcbi.1003580>

Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt—PubMed. (s. f.). Recuperado 16 de septiembre de 2020, de <https://pubmed.ncbi.nlm.nih.gov/19617889/>

Mardinoglu, A., Agren, R., Kampf, C., Asplund, A., Nookaew, I., Jacobson, P., Walley, A. J., Froguel, P., Carlsson, L. M., Uhlen, M., & Nielsen, J. (2013). Integration of clinical data with a genome-scale metabolic model of the human adipocyte. *Molecular Systems Biology*, 9, 649. <https://doi.org/10.1038/msb.2013.5>

Mardinoglu, A., Agren, R., Kampf, C., Asplund, A., Uhlen, M., & Nielsen, J. (2014). Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease. *Nature Communications*, 5, 3083. <https://doi.org/10.1038/ncomms4083>

Martín-Jiménez, C. A., Salazar-Barreto, D., Barreto, G. E., & González, J. (2017). Genome-Scale Reconstruction of the Human Astrocyte Metabolic Network. *Frontiers in Aging Neuroscience*, 9. <https://doi.org/10.3389/fnagi.2017.00023>

---

Martin-Jiménez, C., González, J., Vesga, D., Aristizabal, A., & Barreto, G. E. (2020). Tibolone Ameliorates the Lipotoxic Effect of Palmitic Acid in Normal Human Astrocytes. *Neurotoxicity Research*, 38(3), 585–595. <https://doi.org/10.1007/s12640-020-00247-4>

Masid, M., Ataman, M., & Hatzimanikatis, V. (2020). Analysis of human metabolism by reducing the complexity of the genome-scale models using redHUMAN. *Nature Communications*, 11(1), 2821. <https://doi.org/10.1038/s41467-020-16549-2>

McBean, G. J. (2017). Cysteine, Glutathione, and Thiol Redox Balance in Astrocytes. *Antioxidants*, 6(3). <https://doi.org/10.3390/antiox6030062>

mixOmics: An R package for ‘omics feature selection and multiple data integration. (s. f.). Recuperado 16 de septiembre de 2020, de <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005752>

N, L., Ja, D., Ja, T., V, R., Ai, R., Bm, S., Jd, H., Mn, G., Rm, S., & C, K. (2013, abril 15). EBSeq: An empirical Bayes hierarchical model for inference in RNA-seq experiments. *Bioinformatics* (Oxford, England); Bioinformatics. <https://doi.org/10.1093/bioinformatics/btt087>

Navid, A., & Almaas, E. (2012). Genome-level transcription data of *Yersinia pestis* analyzed with a New metabolic constraint-based approach. *BMC Systems Biology*, 6(1), 150. <https://doi.org/10.1186/1752-0509-6-150>

Neurodegenerative diseases—Latest research and news | Nature. (s. f.). Recuperado 12 de septiembre de 2020, de <https://www.nature.com/subjects/neurodegenerative-diseases>

Oliveira, A. de A. B., Melo, N. de F. M., Vieira, É. dos S., Nogueira, P. A. S., Coope, A., Velloso, L. A., Dezone, R. S., Ueira-Vieira, C., Botelho, F. V., Gomes, J. de A. S., & Zanon, R. G. (2018). Palmitate treated-astrocyte conditioned medium contains increased glutathione and interferes in hypothalamic synaptic network in vitro. *Neurochemistry International*, 120, 140-148. <https://doi.org/10.1016/j.neuint.2018.08.010>

Osorio, D., Pinzón, A., Martín-Jiménez, C., Barreto, G. E., & González, J. (2020). Multiple Pathways Involved in Palmitic Acid-Induced Toxicity: A System Biology Approach. *Frontiers in Neuroscience*, 13. <https://doi.org/10.3389/fnins.2019.01410>

Osorio, D., Botero, K., Gonzalez, J., & Pinzon, A. (2016). “*exp2flux*” Convert Gene EXPression Data to FBA FLUXes. <https://doi.org/10.13140/RG.2.2.14401.56168>

