



Reposicionamiento de fármacos en Enfermedades Mitocondriales

Pablo Serrano Lorenzo

Máster universitario en Bioinformática y bioestadística

TFM - Bioinformática y Bioestadística

Nombre de la consultora:

Marta Enciso Carrasco

Profesor responsable de la asignatura:

Javier Luis Cánovas Izquierdo

Fecha Entrega: 08-01-2020



Esta obra está sujeta a una licencia de Reconocimiento-NoComercial-SinObraDerivada [3.0 España de Creative Commons](https://creativecommons.org/licenses/by-nc-nd/3.0/es/)

Copyright © 2020 Pablo Serrano-Lorenzo

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts.

A copy of the license is included in the section entitled "GNU Free Documentation License".

FICHA DEL TRABAJO FINAL

Título del trabajo:	<i>Reposicionamiento de fármacos en Enfermedades Mitocondriales</i>
Nombre del autor:	<i>Pablo Serrano Lorenzo</i>
Nombre del consultor/a:	<i>Marta Enciso Carrasco</i>
Nombre del PRA:	<i>Javier Luis Cánovas Izquierdo</i>
Fecha de entrega (mm/aaaa):	01/2020
Titulación::	<i>Máster universitario en Bioinformática y bioestadística</i>
Área del Trabajo Final:	<i>TFM - Bioinformática y Bioestadística</i>
Idioma del trabajo:	<i>Español</i>
Palabras clave	<i>Drug repurposing, Mitochondrial diseases, Machine Learning</i>
<p>Resumen del Trabajo (máximo 250 palabras): <i>Con la finalidad, contexto de aplicación, metodología, resultados i conclusiones del trabajo.</i></p>	
<p>Las enfermedades mitocondriales (EM) son un conjunto de enfermedades de base genética caracterizadas por defectos en el sistema de fosforilación oxidativa (OXPHOS) u otros trastornos metabólicos en los que participan las mitocondrias. Actualmente no existe tratamiento para la inmensa mayoría de estas enfermedades, por lo que el reposicionamiento de fármacos es una estrategia interesante y prometedora para el descubrimiento de nuevas terapias. En el presente trabajo se ha desarrollado una metodología, mediante el uso de técnicas de <i>Machine Learning</i> (ML), que permite la priorización de fármacos con posibles efectos beneficiosos en el ámbito de las EM.</p> <p>Mediante el uso secuencial de dos modelos, uno para la búsqueda de fármacos con propiedades físico-químicas y estructurales similares a las de los fármacos que se usan actualmente en EM, y otro modelo para la predicción de nuevas interacciones fármaco-diana mitocondrial se han priorizado un total de 68 fármacos (no presentes en el set de fármacos con los que se entrenaron los modelos).</p> <p>Tras una revisión bibliográfica sobre los 68 fármacos priorizados se han encontrado trabajos sobre 7 fármacos del set que apoyan su posible efecto beneficioso en función mitocondrial (D-Cisteína, Ácido Málico, L-Glutamina, Ácido Fólico, Beta-Alanina, Glicina y Beta-Amino Isobutirato). Sólo se ha hallado una referencia sobre uno de los fármacos priorizados en el que se describe un posible efecto perjudicial sobre la actividad mitocondrial (D-Serina). Finalmente concluimos que el conjunto de 68 fármacos priorizados podría contener nuevos fármacos con un efecto potencialmente beneficioso en EM.</p>	

Abstract (in English, 250 words or less):

Mitochondrial diseases (MD) are a group of genetic disorders characterized by defects in the oxidative phosphorylation system (OXPHOS) or other metabolic disorders involving mitochondria. Currently there is no treatment for the majority of these diseases, so drug repositioning is an interesting and promising way to discover new therapies. In the present work a methodology has been developed, using *Machine Learning* (ML) techniques, which allows the prioritization of drugs with potential effects in MD.

Through a ML/ML tandem use of two models, one for searching drugs with chemical and structural properties similar to the drugs currently used in MD, and another model for the prediction of new mitochondrial drug-target interactions, a total of 68 drugs (not present in the original set of drugs for model training) were prioritized.

After a review of the 68 prioritized drugs, research studies have been found about seven drugs (D-Cysteine, Malic Acid, L-Glutamine, Folic acid, Beta-Alanine, Glycine and Beta-Amino Isobutyrate) that supports their potential beneficial effect on mitochondrial function and one work has been found about one prioritized drug (D-Serine) in which a possible detrimental effect on mitochondrial activity is described. Overall we concluded that the set of 68 prioritized drugs could contain new drugs with a potential beneficial effect for the treatment of MD.

Índice

Índice.....	iv
Lista de figuras.....	vii
Lista de tablas.....	7
1. Introducción.....	1
1.1 Contexto y justificación del Trabajo.....	1
1.2 Objetivos del Trabajo.....	2
1.3 Enfoque y método seguido.....	3
1.4 Planificación del Trabajo (Tareas).....	5
1.4 Planificación del Trabajo (Hitos).....	6
1.5 Breve resumen de productos obtenidos.....	6
1.6 Breve descripción de los otros capítulos de la memoria.....	7
2. Selección de los fármacos y proteínas que se incluirán en el trabajo.....	8
3. Creación de los predictores moleculares de los fármacos.....	9
4. Creación del dataset de interacciones.....	11
5. Creación de los predictores de interacción.....	13
6. Creación y evaluación de modelos basados en fármacos.....	15
7. Creación y evaluación de los modelos de interacción fármaco-diana.....	18
8. Predicción de fármacos con potencial efecto positivo mitocondrial.....	21
9. Análisis de los resultados obtenidos.....	23
10. Conclusiones.....	34
11. Glosario.....	36
12. Bibliografía.....	37
13. Anexos.....	41

Lista de figuras

Imagen 1 : Cronograma Planificación de tareas	6
Imagen 2 : Gráfico de correlación predictores moleculares	10
Imagen 3 : Distribución predictor ALogP	11
Imagen 4 : Distribución predictor Zagreb	11
Imagen 5 : Rendimiento de los modelos en función del número de variables (RFE)	14
Imagen 6 : Distribución predictor XLogP, seleccionado por el algoritmo RFE	15
Imagen 7 : Distribución predictor nRotB, desechado por el algoritmo RFE	15
Imagen 8 : Matriz de confusión del modelo final KNN	16
Imagen 9 : Matriz de confusión del modelo final ANN	17
Imagen 10 : Matriz de confusión del modelo final SVM	17
Imagen 11 : Matriz de confusión del modelo final RF	18
Imagen 12 : Matriz de confusión del modelo final KNN Interacciones	19
Imagen 13 : Matriz de confusión del modelo final ANN Interacciones	20
Imagen 14 : Matriz de confusión del modelo final SVM Interacciones	20
Imagen 15 : Matriz de confusión del modelo final RF Interacciones	21
Imagen 16 : Cluster fármacos mitoplus (rojo) vs selección fármacos DrugBank (azul)	24
Imagen 17 : Cluster fármacos Inter (rojo) vs selección fármacos DrugBank (azul)	25
Imagen 18 : Cluster fármacos Mito+ (rojo) vs MitoTrain (azul)	26
Imagen 19 : Cluster fármacos mitoplus (rojo) vs MitoTrain (azul)	27
Imagen 20 : Imágenes estructuras fármacos Cluster 140, 87	28
Imagen 21 : Imágenes estructuras fármacos Cluster 136, 8	28
Imagen 22 : Imágenes estructuras fármacos Cluster 33, 34, 60	29
Imagen 23 : Imágenes estructuras fármacos Cluster 16, 44, 130	29
Imagen 24 : Imágenes estructuras fármacos Cluster 2, 82, 133	29
Imagen 25 : Imágenes estructuras fármacos Cluster 138, 85, 103	29
Imagen 26 : Imágenes estructuras fármacos Cluster 57, 72, 16	30
Imagen 27 : Imágenes estructuras fármacos Cluster 74, 5	30
Imagen 28 : Imágenes estructuras fármacos Cluster 66, 43	30
Imagen 29 : Imágenes estructuras fármacos Cluster 50, 18, 10	31
Imagen 30 : Imágenes estructuras fármacos Cluster 9, 58	31

Lista de tablas

Tabla 1 : Ejemplos de los predictores moleculares utilizados	9
Tabla 2 : Resumen del dataset de interacciones	12
Tabla 3 : Predictores proteicos utilizados	13
Tabla 4 : Resumen del rendimiento de los diferentes modelos	18
Tabla 5 : Resumen del rendimiento de los diferentes modelos de interacciones	21
Tabla 6 : Interacciones de las enzimas <i>mt-aaRS2</i> , según el modelo de interacciones	32

1. Introducción

1.1 Contexto y justificación del Trabajo

Las mitocondrias son orgánulos celulares que desempeñan diferentes funciones relacionadas con el metabolismo y la homeostasis celular, siendo la producción de energía celular mediante la fosforilación oxidativa (OXPHOS) su función mejor conocida. Las enfermedades mitocondriales (EM) son un conjunto heterogéneo de enfermedades de base genética caracterizadas por defectos en el sistema OXPHOS u otros trastornos metabólicos en los que participa la mitocondria (ciclo de Krebs o metabolismo de algunos cofactores y vitaminas, por ejemplo). Existen múltiples EM, siendo la presentación clínica y la edad de aparición de los síntomas muy variados, pudiendo iniciarse en la fase neonatal, infancia o la edad adulta. Igualmente los síntomas pueden afectar a un solo órgano o bien ser multisistémicas, siendo generalmente los órganos más afectados aquellos que requieren una mayor demanda energética, como el cerebro o el músculo esquelético o cardíaco.¹

Una particularidad de las mitocondrias es que cuentan con su propio genoma, que codifica 37 genes, mientras que el resto de las aproximadamente 1500 proteínas que forman el proteoma mitocondrial son codificadas en el genoma nuclear. Se han descrito mutaciones en más de 350 genes, tanto nucleares como mitocondriales, asociadas a patología mitocondrial, no existiendo, en general, una clara relación fenotipo-genotipo. Por lo tanto es común que un mismo síndrome sea causado por mutaciones en diferentes genes (por ejemplo el síndrome de Leigh se ha asociado a mutaciones en aproximadamente 80 genes distintos) o bien que mutaciones en el mismo gen den lugar a fenotipos diferentes. Todo esto, unido a un conocimiento incompleto de los mecanismos involucrados en las diferentes patologías, hace que el diagnóstico de estas enfermedades sea un proceso complejo.^{1,2}

Los tratamientos existentes en la actualidad para las EM se basan principalmente en agentes capaces de mejorar la transferencia electrónica de la cadena respiratoria mitocondrial (CRM) como la coenzima Q10 o la riboflavina, entre otros. Otras estrategias terapéuticas son el suministro de antioxidantes o sustancias que aumentan la biosíntesis mitocondrial (ácido lipoico, vitamina C, bezafibrato).³ Sin embargo la existencia de terapias más específicas es prácticamente nula, existiendo en la actualidad solo un tratamiento orientado al defecto genético subyacente (en el déficit de TK2).⁴

Entre las dificultades que presenta el desarrollo de nuevas terapias está la heterogeneidad clínica y genética mencionada anteriormente así como el difícil acceso a la doble membrana mitocondrial donde se ubican los complejos OXPHOS, la escasez de modelos animales con EM o un número escaso de pacientes genéticamente caracterizados para poder llevar a cabo ensayos clínicos significativos.² La obtención de tratamientos específicos para estas enfermedades sigue siendo, por lo tanto, objeto de investigación. En este

sentido el reposicionamiento de fármacos es una estrategia interesante y que de hecho ya se está usando en el ámbito de las EM.⁵

El reposicionamiento de fármacos consiste en la búsqueda de nuevas indicaciones y dianas terapéuticas para fármacos ya aprobados y existentes en el mercado. Dado lo costoso en términos económicos y de tiempo que es el desarrollo de un nuevo fármaco, esta estrategia está siendo cada vez más utilizada en la industria farmacéutica por las ventajas que ofrece: la fase de ensayo preclínico y el desarrollo de los perfiles tóxicos, farmacocinéticos y farmacodinámicos de los fármacos ya han sido realizados por lo que el coste y el tiempo de todo el proceso se ve muy reducido.⁶ Por otra parte el reposicionamiento de fármacos resulta especialmente interesante en ciertas áreas, como es el de las enfermedades raras, dentro de las cuales se engloban las diferentes EM, ya que el bajo número de pacientes afectados por estas enfermedades hace que exista una menor inversión por parte de las compañías farmacéuticas en el desarrollo de nuevos tratamientos para ellas.⁷

No se ha publicado hasta la fecha ningún estudio de reposicionamiento de fármacos en EM desde una aproximación computacional o *in silico*. Los estudios de reposicionamiento *in silico* tienen, respecto a los estudios de reposicionamiento experimentales (como el citado anteriormente)⁵ la ventaja de que son menos costosos y más rápidos, permitiendo, por ejemplo, el cribado de un número mayor de fármacos.⁶ Sin embargo, tanto en la aproximación *in silico* como en la experimental, siempre va a ser necesaria una validación clínica de los resultados obtenidos.

1.2 Objetivos del Trabajo

El objetivo del TFM es la creación de modelos de predicción computacional, basados en técnicas de *Machine Learning* (ML) convencional (K-Nearest Neighbors, Neural Networks, Support Vector Machines y Random Forest), para la priorización de fármacos con un posible efecto beneficioso en EM. Para ello se crearán modelos que usarán predictores moleculares basados en las propiedades físico-químicas y estructurales de los compuestos que se usan en la actualidad en EM o que han demostrado un efecto positivo en función mitocondrial para la búsqueda de fármacos con propiedades similares (modelo basado en fármacos).

Por otra parte se crearán modelos mediante el uso de predictores moleculares de los fármacos y predictores proteicos de sus dianas mitocondriales para la predicción de nuevas interacciones fármaco-diana mitocondrial aprovechando el conocimiento que existe en la actualidad sobre las redes de interacciones fármaco-proteína y del proteoma mitocondrial.

Como el trabajo está enfocado en EM se elegirá como espacio biológico de estudio las aproximadamente 1150 proteínas que componen el proteoma mitocondrial.⁸

Para la obtención de la información necesaria para la creación de los modelos se usaran bases de datos públicas de propiedades, estructuras e interacciones fármaco-proteína (DrugBank, PubChem) y bases de datos de proteínas enriquecidas con información funcional (UniProt).

Aunque no entra dentro de los objetivos del TFM, cabe señalar que si al finalizar el TFM se obtuviese un resultado positivo existiría la posibilidad de validar experimentalmente los fármacos candidatos en diferentes modelos celulares de EM.

1.3 Enfoque y método seguido

El reposicionamiento de fármacos in silico puede abordarse usando diferentes estrategias. Tradicionalmente se han usado técnicas centradas en dianas terapéuticas individuales, para las cuales se buscaban moléculas que tuvieran el mayor grado de afinidad posible en función de las estructuras y propiedades físico-químicas de la diana y los fármacos. Este enfoque se podría resumir mediante la teoría “un fármaco - una diana terapéutica - una enfermedad”.⁹

Sin embargo la evidencia ha demostrado que no siempre una alta afinidad entre un fármaco y su diana se relacionan con su eficacia terapéutica.¹⁰ Al contrario, muchos fármacos han mostrado tener efectos sobre diferentes dianas, lo que probablemente resulte más eficiente que actuar solo en una, ya que el metabolismo celular suele ser muy robusto a la alteración de sus componentes individuales y tiende a compensarlo mediante diferentes mecanismos celulares.⁹ Estas evidencias han hecho que recientemente la búsqueda de nuevos fármacos tienda a hacerse mediante aproximaciones más holísticas, basadas en biología de sistemas, en lugar de centrarse únicamente en la afinidad fármaco-diana.¹¹

Desde el punto de vista de las EM, y aunque la mayoría de éstas son enfermedades monogénicas, es decir, causadas por un defecto concreto en una determinada proteína, se ha decidido seguir una aproximación general o sistémica en lugar de una centrada en dianas únicas por diferentes motivos:

- a) El ensamblaje de los diferentes complejos OXPHOS es un proceso extremadamente intrincado y regulado¹², y por lo tanto es muy improbable que pueda ser corregido mediante un fármaco basándose solamente en una afinidad elevada entre el fármaco y la proteína defectuosa.
- b) Mutaciones en ciertos genes nucleares afectan rutas bioquímicas complejas (por ejemplo genes de transcripción y traducción mitocondrial, genes de metabolismo Fe/S, genes de mantenimiento del mtDNA, genes involucrados en dinámica mitocondrial, etc).
- c) La accesibilidad a la doble membrana mitocondrial donde se encuentran los complejos OXPHOS es muy reducida.²

Por lo tanto se considera que podrían tener un mayor efecto beneficioso aquellas moléculas con potencial para modular simultáneamente diferentes

dianas involucradas en rutas metabólicas y mecanismos de señalización celular mitocondriales que puedan estar afectados debido a la enfermedad, en lugar de aquellas que simplemente muestren una alta afinidad por la proteína afectada.

Con este planteamiento en mente se pretende seleccionar un espacio terapéutico lo más amplio posible, eligiendo, como mínimo, todas las proteínas que forman parte de los diferentes complejos OXPHOS así como sus diferentes factores de ensamblaje (en total 150 proteínas aproximadamente). Idealmente, y si es posible, se elegirá como espacio terapéutico las proteínas incluidas en el repositorio MitoCarta 2.0 (un listado de 1158 proteínas de localización mitocondrial).⁸

Como espacio farmacológico se elegirá el mayor número de compuestos posible, teniendo en cuenta que probablemente haya que descartar un número elevado debido a la falta de información necesaria sobre ellos para construir los modelos. Se intentará en la medida de lo posible incluir todos aquellos fármacos que hayan demostrado tener un efecto positivo en enfermedades mitocondriales^{3,13} y en caso de no estar disponible la información necesaria se planteará la posibilidad de curar manualmente la entrada de dichos fármacos para hacer posible su inclusión en los modelos. Dicha inclusión es importante por varios factores:

- a) para el entrenamiento de los modelos basados en fármacos.
- b) es necesaria una red de interacciones fármaco-proteína ya establecida que servirá como respuesta para el entrenamiento del modelo de nuevas interacciones fármaco-proteína.
- c) para la evaluación de los modelos.

Respecto a las herramientas bioinformáticas que se usarán para desarrollar los modelos se plantea la posibilidad de usar como herramienta principal el paquete de R/Bioconductor Rcpí.¹⁴ Mediante esta herramienta se pueden obtener múltiples parámetros relacionados con la composición, estructura y propiedades físico-químicas de fármacos y proteínas, así como el cálculo de varios descriptores de interacciones proteína-proteína y fármaco-proteína.

A partir de los diferentes predictores obtenidos con el paquete Rcpí se plantea la generación de dos tipos de modelos diferentes:

1. Creación de un modelo para la priorización de fármacos con propiedades físico-químicas y estructurales semejantes a los usados en la actualidad en EM o que hayan demostrado tener un efecto beneficioso en modelos celulares humanos derivados de pacientes de EM.
2. Creación de un modelo para la predicción de nuevas interacciones fármaco-diana mitocondrial.

La estrategia a seguir será la de priorizar solo los fármacos clasificados como de interacción positiva con diana mitocondrial que además tuviesen una estructura similar a los ya utilizados en EM (provenientes del modelo basado en

fármacos). Por último se analizarán los resultados obtenidos y se evaluará el rendimiento de los diferentes modelos.

1.4 Planificación del Trabajo (Tareas)

- a) Selección de los fármacos que se incluirán en el trabajo y las bases de datos de las que se obtendrá la información necesaria. También se determinará que clase de predictores asociados al espacio farmacológico se utilizarán para desarrollar los modelos y así poder evaluar si las bases de datos que se consultarán cuentan con la información necesaria sobre ellos. Como se comentó en el punto “Enfoque y método seguido” se pondrá especial interés en incluir todos aquellos fármacos que hayan demostrado tener efecto en EM pues su presencia es indispensable para poder desarrollar los modelos. Periodo estimado para la realización de la tarea: 7 días (15-21 Octubre de 2019).
- b) Selección de las proteínas que se incluirán en el trabajo. También se determinará que clase de predictores asociados al espacio terapéutico se utilizarán para crear los modelos y así poder evaluar si las bases de datos cuentan con la información necesaria para su desarrollo. El objetivo es poder incluir el mayor número posible de proteínas del repositorio MitoCarta 2.0⁸, especialmente aquellas que forman parte o intervienen en el ensamblaje de los diferentes complejos OXPHOS. Periodo estimado para la realización de la tarea: 7 días (22-29 Octubre de 2019).
- c) Obtención de los predictores asociados a los espacios farmacológico y terapéuticos usando la librería de R/Bioconductor RcpI. Periodo estimado para la realización de la tarea: 12 días (30 Octubre de 2019 - 10 Noviembre 2019).
- d) Creación de los modelos basados en fármacos y de los modelos de interacciones fármaco-diana. Se intentará crear el mayor número de modelos posible en el tiempo estimado para esta tarea, así como una puesta a punto de los mismos probando diferentes parámetros de ajuste. Se usará como herramienta principal para ello el paquete de R *caret*.¹⁵ Periodo estimado para la realización de la tarea: 15 días (11-25 Noviembre 2019).
- e) Evaluación de los modelos obtenidos en la tarea anterior. Para ello se crearán un conjunto de train y test set y se calcularán parámetros para medir el rendimiento de los modelos sobre el test set tales como la precisión o sensibilidad. Periodo estimado para la realización de la tarea: 5 días (26-30 Noviembre 2019).
- f) Análisis de resultados: Se analizarán los resultados obtenidos por los distintos modelos: se comprobará si el modelo basado en fármacos es capaz de seleccionar fármacos de estructura similar a los usados en EM (train set) mediante gráficas de agrupamiento jerárquico, se

analizarán los tipos de fármacos (según estructura) que arrojen tanto el modelo basado en fármacos como el de interacciones, se realizará una búsqueda bibliográfica de los fármacos priorizados por los modelos para comprobar si ya se ha descrito algún tipo de efecto sobre la actividad mitocondrial de dichos fármacos. Periodo estimado para la realización de la tarea: 15 días (1-16 Diciembre 2019).

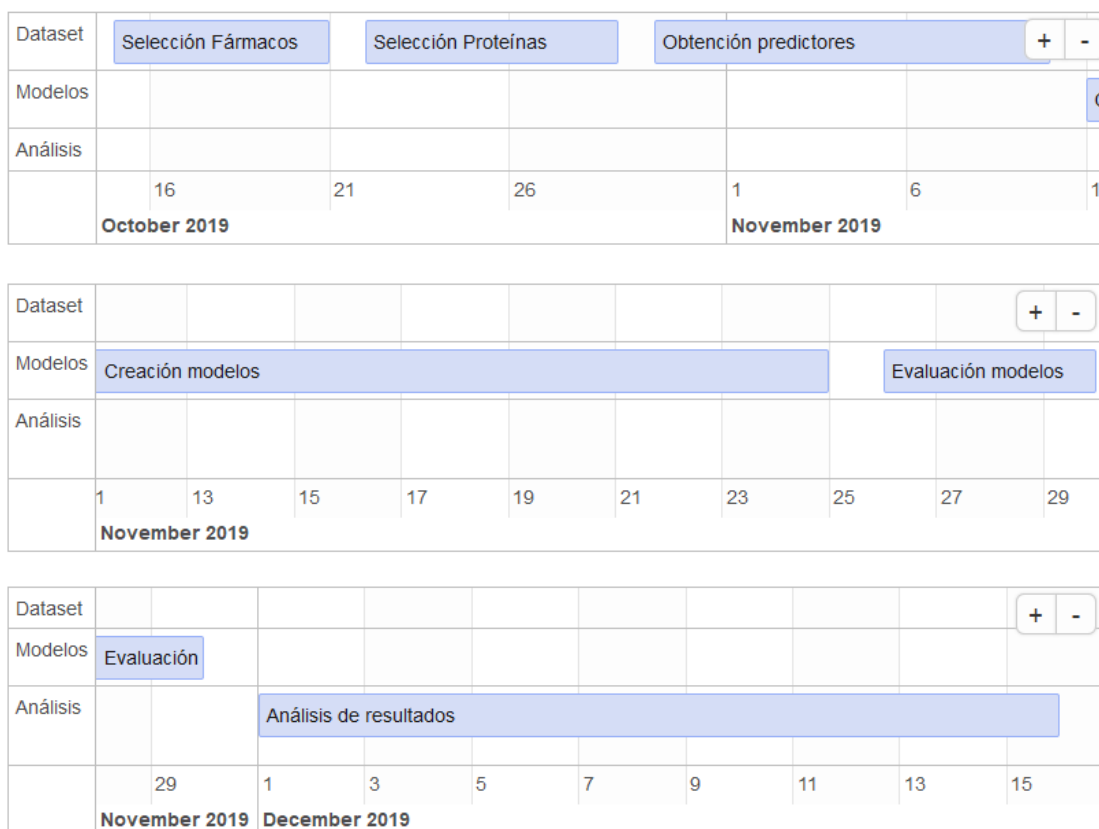


Imagen 1: Cronograma Planificación de tareas

1.4 Planificación del Trabajo (Hitos)

- a) Obtención de los espacios farmacológico y terapéutico. Verificación de que en las bases de datos públicas se pueden obtener los datos necesarios para la creación de los modelos.
- b) Desarrollo de los modelos ML (fármacos e interacciones) usando los datos y predictores obtenidos.
- c) Análisis y conclusiones de los resultados obtenidos.

1.5 Breve sumario de productos obtenidos

En la primera parte del trabajo se ha obtenido una lista de fármacos usados en la actualidad en EM o que han demostrado tener un efecto positivo sobre la función mitocondrial (ver bibliografía) para el entrenamiento de los modelos,

una lista de proteínas mitocondriales (dianas) para el modelo de interacciones, así como los diferentes parámetros y descriptores asociados generados mediante la librería de R RcpI.

En la segunda parte del trabajo se han creado los diferentes modelos y se ha evaluado su rendimiento. Utilizando los modelos que mejor rendimiento obtuvieron se ha priorizado una lista de 68 fármacos con propiedades físico-químicas y estructurales semejantes a los usados en EM (obtenidos con el modelo basado en fármacos) y que además interactúan con dianas mitocondriales según las predicciones del modelo de interacciones. Se ha analizado la lista de fármacos obtenida (tipos estructurales, consulta bibliográfica de cada uno de ellos, tipos de interacciones más frecuentes).