Pandey, V., Hadadi, N., & Hatzimanikatis, V. (2019). Enhanced flux prediction by integrating relative expression and relative metabolite abundance into thermodynamically consistent metabolic models. *PLOS Computational Biology*, 15(5), e1007036. <https://doi.org/10.1371/journal.pcbi.1007036>

Pfefferkorn, C., Kallfass, C., Lienenklaus, S., Spanier, J., Kalinke, U., Rieder, M., Conzelmann, K.-K., Michiels, T., & Staeheli, P. (2016). Abortively Infected Astrocytes Appear To Represent the Main Source of Interferon Beta in the Virus-Infected Brain. *Journal of Virology*, 90(4), 2031-2038. <https://doi.org/10.1128/JVI.02979-15>

Phatnani, H., & Maniatis, T. (2015a). Astrocytes in Neurodegenerative Disease. *Cold Spring Harbor Perspectives in Biology*, 7(6). <https://doi.org/10.1101/cshperspect.a020628>

Phatnani, H., & Maniatis, T. (2015b). Astrocytes in Neurodegenerative Disease. *Cold Spring Harbor Perspectives in Biology*, 7(6). <https://doi.org/10.1101/cshperspect.a020628>

Prah, J., Winters, A., Chaudhari, K., Hersh, J., Liu, R., & Yang, S.-H. (2019). A Novel Serum Free Primary Astrocyte Culture Method that Mimic Quiescent Astrocyte Phenotype. *Journal of Neuroscience Methods*, 320, 50. <https://doi.org/10.1016/j.jneumeth.2019.03.013>

R, A., Sg, V.-B., & K, R. (2011, agosto 15). Metab: An R package for high-throughput analysis of metabolomics data generated by GC-MS. *Bioinformatics* (Oxford, England); *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/btr379>

---

R, W., G, W., & F, M. (2014, enero 9). metaMS: An open-source pipeline for GC-MS-based untargeted metabolomics. *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences; J Chromatogr B Analyt Technol Biomed Life Sci.* <https://doi.org/10.1016/j.jchromb.2014.02.051>

Ramus, C., Hovasse, A., Marcellin, M., Hesse, A.-M., Mouton-Barbosa, E., Bouys-sié, D., Vaca, S., Carapito, C., Chaoui, K., Bruley, C., Garin, J., Cianféroni, S., Ferro, M., Van Dorssaeler, A., Burlet-Schiltz, O., Schaeffer, C., Couté, Y., & Gonzalez de Peredo, A. (2016). Benchmarking quantitative label-free LC-MS data processing workflows using a complex spiked proteomic standard dataset. *Journal of Proteomics*, 132, 51-62. <https://doi.org/10.1016/j.jprot.2015.11.011>

Robinson, M. (s. f.). flagme: Analysis of Metabolomics GC/MS Data version 1.44.0 from Bioconductor. Recuperado 16 de septiembre de 2020, de <https://rdrr.io/bioc/flagme/>

Russell, W. M. S., & Burch, R. L. (1959). *The principles of humane experimental technique*. Methuen.

Sajitz-Hermstein, M., Töpfer, N., Kleessen, S., Fernie, A. R., & Nikoloski, Z. (2016). iReMet-flux: Constraint-based approach for integrating relative metabolite levels into a stoichiometric metabolic models. *Bioinformatics*, 32(17), i755-i762. <https://doi.org/10.1093/bioinformatics/btw465>

Schellenberger, J., Lewis, N. E., & Palsson, B. Ø. (2011). Elimination of Thermodynamically Infeasible Loops in Steady-State Metabolic Models. *Biophysical Journal*, 100(3), 544–553. <https://doi.org/10.1016/j.bpj.2010.12.3707>

Schmidt, B. J., Ebrahim, A., Metz, T. O., Adkins, J. N., Palsson, B. Ø., & Hyduke, D. R. (2013). GIM3E: Condition-specific models of cellular metabolism developed from metabolomics and expression data. *Bioinformatics*, 29(22), 2900-2908. <https://doi.org/10.1093/bioinformatics/btt493>