1.6 Breve descripción de los otros capítulos de la memoria

2. Selección de los fármacos y proteínas que se incluirán en el trabajo.
3. Creación de los predictores moleculares de los fármacos.
4. Creación del dataset de interacciones.
5. Creación de los predictores de interacción.
6. Creación y evaluación de modelos basados en fármacos.
7. Creación y evaluación de modelos de interacciones fármaco-diana.
8. Predicción de fármacos con potencial efecto positivo mitocondrial.
9. Análisis de los resultados obtenidos.
10. Conclusiones.
11. Glosario.
12. Bibliografía.
13. Anexos.

2. Selección de los fármacos y proteínas que se incluirán en el trabajo.

Los fármacos seleccionados se dividirán entre aquellos con un efecto positivo en enfermedades mitocondriales (*mitop*)^{2-5,16-23} vs fármacos con un efecto tóxico mitocondrial (*miton*).²³⁻²⁶ Para la selección de fármacos *mitop* se ha hecho una búsqueda bibliográfica relacionada con:

- a) Compuestos usados actualmente para el tratamiento de enfermedades mitocondriales.
- b) Compuestos que están en proceso de ensayo clínico para el tratamiento de enfermedades mitocondriales.
- c) Compuestos que han demostrado tener un efecto funcional positivo en modelos celulares humanos derivados de pacientes de enfermedad mitocondrial.
- d) Compuestos de uso compasivo en los que se ha publicado un efecto positivo en los pacientes tratados.

La selección de fármacos *mitop* y *miton* (Total = 82 fármacos), con la que se entrenaron los modelos basados en fármacos se puede consultar en el Anexo: PEC – 2: selecciones *drug_mitop* y *drug_miton*.

Para la obtención de la información de los cinco fármacos *mitop* que no estaban en la base de datos Drugbank (KH-176, KL1333, Molibresib, JP4-039 y Mitochonic acid 5) se ha creado manualmente un archivo adicional con el Nombre, SMILES (Simplified Molecular-Input Line-Entry System, necesario para crear los predictores moleculares) y PubChem ID de cada compuesto (información obtenida de la base de datos PubChem).²⁷ La información del resto de fármacos *mitop* y *miton* se ha obtenido de la base de datos DrugBank.²⁸

El set de fármacos que se intentará reposicionar y la información necesaria para crear los predictores moleculares (SMILES) se ha obtenido de la base de datos DrugBank.²⁸ Se usarán todos los fármacos incluidos en las secciones: Approved, Experimental, Nutraceutical, Investigational (aproximadamente 11.000 fármacos).

Como espacio terapéutico se han seleccionado todas las proteínas de localización mitocondrial recopiladas en el repositorio MitoCarta 2.0 con entrada curada en UniProt (Total = 456 proteínas).²⁹ El proceso para la obtención de proteínas con entrada curada en UniProt se puede consultar en el Anexo: PEC - 2: selección *UP_curated*. Se seleccionan solo proteínas curadas debido a que cuentan con una secuencia de aminoácidos canónica, mientras que proteínas no curadas cuentan con diferentes entradas y diferentes secuencias de aminoácidos (obtenidas supuestamente a través de splicings

alternativos del gen que las codifica). Como la obtención de predictores proteicos se basa en la secuencia de aminoácidos de la proteína se ha preferido usar para los modelos solo aquellas que cuenten con una secuencia única y curada manualmente como la secuencia canónica de la proteína.

3. Creación de los predictores moleculares de los fármacos.

Para la creación de los predictores moleculares de los fármacos se usó la librería de R/Bioconductor Rcpí.¹⁴ Los descriptores se crean a partir de la información de los compuestos en formato MOL. Previamente se usa la función *readMolFromSmi()* de la librería Rcpí para transformar la información original en formato SMILES a formato MOL. El proceso detallado de como se crean los predictores se puede consultar en el Anexo: PEC – 2: Creación de los predictores moleculares de los fármacos.

A modo de resumen: se crearon un total de 67 predictores moleculares obtenidos a partir de 28 funciones distintas proporcionadas por el paquete Rcpí. En la siguiente tabla se muestran algunos ejemplos de las funciones utilizadas y el tipo de descriptor que originan (extraídas directamente de la documentación del paquete Rcpí):

Función	Tipo descriptor
<code>extractDrugALOGP()</code>	Atom additive logP and molar refractivity values descriptor
<code>extractDrugApol()</code>	Sum of the atomic polarizabilities
<code>extractDrugAutocorrelationMass()</code>	Moreau-Broto autocorrelation descriptors using atomic weight
<code>extractDrugBPol()</code>	Sum of the absolute value of the difference between atomic polarizabilities of all bonded atoms in the molecule
<code>extractDrugCarbonTypes()</code>	Topological descriptor characterizing the carbon connectivity in terms of hybridization
<code>extractDrugHBondAcceptorCount()</code>	Number of hydrogen bond acceptors
<code>extractDrugKappaShapeIndices()</code>	Kier & Hall Kappa molecular shape indices
<code>extractDrugLargestChain()</code>	Number of atoms in the largest chain
<code>extractDrugMannholdLogP()</code>	LogP based on the number of carbons and hetero atoms
<code>extractDrugTPSA()</code>	Topological Polar Surface Area (TPSA)

Tabla 1: Ejemplos de los predictores moleculares utilizados

Se eliminaron tres de los 67 predictores en los que se observó que apenas existía variación entre los valores que tomaban (varianza ≈ 0). Para ello se utilizó la función *nearZeroVar()* del paquete *caret* que avisa sobre este tipo de variables con varianza cero o *cuasicero*.

Ante la sospecha de que muchos de los predictores podrían guardar relación entre sí, se realizó un estudio de correlación entre todas las variables (código para obtener el gráfico y la lista de los predictores con correlación > 0.9 en el

Anexo: Correlación variables Dataset Fármacos.R). A continuación se muestra el gráfico de correlación obtenido al comparar los 64 predictores utilizados (una vez eliminados los 3 predictores de baja variabilidad) para la creación de los modelos basados en Fármacos:

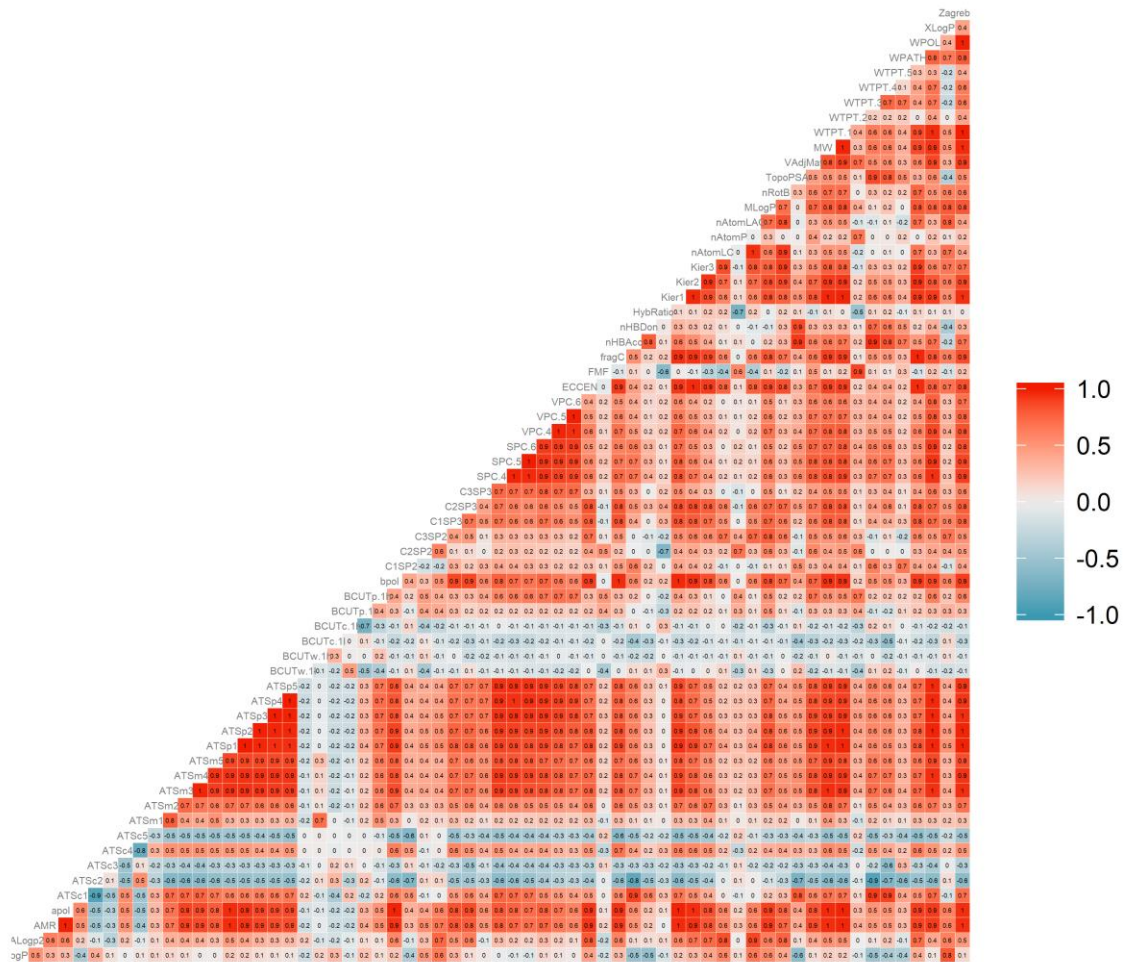


Imagen 2: Gráfico de correlación predictores moleculares

A continuación se muestra la representación gráfica de dos de los predictores mediante un gráfico de densidad para ver si se aprecia alguna diferencia en su distribución respecto a la variable a predecir “Mito” (0 = *miton* y 1 = *mitop*). El código para obtener las imágenes se puede consultar en el Anexo: PEC – 3: Modelos basados en Fármacos. Lectura, exploración de los datos y obtención de los muestras de train y test:

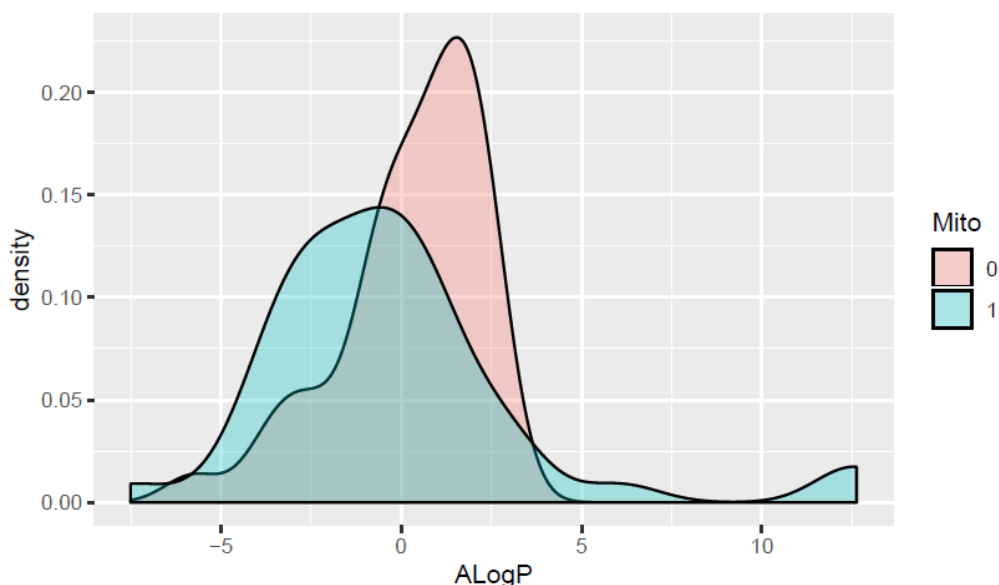


Imagen 3: Distribución predictor ALogP

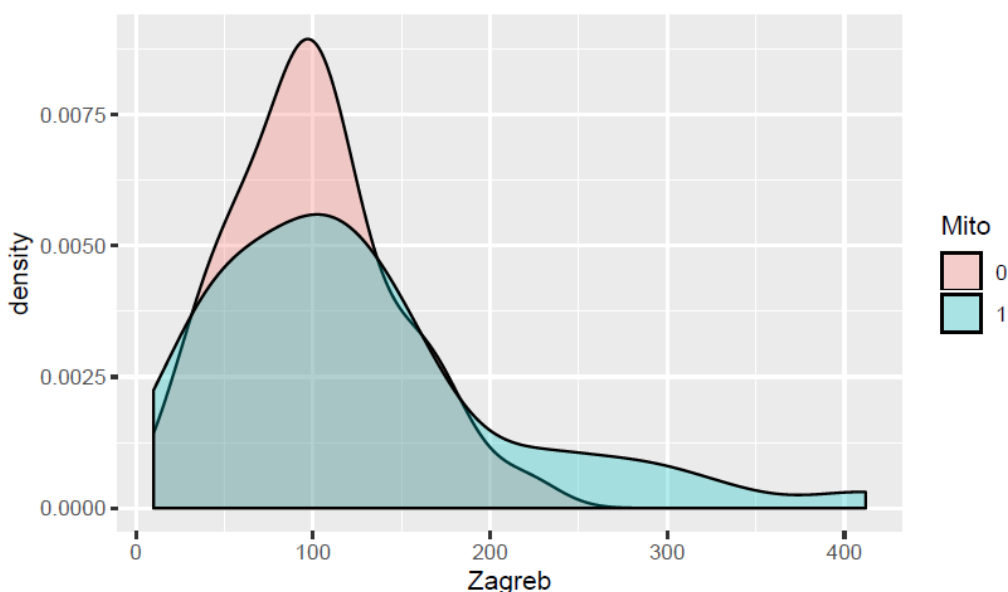


Imagen 4: Distribución predictor Zagreb

4. Creación del dataset de interacciones.

Como dataset de interacciones fármaco-diana conocidas se ha seleccionado la presente en la base de datos DrugBank.²⁸ Se usarán todos los fármacos de la base de datos (dataset *all*, que incluye los fármacos de las siguientes categorías: Approved, Small Molecule, Biotech, Experimental, Nutraceutical, Illicit, Withdrawn, Investigational).

Como tipo de respuesta se ha optado por interacción positiva/interacción negativa (o interacción SI/NO). No se va a tener en cuenta los tipos de interacciones posibles (por ejemplo de activación o inhibición) para evitar

construir el modelo con grupos de escasa n (pues el número de interacciones fármaco-diana mitocondrial conocidas no es muy elevado). Se intentará compensar esta limitación priorizando solo los fármacos clasificados como de interacción positiva con diana mitocondrial que además tuviesen una estructura similar a los ya utilizados en enfermedades mitocondriales (provenientes del modelo basado en fármacos).

El proceso detallado de como se construyó el dataset de interacciones positivas se encuentra en el Anexo: PEC – 2: Creación del dataset de interacciones positivas.

La tabla final de interacciones obtenida tiene el siguiente aspecto:

Symbol	Drug	Sequence
TK2	DB02452	MLLWPLRGWAARALRCFGPGSRGS...
TK2	DB02594	MLLWPLRGWAARALRCFGPGSRGS...
TK2	DB04485	MLLWPLRGWAARALRCFGPGSRGS...
NDUFS8	DB00157	MRCLTTPMLLRALAQAARAGPPGGR...
NDUFA4	DB00157	MLRQIIGQAKKHPSLIPLFVFIGTGAT...

Tabla 2: Resumen del dataset de interacciones

Donde *Symbol* es el nombre del gen, según la nomenclatura del HUGO Gene Nomenclature Committee (HGNC).³⁰ *Drug* es el código DrugBank de cada fármaco y *Sequence* es la secuencia de aminoácidos canónica que codifica cada gen (obtenida de las entradas curadas de cada gen en la base de datos UniProt). Dicha tabla se completará añadiendo los predictores moleculares de los fármacos a partir de su código SMILES (de manera análoga a como se realizó para el dataset de los modelos basados en fármacos, ver capítulo 3 de la memoria) y los predictores proteicos a partir de la secuencia de aminoácidos (descrito en detalle en el capítulo 5 de la memoria: Creación de los predictores de interacción).

Para la creación del dataset de interacciones negativas se seleccionaron todos los fármacos presentes en la base de datos DrugBank excepto aquellos con interacciones conocidas con proteínas MitoCarta, según el dataset de interacciones de DrugBank. Posteriormente se combinó este set de fármacos sin interacción con proteínas mitocondriales de forma aleatoria con las proteínas MitoCarta con entrada curada en la base de datos UniProt. El proceso detallado de como se construyó el dataset de interacciones negativas se encuentra en el Anexo: PEC – 2: Creación del dataset de interacciones negativas.

Las dimensiones finales del dataset de interacciones, una vez eliminadas las entradas defectuosas (como por ejemplo fármacos con el campo SMILES vacío), es de 1.177 entradas: 594 interacciones positivas y 583 interacciones negativas.

5. Creación de los predictores de interacción.

Los predictores de interacción son, por una parte, los mismos predictores moleculares asociados a los fármacos que se usaron para los modelos basados en fármacos (ver capítulo 3 de la memoria) y, por otra parte, los predictores proteicos basados en la secuencia de aminoácidos de cada proteína. Para la elección de los predictores proteicos que se usaron en los modelos se ha seguido una aproximación similar al del artículo de Ong *et al.*³¹ Se utilizó un set de predictores de cada categoría principal (aquella que según el trabajo de Ong tuvo un mejor rendimiento dentro de su categoría), siendo lo ideal la utilización de todos los predictores en su conjunto. En la siguiente tabla se muestran todas las funciones utilizadas y el tipo de descriptor proteico que originan (extraídas directamente de la documentación del paquete Rcp):

Función	Tipo descriptor
extractProtAAC()	Amino acid composition
extractProtDC()	Dipeptide composition
extractProtPAAC()	Pseudo-amino acid composition
extractProtCTDC()	Combination of amino acid and Dipeptide composition
extractProtQS0()	Quasi-sequence-order descriptors
extractProtMoreauBroto()	Correlation of physicochemical properties (Hydrophobicity scale, average flexibility index, polarizability parameter, etc)

Tabla 3: Predictores proteicos utilizados

El número total de predictores en el set de interacciones (moleculares + proteicos) es de 898. Debido a su elevado número, y tras haberse comprobado de manera análoga a como se hizo con el set de predictores moleculares de los modelos basados en fármacos que existe un alto índice de correlación entre las variables (ver capítulo 3 de la memoria, el gráfico equivalente no se muestra por ser el número de variables muy elevado) se decide reducir el número de predictores mediante técnicas de *Feature Selection*.

Como primera aproximación, igual que como se realizó en el dataset de fármacos, se eliminan aquellos predictores con varianza nula o cuasi-nula, debido a su bajo valor discriminante. En segundo lugar se utilizará el método de *Recursive Feature Elimination* (RFE) usando para ello las herramientas proporcionadas por el paquete *caret*. Mediante este método se ajustan varios modelos tomando un número diferente de predictores cada vez y comparando el rendimiento de cada modelo se puede establecer la importancia relativa de cada variable, permitiendo descartar aquellas que apenas tengan peso en los diferentes modelos. Para poder seleccionar el número óptimo de variables a utilizar es necesario realizar una *cross-validation* para evaluar diferentes grupos de variables y así poder seleccionar el conjunto que mejor rendimiento obtenga. En el trabajo se usó el algoritmo Random Forest para construir los modelos con una *cross-validation* de 10 iteraciones. El proceso detallado de como se realizó todo el proceso de *Feature Selection* (incluida la técnica de RFE) se encuentra en el Anexo: PEC – 3: Modelos basados en interacciones fármaco-diana - Lectura, exploración de los datos y obtención de los muestras de train y test).

Se seleccionó como mejor modelo el que usó 200 variables, por obtener la precisión más alta de entre todos los modelos probados (Accuracy = 0.862). Las 5 variables que más peso tuvieron en los modelos fueron: XLogP, MLogP, C2SP2, WTPT.4 y FMF. En la selección de las 200 variables de mayor peso hay 37 predictores moleculares de fármacos y 163 predictores proteicos.

A continuación se muestra una gráfica que representa el rendimiento de cada modelo en función de las variables con las que se construyó:

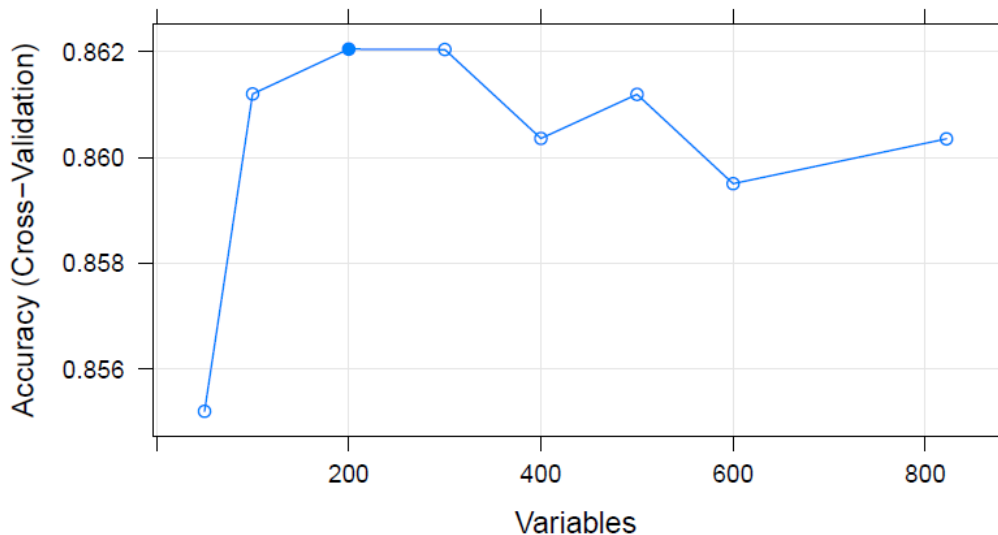


Imagen 5: Rendimiento de los modelos en función del número de variables (RFE)

Igual a como se hizo en el caso de los predictores basados en fármacos se hará una representación mediante gráficos de densidad de uno de los predictores mejor valorados por el algoritmo RFE vs un predictor calificado como de bajo valor (deshechado por el modelo RFE) para ver si se aprecian diferencias gráficas en su distribución respecto a la variable a predecir “outcome” (0 = “Interacción Negativa” y 1 = “Interacción Positiva”):

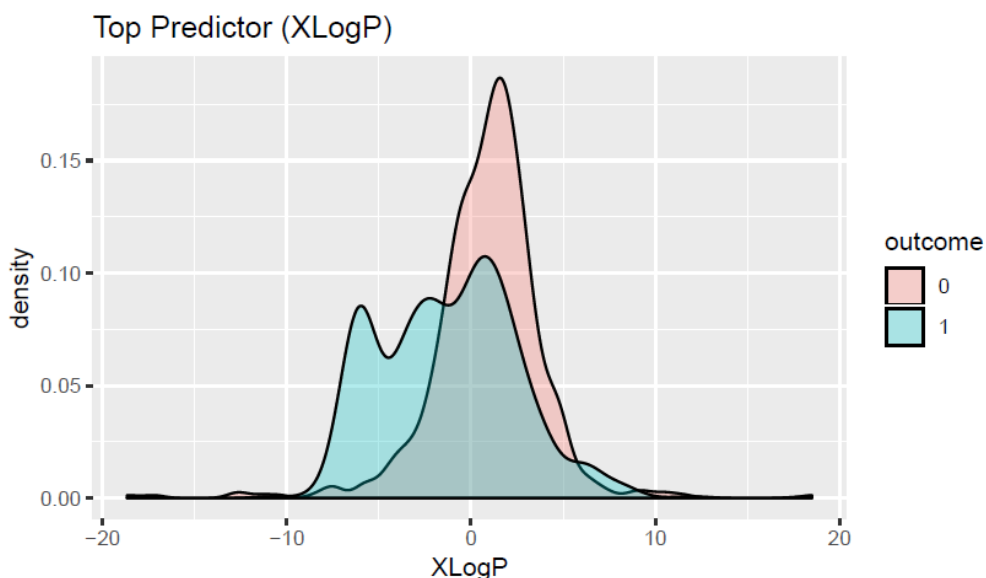


Imagen 6: Distribución predictor XLogP, seleccionado por el algoritmo RFE
Predictor desechado (nRotB)

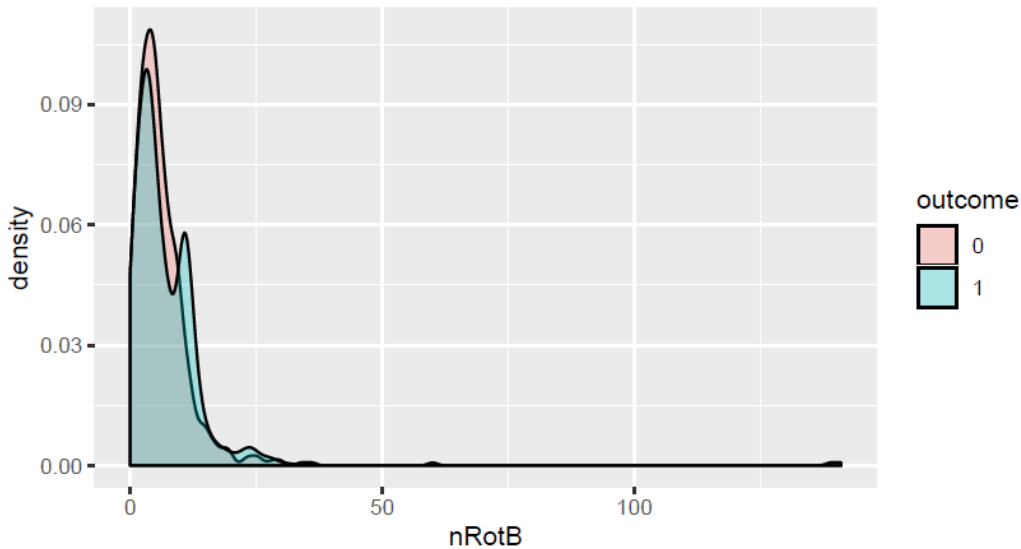


Imagen 7: Distribución predictor nRotB, desechado por el algoritmo RFE

6. Creación y evaluación de modelos basados en fármacos.

Las dimensiones del dataset para la creación de los modelos basados en fármacos es de 82 entradas (fármacos) por 64 variables o predictores (más la variable resultado o *outcome*). Para el entrenamiento y evaluación del rendimiento de los modelos se creó un training set con el 70% de los datos y el test set con el 30% restante de los datos.