Segrè, D., Vitkup, D., & Church, G. M. (2002). Analysis of optimality in natural and perturbed metabolic networks. *Proceedings of the National Academy of Sciences*, 99(23), 15112-15117. <https://doi.org/10.1073/pnas.232349399>

Shinfuku, Y., Sorpitiporn, N., Sono, M., Furusawa, C., Hirasawa, T., & Shimizu, H. (2009). Development and experimental verification of a genome-scale metabolic model for *Corynebacterium glutamicum*. *Microbial Cell Factories*, 8(1), 43. <https://doi.org/10.1186/1475-2859-8-43>

Skogerson, K., Wohlgemuth, G., Barupal, D. K., & Fiehn, O. (2011). The volatile compound BinBase mass spectral database. *BMC Bioinformatics*, 12(1), 321. <https://doi.org/10.1186/1471-2105-12-321>

Stare, T., Stare, K., Weckwerth, W., Wienkoop, S., & Gruden, K. (2017). Comparison between Proteome and Transcriptome Response in Potato (*Solanum tuberosum* L.) Leaves Following Potato Virus Y (PVY) Infection. *Proteomes*, 5(3). <https://doi.org/10.3390/proteomes5030014>

Swainston, N., Smallbone, K., Mendes, P., Kell, D. B., & Paton, N. W. (2011). The SuBliMinaL Toolbox: Automating steps in the reconstruction of metabolic networks. *Journal of Integrative Bioinformatics*, 8(2), 187–203. <https://doi.org/10.1515/jib-2011-186>

Swainston, N., Smallbone, K., Hefzi, H., Dobson, P. D., Brewer, J., Hanscho, M., Zielinski, D. C., Ang, K. S., Gardiner, N. J., Gutierrez, J. M., Kyriakopoulos, S., Lakshmanan, M., Li, S., Liu, J. K., Martínez, V. S., Orellana, C. A., Quek, L.-E., Thomas, A., Zanghellini, J., ... Mendes, P. (2016). Recon 2.2: From reconstruction to model of human metabolism. *Metabolomics*, 12(7), 109. <https://doi.org/10.1007/s11306-016-1051-4>

Tejera, N., Crossman, L., Pearson, B., Stoakes, E., Nasher, F., Djeghout, B., Poolman, M., Wain, J., & Singh, D. (2020). Genome-Scale Metabolic Model Driven Design of a Defined Medium for *Campylobacter jejuni* M1cam. *Frontiers in Microbiology*, 11, 1072. <https://doi.org/10.3389/fmicb.2020.01072>

Thiele, I., Vlassis, N., & Fleming, R. M. T. (2014). fastGapFill: Efficient gap filling in metabolic networks. *Bioinformatics (Oxford, England)*, 30(17), 2529–2531. <https://doi.org/10.1093/bioinformatics/btu321>

Välikangas, T., Suomi, T., & Elo, L. L. (2018). A comprehensive evaluation of popular proteomics software workflows for label-free proteome quantification and imputation. *Briefings in Bioinformatics*, 19(6), 1344–1355. <https://doi.org/10.1093/bib/bbx054>

Väremo, L., Scheele, C., Broholm, C., Mardinoglu, A., Kampf, C., Asplund, A., Nookaew, I., Uhlén, M., Pedersen, B. K., & Nielsen, J. (2015). Proteome- and transcriptome-driven reconstruction of the human myocyte metabolic network and its use for identification of markers for diabetes. *Cell Reports*, *11*(6), 921–933. <https://doi.org/10.1016/j.celrep.2015.04.010>

Vieira, V., Ferreira, J., Rodrigues, R., Liu, F., & Rocha, M. (2019). A Model Integration Pipeline for the Improvement of Human Genome-Scale Metabolic Reconstructions. *Journal of Integrative Bioinformatics*, *16*(1). <https://doi.org/10.1515/jib-2018-0068>

Yang, Z., & Xiong, H.-R. (2012). Culture Conditions and Types of Growth Media for Mammalian Cells. In *Biomedical Tissue Culture*. IntechOpen. <https://doi.org/10.5772/52301>