Se usaron diferentes algoritmos: k-Nearest Neighbour, Artificial Neural Network (ANN), Support Vector Machines (SVM), Random Forest y se hizo una puesta a punto de cada modelo probando diferentes parámetros de ajuste y seleccionando como modelo final para comparar con el resto de algoritmos aquellos cuyos parámetros obtuvieron un rendimiento mayor. Los modelos se pusieron a punto usando la función *trainControl* del paquete *caret* mediante un muestreo del tipo *repeated cross-validation* (una *10 fold-cross-validation* repetida 5 veces). El proceso detallado de creación y evaluación de los modelos se encuentra en el Anexo: PEC – 3: Modelos basados en Fármacos).

- Resumen de los parámetros de ajuste y rendimientos obtenidos por los modelos:

K-Nearest Neighbour: Se probaron modelos con dos formas diferentes de transformar las variables (variables normalizadas y variables escaladas), usando el algoritmo KNN con el que cuenta el paquete *caret* (dependiente de la librería *e1071*). Los modelos se ajustaron con diferentes valores de K (número de vecinos) y se eligieron aquellos que mejor rendimiento tuvieron en función de la precisión o *accuracy*.

Tanto con variables normalizadas como con variables escaladas la $K = 4$ es la que ofrece una mejor precisión. Como posible inconveniente de usar una K relativamente tan baja está un posible “sobreajuste” a los datos del training set. Para elegir el mejor modelo (variables normalizadas vs escaladas) se probaron en el test set usando las K que ofrecieron un mejor resultado ($K = 4$), siendo en ambos casos la precisión obtenida idéntica (Precisión = 0.731). Finalmente se usará para comparar con el resto de algoritmos el modelo con variables escaladas por haber obtenido un comportamiento ligeramente superior en las predicciones sobre el training set.

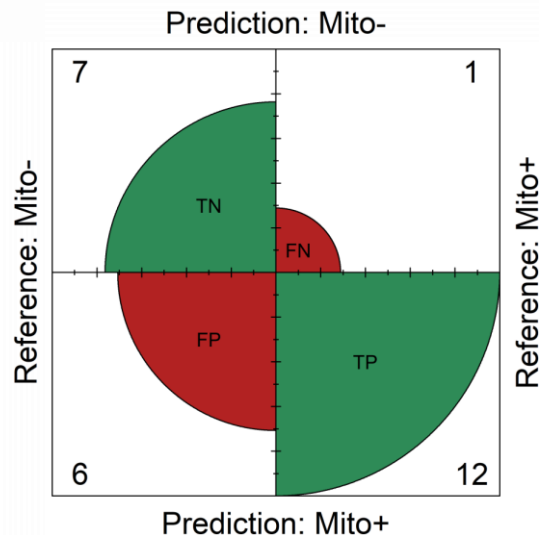


Imagen 8: Matriz de confusión del modelo final KNN

Artificial Neural Network (ANN): Para construir los modelos basados en redes neuronales se usó el método “*nnet*” del paquete *caret*. Para poner a punto los modelos, además de seleccionar el número de nodos en la capa oculta (*size*), hay que seleccionar un valor para el parámetro *decay*. Este parámetro es una penalización usada a la hora de calcular los *weights* o pesos con el objetivo de reducir el sobreajuste a los datos del train set. Respecto al tratamiento de las variables, al igual que en el caso anterior, se probaron modelos con las variables normalizadas y modelos con las variables escaladas.

Al igual que en el caso del algoritmo KNN las predicciones en el train set usando variables normalizadas o escaladas fue muy similar en los algoritmos ANN. Para elegir el mejor modelo ANN se prueban en el test set usando los parámetros de *size* y *decay* que mejor resultados obtuvieron en la cross-validation del train set, siendo en este caso el modelo con mejor rendimiento el de variables normalizadas (*size* = 10 y *decay* = 0.01) que obtuvo una Precisión = 0.808.

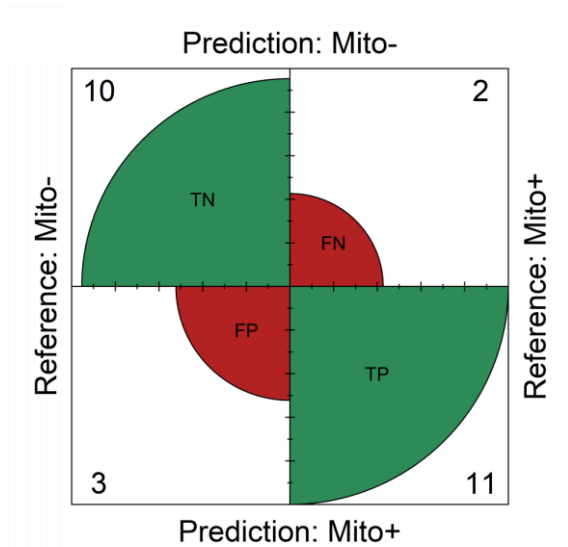


Imagen 9: Matriz de confusión del modelo final ANN

Support Vector Machines (SVM): Para construir los modelos SVM se usó el método “svm” del paquete *caret* (dependiente de la librería *kernelab*). En este caso se probaron dos tipos de modelos: uno usando una función de kernel lineal (“svmLinear”) y otro usando un kernel radial (“svmRadial”), usando solo variables normalizadas. Para encontrar los mejores parámetros de C y sigma se utilizó un rango alrededor de los valores usados habitualmente para estos parámetros.³²

Las predicciones en el training set del modelo con kernel radial fueron superiores a las del modelo de kernel lineal. Al probar los dos tipos de modelos en el test set también obtuvo un mejor rendimiento el modelo SVM con kernel radial (sigma = 0.08 y C = 3). La precisión obtenida fue 0.769.

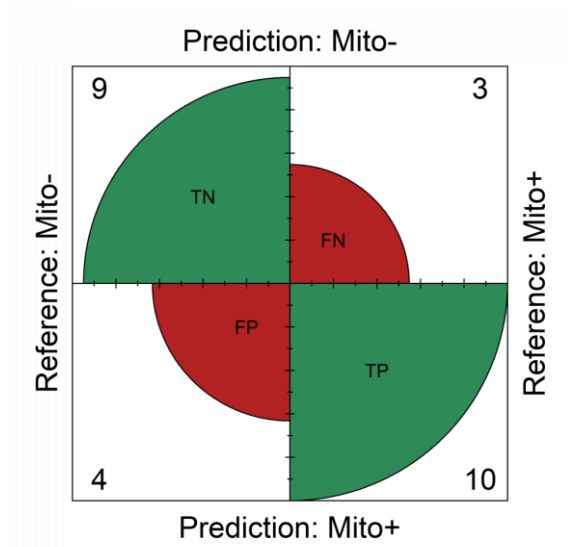


Imagen 10: Matriz de confusión del modelo final SVM

Random Forest (RF): Para construir los modelos Random Forest se usó el método “*rf*” del paquete *caret* (dependiente de la librería *randomForest*). En este caso se probó un único tipo de modelo con las variables sin transformar. Se utilizó la opción *tuneLength* del paquete *caret* para que el algoritmo pruebe diferentes valores por defecto del parámetro principal *mtry* (número de predictores, escogidos aleatoriamente, que se usan en cada nodo del árbol de decisión).

El parámetro de *mtry* que mejor resultado obtuvo en la *cross-validation* del train set fue *mtry* = 51 (Precisión = 0.686). La precisión del modelo eligiendo este parámetro de *mtry* en el test set fue 0.731.

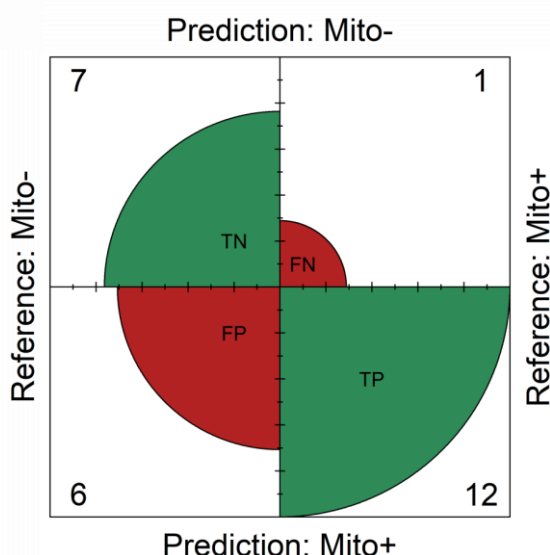


Imagen 11: Matriz de confusión del modelo final RF

En la siguiente tabla se resumen el rendimiento de los diferentes modelos en el training set y test set:

	Precisión (test set)	Sensibilidad (test set)	Precisión (training set)
KNN	0.731	0.923	0.681
ANN	0.808	0.846	0.698
SVM	0.731	0.769	0.725
RF	0.731	0.923	0.686

Tabla 4: Resumen del rendimiento de los diferentes modelos

7. Creación y evaluación de los modelos de interacción fármaco-diana.

Las dimensiones del dataset de interacciones es de 1.177 entradas (594 interacciones positivas y 583 interacciones negativas) y 899 variables: 898 predictores y la variable “inter” que codifica el resultado (interacción positiva =

1, interacción negativa = 0). Los 898 predictores originales se redujeron a 200 eliminando aquellos con varianza nula o muy baja y mediante el uso de una técnica de RFE (en capítulo 5: Creación de los predictores de interacción).

Para la predicción de nuevas interacciones fármaco-diana se probaron los mismos algoritmos que para los modelos basados en fármacos (Capítulo 6 de la memoria): k-Nearest Neighbour, Artificial Neural Network (ANN), Support Vector Machines (SVM) y Random Forest (RF), siguiéndose una metodología similar para la puesta a punto de los modelos, aunque en esta ocasión se usó el 75% de los datos para el training set y el 25% restante de los datos para la evaluación de los modelos. También se cambió el método de muestreo, para hacerlo menos exigente computacionalmente, a una *cross-validation* simple de 10 iteraciones. El proceso detallado de creación y evaluación de los modelos se encuentra en el Anexo: PEC – 3: Modelos basados en interacciones fármaco-diana).

- Resumen de los parámetros de ajuste y rendimientos obtenidos por los modelos:

K-Nearest Neighbour: En este caso solo se probaron modelos con variables normalizadas, ya que no se apreciaron diferencias significativas en los modelos previos entre usar variables normalizadas o escaladas. De todos los valores de K probados el modelo con mayor rendimiento fue aquel con una K = 10. No se probaron modelos con K más baja ya que sería probable que se obtuviese un modelo demasiado “sobreajustado” a los datos del training set. La precisión del modelo con K = 10 en el test set fue de 0.671.

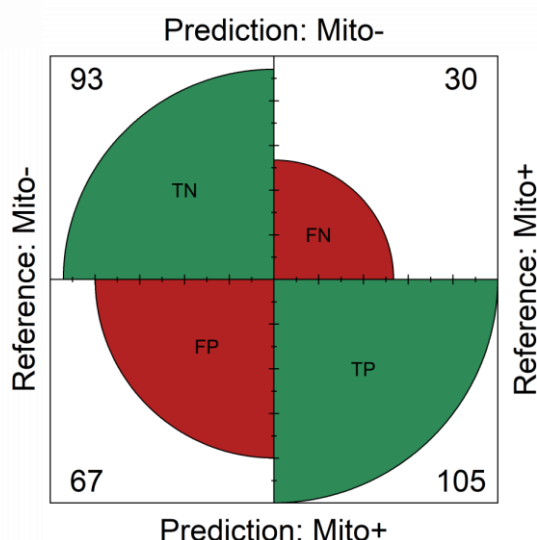


Imagen 12: Matriz de confusión del modelo final KNN Interacciones

Artificial Neural Network (ANN): Para construir el modelo se usarán solo variables normalizadas. Se ajusta el parámetro *MaxNWts* para incrementarlo a 3637 ya que es el número de pesos (weights) necesario para crear la red debido a su topología y a las dimensiones del dataset: (número máximo de

nodos en la capa oculta \times [número de columnas + 1] + (número máximo de nodos en la capa oculta + 1) = 3637 (número de columnas o variables = 200; número máximo de nodos en la capa oculta o size usados para poner a punto el modelo = 18). Los parámetros de size = 14 y decay = 0.18 son los que obtuvieron un mejor rendimiento en el train set. La precisión en el test set del modelo ANN usando estos parámetros fue de 0.793.

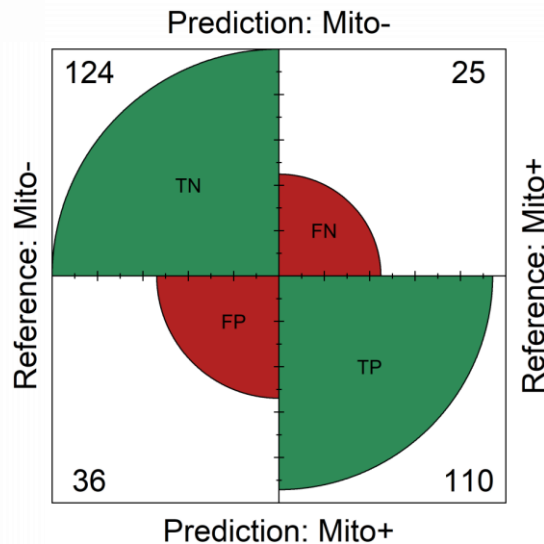


Imagen 13: Matriz de confusión del modelo final ANN Interacciones

Support Vector Machines (SVM): Para construir los modelos también se usaron solo variables normalizadas. Se probaron dos tipos de modelos, “svmLinear” y “svmRadial”. El modelo con kernel radial tuvo un rendimiento superior, siendo los parámetros de sigma = 0.02 y C = 6 los que obtuvieron un mejor rendimiento en el train set. La precisión en el test set del modelo SVM con kernel radial usando estos parámetros fue de 0.783.

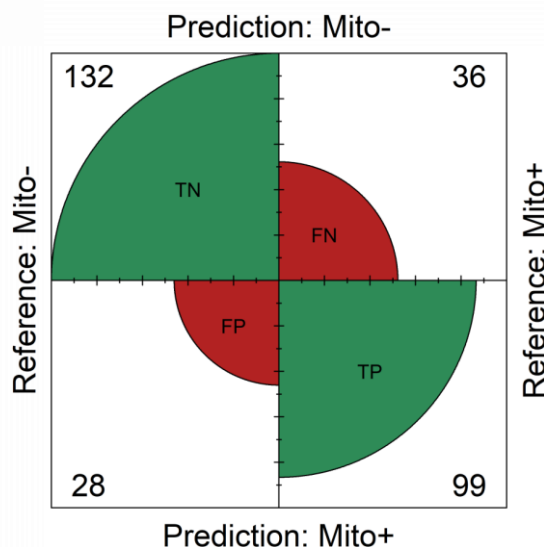


Imagen 14: Matriz de confusión del modelo final SVM Interacciones

Random Forest (RF): Se probó un único tipo de modelo con las variables sin transformar. Se utilizó la opción *tuneLength* del paquete *caret* para que el algoritmo pruebe diferentes valores por defecto del parámetro *mtry*. El valor de *mtry* que mayor rendimiento tuvo en el train set fue *mtry* = 115. La precisión del modelo RF en el test set usando este valor de *mtry* fue de 0.874.

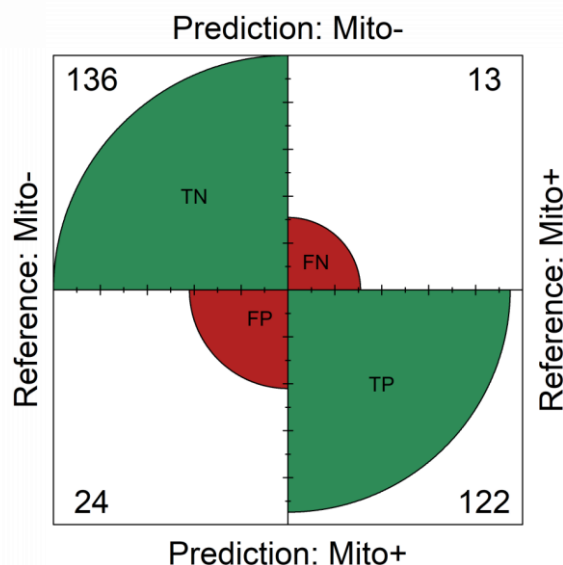


Imagen 15: Matriz de confusión del modelo final RF Interacciones

En la siguiente tabla se resumen el rendimiento de los diferentes modelos de interacciones en el training set y test set:

	Precisión (test set)	Sensibilidad (test set)	Precisión (training set)
KNN	0.671	0.778	0.709
ANN	0.793	0.815	0.817
SVM	0.783	0.733	0.809
RF	0.874	0.904	0.860

Tabla 5: Resumen del rendimiento de los diferentes modelos de interacciones

8. Predicción de fármacos con potencial efecto positivo mitocondrial.

Predicción de fármacos con posible efecto mitocondrial positivo (Mito+):

Como dataset para realizar las predicciones de fármacos con posible efecto mitocondrial positivo (Mito+) se usaron todos los fármacos de la base de datos DrugBank. Los predictores se crearon en scripts aparte debido al alto volumen de datos y tiempo de ejecución necesario para crearlo. (Anexos: Dataset completo farmacos.R, Dataset completo farmacos_resto de farmacos.R y Union de los subsets farmacos.R). También fue necesario procesar el dataset completo de fármacos DrugBank para dejarlo con el mismo número de

predictores que el dataset con el que se crearon los modelos basados en fármacos así como eliminar las entradas que contengan entradas vacías (NAs) en alguna de sus variables (código con el que se procesó el dataset completo de fármacos DrugBank en Anexo PEC-3 - 2).

Para la predicción de nuevos fármacos Mito+ se utilizaron los 2 modelos que mejor rendimiento obtuvieron: ANN con variables normalizadas y Random Forest (la selección de modelos se priorizó usando el siguiente criterio: Precisión en test set > Sensibilidad > Precisión en training set). Se seleccionaron como fármacos candidatos aquellos clasificados con mayor probabilidad de pertenecer a la clase Mito+ por cada modelo (ANN y RF). Código para obtener la lista de fármacos que han sido clasificados con alta probabilidad de tener un efecto Mito+ por los modelos ANN y RF en Anexo PEC 3 - 2.

Predicción de nuevas interacciones fármaco-diana mitocondrial:

Para realizar las predicciones de nuevas interacciones fármaco-diana mitocondrial se hicieron 5 nuevos datasets con combinaciones aleatorias de fármacos extraídos de la base de datos DrugBank con proteínas mitocondriales extraídas de MitoCarta y curadas por la base de datos UniProt.

Cada uno de los 5 datasets de interacciones aleatorias contenía 4000 interacciones, de las cuales para el 50% (2000 interacciones) se usaron fármacos de la base de datos completa DrugBank y para el 50% restante se usaron fármacos de la lista proporcionada por los modelos basados en fármacos con potencial efecto Mito+. También fue necesario crear y seleccionar los predictores que se usaron para el entrenamiento de los modelos de interacciones y descartar el resto. El código con el que se obtuvieron los 5 datasets de interacciones para hacer las predicciones de nuevas interacciones fármaco-diana mitocondrial se puede consultar en el Anexo PEC 3 - 2.Rmd (chunks *preparacion_interacciones* y *datasets_interacciones_aleatorias*).

Se consideraron como predicciones de mayor calidad (fármacos *mitoplus*) aquellas que cumplieron los siguientes criterios:

- i) Haber sido seleccionados como Mito+ por los modelos ANN y RF basados en fármacos.
- ii) Ser clasificados como interacción positiva fármaco-diana mitocondrial por el modelo basado en interacciones (RF, pues fue el que mejor rendimiento obtuvo).

9. Análisis de los resultados obtenidos.

1. Análisis de los grupos estructurales obtenidos por los modelos:

En esta sección se agrupará y evaluarán los fármacos priorizados por los modelos según su tipo estructural. Inicialmente se intentará visualizar, mediante gráficos que representen y agrupen jerárquicamente las estructuras moleculares de los fármacos, los siguientes grupos de fármacos:

i) los fármacos que hayan sido seleccionados tanto por el modelo basado en fármacos como por el modelo basado en interacciones con los criterios que se han descrito (*mitoplus*) vs una selección aleatoria de fármacos de la base de datos DrugBank completa.

ii) los fármacos seleccionados por el modelo basado en interacciones pero que NO coincidan con fármacos *mitoplus* (fármacos *Inter*) vs una selección aleatoria de fármacos de la base de datos DrugBank completa.

iii) los fármacos seleccionados por los modelos basados en fármacos (Mito+) vs los fármacos con efecto mitocondrial positivo demostrado que se seleccionaron para construir el train set del modelo basado en fármacos (*mitop* o MitoTrain).

iv) los fármacos *mitoplus* vs los fármacos con efecto mitocondrial positivo demostrado que se seleccionaron para construir el train set del modelo basado en fármacos (MitoTrain).

El objetivo de estas comparaciones es ver si existen sesgos o preferencias respecto al tipo de estructura elegida en los fármacos que han seleccionado los modelos. El código con el que se realizaron los gráficos de agrupación jerárquica de las estructuras de los fármacos se puede consultar en el Anexo PEC 3 - 2.Rmd: Análisis de los grupos estructurales obtenidos por los modelos).

i) Gráfico fármacos mitoplus vs DrugBank

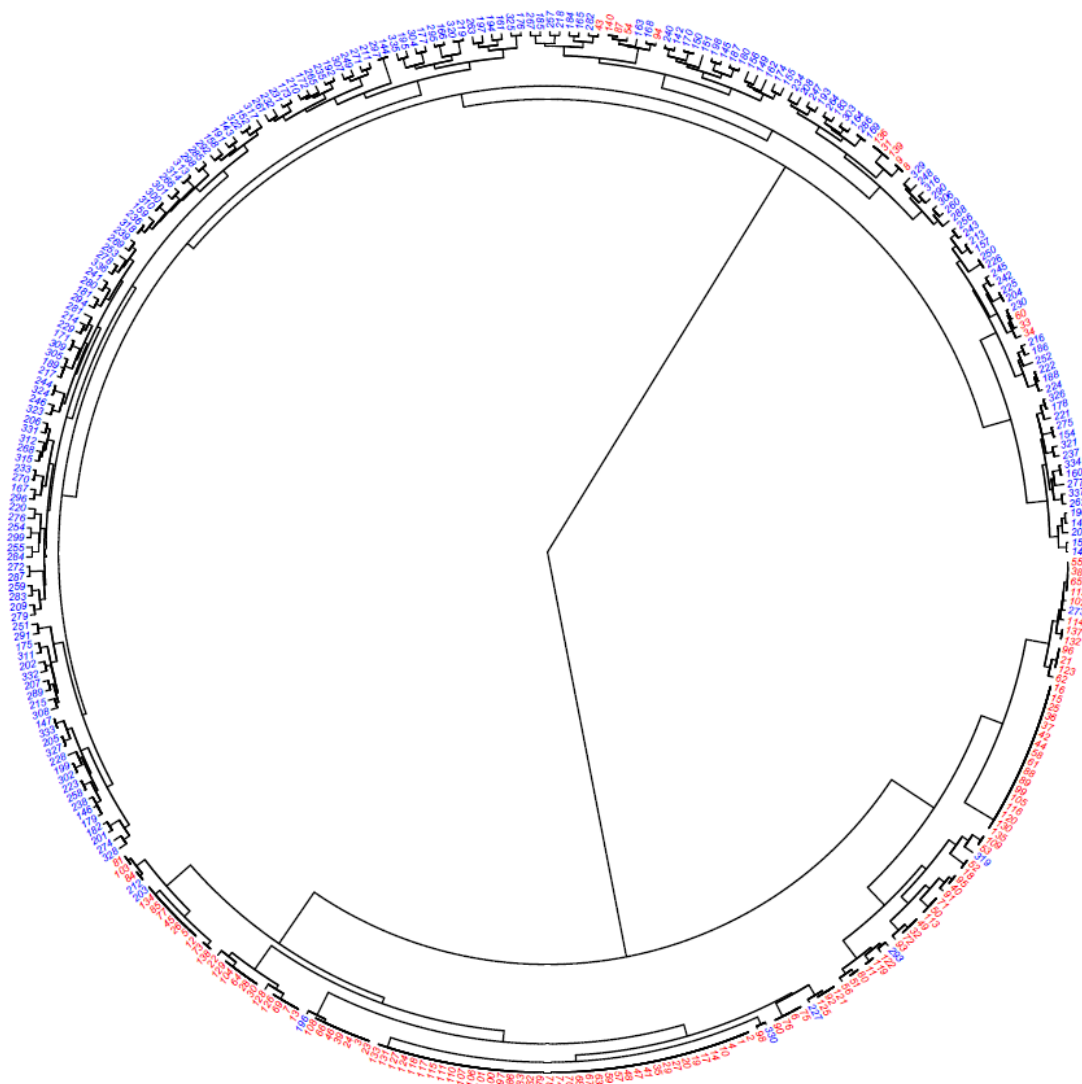


Imagen 16: Cluster fármacos mitoplus (rojo) vs selección fármacos DrugBank (azul)

Al ver el gráfico se observa que los fármacos en rojo (*mitoplus*) se encuentran claramente separados de los azules (selección DrugBank), lo que indica que los modelos han seleccionado fármacos que no representan toda la variedad de estructuras existente, si no que pertenecen a un pequeño subconjunto de estructuras de entre todas las opciones posibles.

ii) Gráfico fármacos Inter vs DrugBank

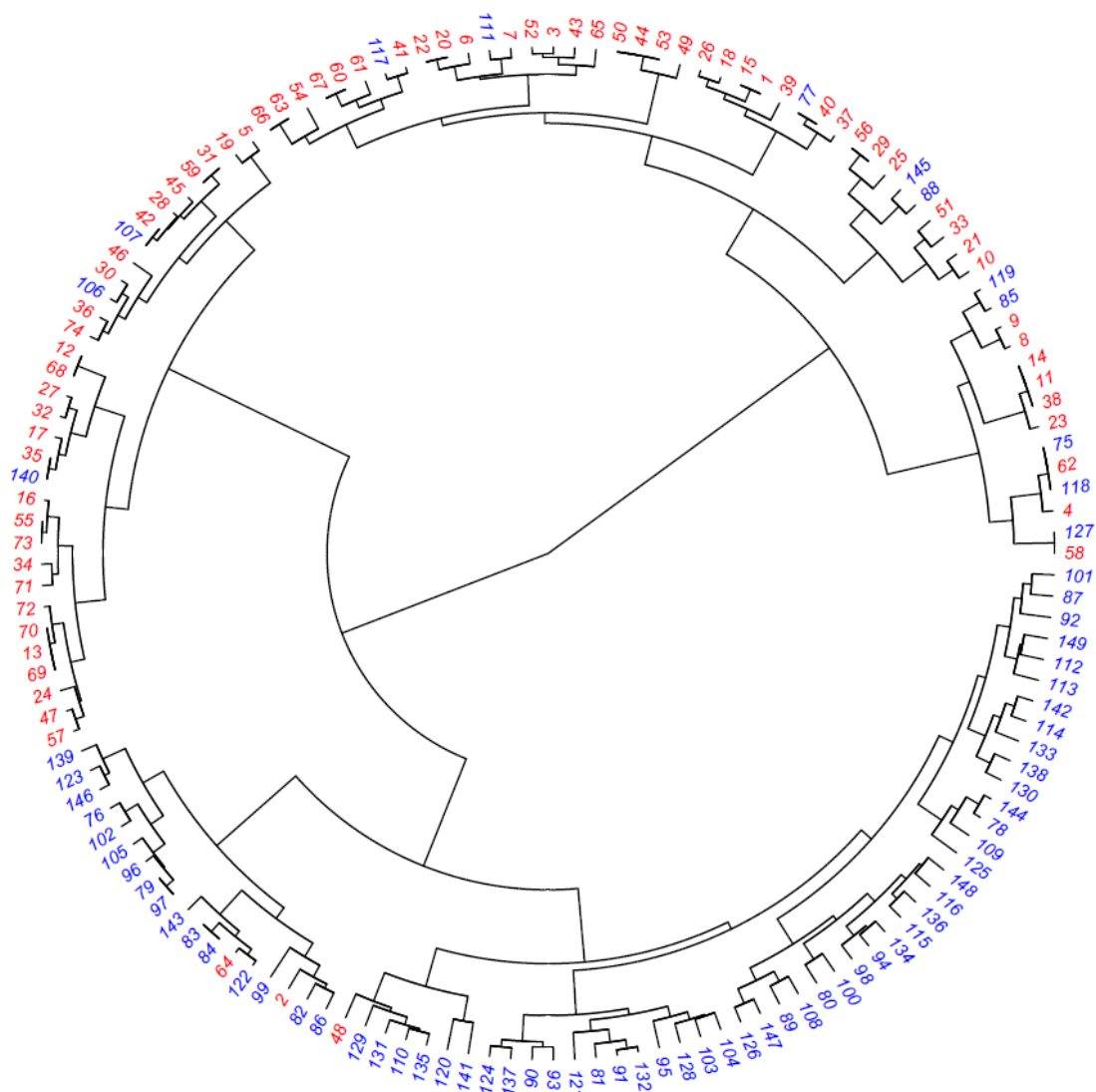


Imagen 17: Cluster fármacos Inter (rojo) vs selección fármacos DrugBank (azul)

Al igual que en el caso anterior, se ve que los fármacos seleccionados por el modelo basado en interacciones (en rojo, sin incluir los seleccionados por los modelos basados en fármacos), tampoco tienen una variedad estructural representativa del conjunto de fármacos de la base de datos DrugBank (azul).

iii) Gráfico fármacos Mito+ vs MitoTrain



Imagen 18: Cluster fármacos Mito+ (rojo) vs MitoTrain (azul)

En esta ocasión sí que se aprecia una distribución más equilibrada entre los tipos estructurales priorizados por los modelos basados en fármacos (Mito+, en rojo) y el training set que se creó para construir los modelos (MitoTrain, en azul). Es decir, los modelos basados en fármacos seleccionaron familias estructurales similares a las de los fármacos que se utilizaron para su creación, que era el comportamiento esperable.

iv) Gráfico fármacos mitoplus vs MitoTrain

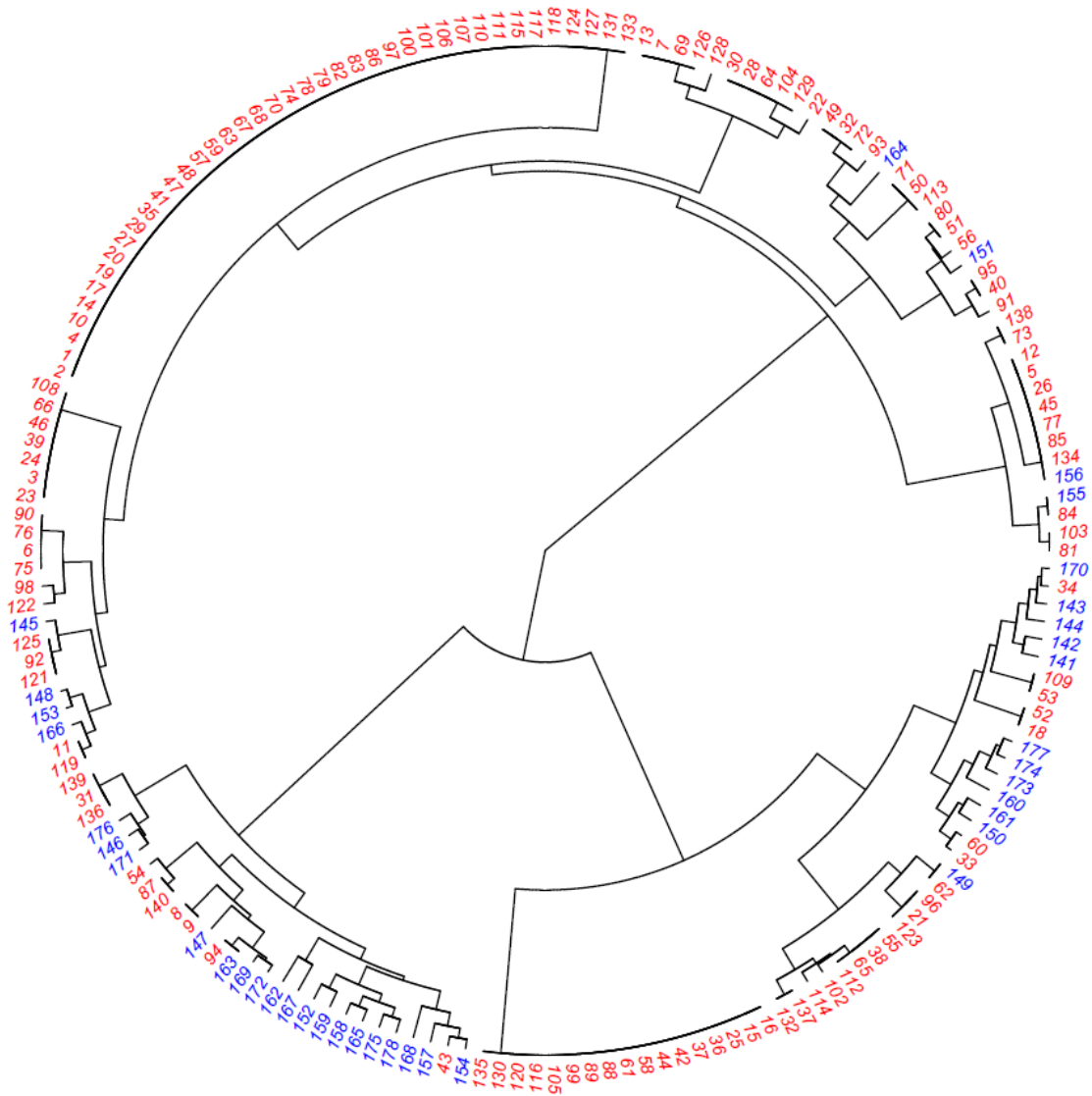


Imagen 19: Cluster fármacos mitoplus (rojo) vs MitoTrain (azul)

En esta ocasión se observa que en la selección de fármacos *mitoplus* (en rojo) están sobrerrepresentados unos tipos estructurales respecto a los que formaban parte de la selección MitoTrain (en azul, train set con el que se entrenaron los modelos basados en fármacos, constituido por fármacos con efecto positivo conocido en función mitocondrial). Si comparamos este resultado con el del gráfico anterior (Mito+ vs MitoTrain), en el que no se apreciaba ningún sesgo claro entre las estructuras de los dos grupos, se concluye que, tras las selecciones de los fármacos priorizados por los modelos basados en fármacos (Mito+) que a su vez se clasificaron como candidatos a interactuar con dianas mitocondriales por el modelo de interacciones (*mitoplus*), solo un pequeño grupo de familias estructurales son las que cumplen estos criterios de

selección. En el siguiente apartado se intentará identificar cuales son estas estructuras priorizadas por la acción de ambos modelos.

2. Identificación de los grupos estructurales obtenidos por los modelos:

A continuación se hará un resumen de los principales subgrupos, según su estructura, de las selecciones *mitoplus* e *Inter*. Para identificar los fármacos y visualizar su posición en los gráficos de agrupación jerárquica, se usará el número que los acompaña en cada gráfico. Para ver un análisis más exhaustivo de los tipos estructurales priorizados por los modelos y los nombres de cada fármaco consultar el Anexo PEC 3 - 2: Identificación de los grupos estructurales obtenidos por los modelos.

- Fármacos *mitoplus* (grafico “Fármacos mitoplus vs DrugBank”):

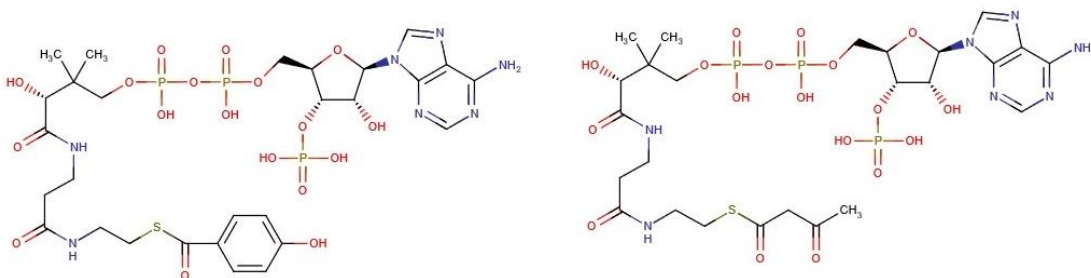


Imagen 20: Imágenes estructuras fármacos Cluster 140, 87

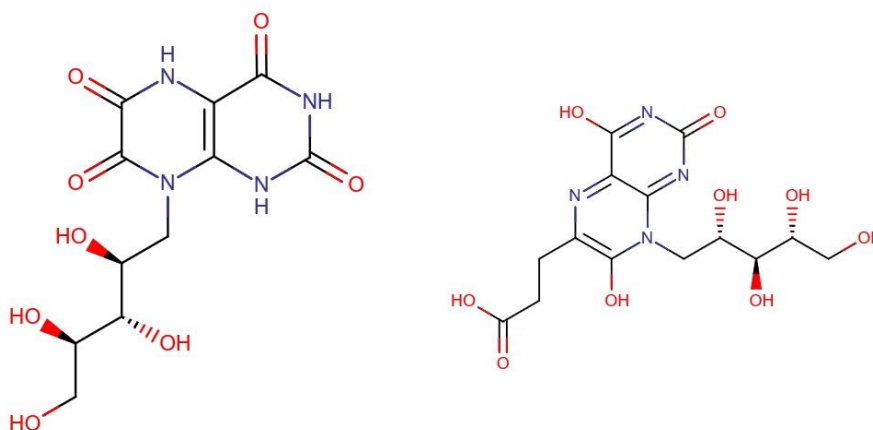


Imagen 21: Imágenes estructuras fármacos Cluster 136, 8

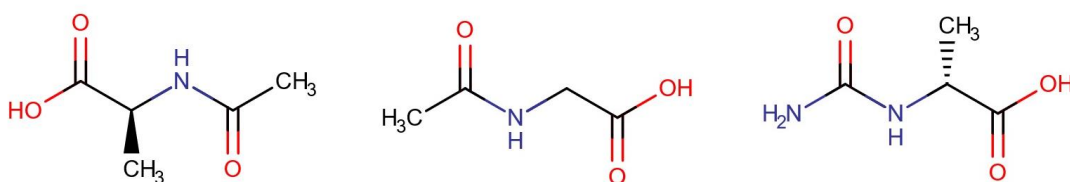


Imagen 22: Imágenes estructuras fármacos Cluster 33, 34, 60

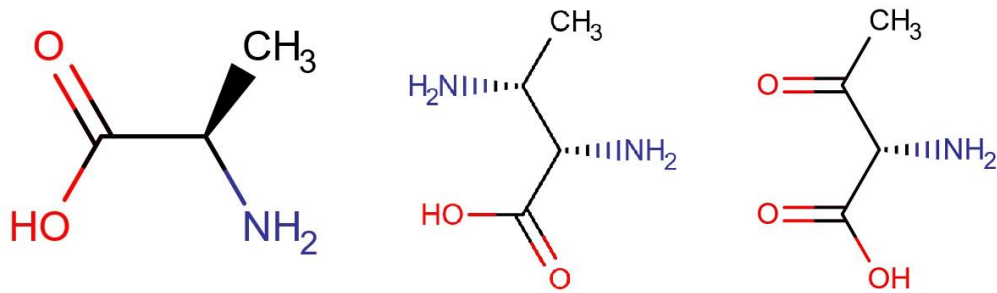


Imagen 23: Imágenes estructuras fármacos Cluster 16, 44, 130

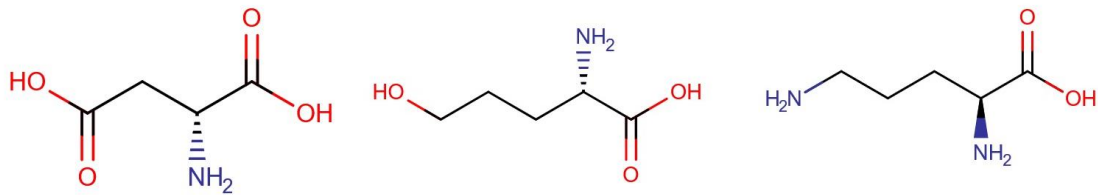


Imagen 24: Imágenes estructuras fármacos Cluster 2, 82, 133

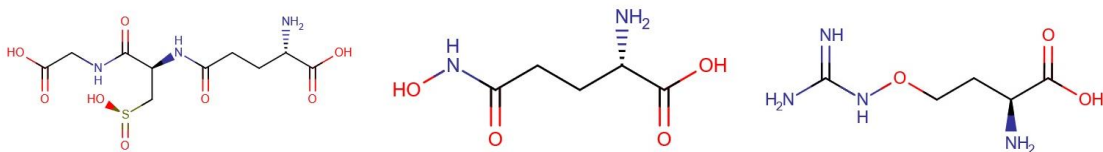


Imagen 25: Imágenes estructuras fármacos Cluster 138, 85, 103

En resumen, las principales estructuras que se encuentran en los fármacos *mitoplus* son:

- Un pequeño grupo de derivados de Acyl-CoA (Ej: 4-hydroxybenzoyl-CoA).
- Un pequeño grupo de pteridinas (Ej: 6,7-dioxo-5H-8-ribitylamino-lumazine).
- Un pequeño grupo de ácidos carboxílicos: (Ej: N-Carbamoyl-Alanine).
- Un gran grupo de aminoácidos y derivados de aminoácidos.

- Farmacos Inter (grafico “Fármacos Inter vs DrugBank”):

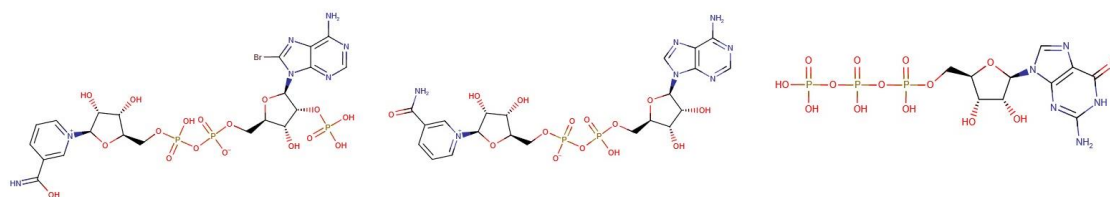


Imagen 26: Imágenes estructuras fármacos Cluster 57, 72, 16

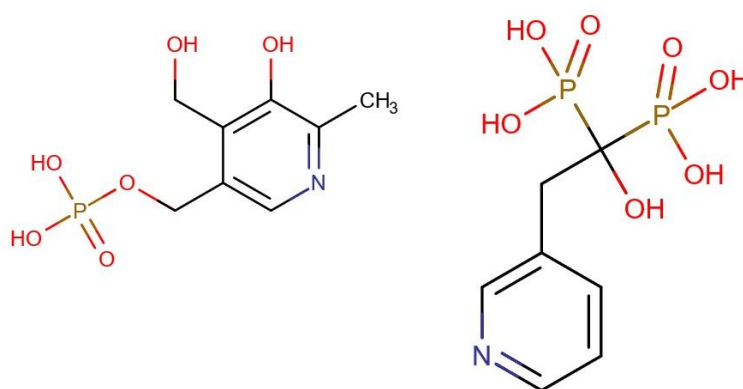


Imagen 27: Imágenes estructuras fármacos Cluster 74, 5

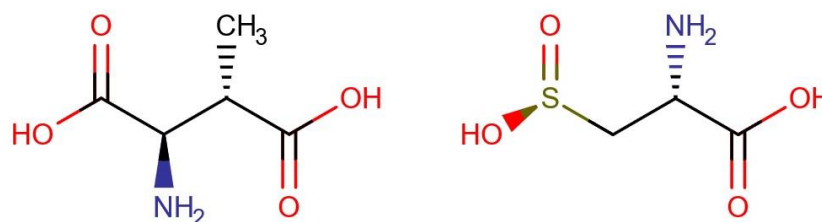


Imagen 28: Imágenes estructuras fármacos Cluster 66, 43

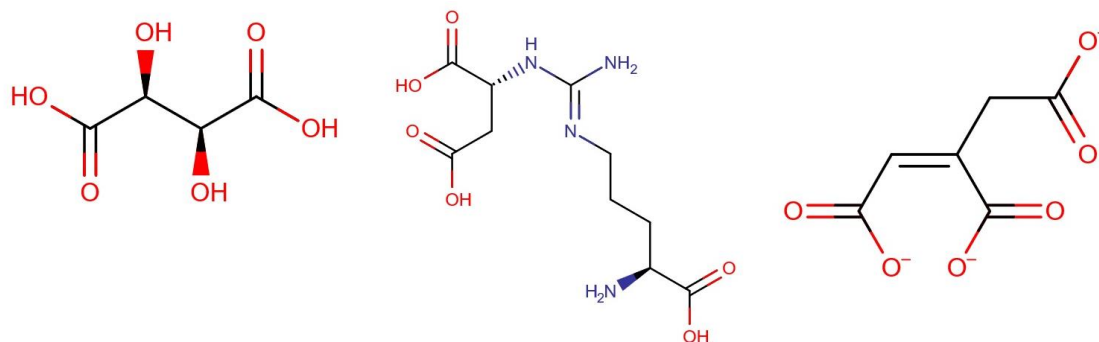


Imagen 29: Imágenes estructuras fármacos Cluster 50, 18, 10

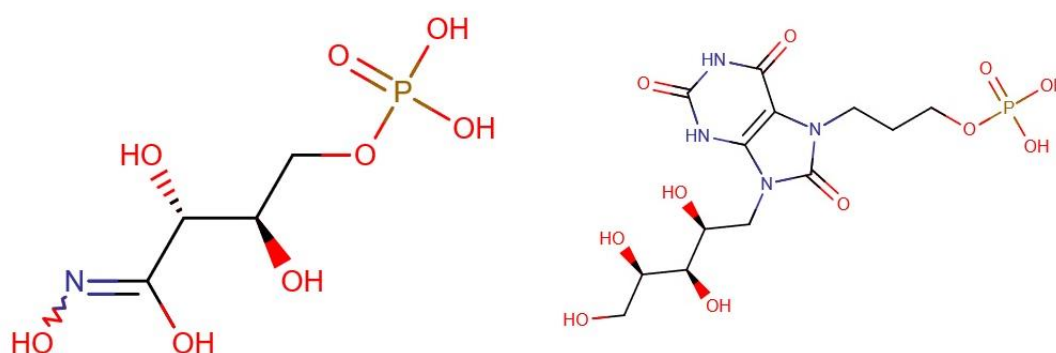


Imagen 30: Imágenes estructuras fármacos Cluster 9, 58

En resumen, las principales estructuras que se encuentran en los fármacos que interactúan con dianas mitocondriales (*Inter*) son:

- Un grupo de nucleótidos y derivados.
- Un pequeño grupo de bifosfonatos (Ej: Risedronic acid).
- Un grupo de aminoácidos y derivados.
- Un pequeño grupo de ácidos orgánicos (hidroxiácidos y ácidos dicarboxílicos).
- Un grupo de carbohidratos y derivados.

En ambas selecciones (*mitoplus* e *Inter*) se observa una gran cantidad de aminoácidos y derivados, especialmente en la selección *mitoplus* obtenida tras enriquecer los fármacos obtenidos por los modelos basados en fármacos con aquellos que presentaban interacción positiva con dianas mitocondriales según el modelo de interacciones. Este resultado sugiere que el modelo de interacciones ha elegido una elevada proporción de aminoácidos y derivados como fármacos de interacción positiva con dianas mitocondriales.

Inspeccionando la lista de proteínas usadas para crear las interacciones *UP_curated* (proteínas MitoCarta con entrada curada en UniProt) e investigando sus funciones biológicas se puede apreciar que un número elevado de estas proteínas son enzimas que catalizan reacciones en las que están involucrados diversos aminoácidos. Por ejemplo familias de enzimas relacionadas con la traducción de proteínas mitocondriales, como factores de iniciación, elongación y terminación, o la familia de tRNAs sintetasas mitocondriales (mt-aaRS2) forman parte de *UP_curated* e interaccionan con diversos aminoácidos para cumplir su función biológica.

A modo de ejemplo, y aprovechando que el nombre de los genes que lo constituyen siguen una nomenclatura similar, se va a comprobar que la familia de genes mt-aaRS2 (19 genes), cuyos nombres siempre terminan en *RS2³³, presentan interacción positiva con aminoácidos según el modelo de interacciones (el código para obtener la tabla se puede consultar en el Anexo PEC 3 - 2: Resumen grupos estructurales obtenidos por los modelos).

Symbol	DrugBank.ID
NARS2	DB01857
CARS2	DB00128
NARS2	DB00148
NARS2	DB02688
MARS2	DB02655
NARS2	DB04193
NARS2	DB03915
NARS2	DB03700

Tabla 6: Interacciones de las enzimas *mt-aaRS2*, según el modelo de interacciones

Los fármacos DB01857, DB00128, DB00148, DB02688, DB02655, DB04193, DB03915, DB03700 son: Phosphoaspartate, Aspartic acid, Creatine, 2,3-Didehydroalanine, D-Aspartic Acid, L-Homoserine, 2-Amino-3- Ketobutyric Acid, D-Threonine respectivamente, todos ellos aminoácidos o derivados de aminoácidos.

Como resumen se puede concluir que los fármacos *mitoplus*, obtenidos tras la priorización secuencial de los modelos basados en fármacos y el modelo basado en interacciones, es rico en fármacos con estructuras de aminoácidos y derivados, aunque también cuenta con otros tipos estructurales como por ejemplo ácidos carboxílicos y derivados de Acil-CoA.

En cuanto a los fármacos obtenidos teniendo en cuenta solo el modelo de interacciones (*Inter*) tenemos una selección más variada de tipos estructurales, con los grupos de aminoácidos y nucleótidos como principales grupos representados, existiendo también otras estructuras representadas tales como ácidos orgánicos y carbohidratos.

En cualquier caso los modelos han priorizado principalmente compuestos orgánicos, no existiendo apenas representación de otro tipo de estructuras.

3. Búsqueda bibliográfica de los fármacos priorizados con potencial efecto positivo mitocondrial:

En este apartado se realizará una búsqueda bibliográfica sobre los fármacos priorizados por los modelos de ML (*mitoplus*) y su posible efecto beneficioso sobre la actividad mitocondrial. Solo se citarán artículos sobre compuestos que no hayan sido incluidos en el set de fármacos MitoTrain, a partir del cual se crearon los train y test set de los modelos basados en fármacos. Para consultar la lista completa de los 68 fármacos *mitoplus* consultar el Anexo PEC 3 - 2: Búsqueda bibliográfica de los fármacos priorizados con potencial efecto positivo mitocondrial.

Se encontraron las siguientes referencias bibliográficas que apoyan un efecto funcional positivo sobre la actividad mitocondrial de los siguientes compuestos *mitoplus*:

- D-Cysteine: En el artículo de Shibuya *et al*³⁴ hallan un efecto reductor de D-Cisteína sobre el estrés oxidativo mitocondrial en cultivo primario de neuronas cerebelares.
- Malic Acid: En el artículo de Wu *et al*³⁵ hallan una mejora significativa de la capacidad física y parámetros bioquímicos relacionados con fatiga muscular en un modelo animal (ratón) tras la administración de ácido málico vía un aumento de producción energética mitocondrial.
- L-Glutamine: En el artículo de Ahmad *et al*³⁶ estudiaron el papel protector de la glutamina en células humanas sometidas a condiciones de estrés oxidativo (hiperoxia). Concluyeron que la suplementación de glutamina aumentaba la supervivencia de las células en condiciones de hiperoxia respecto a las células sin glutamina. También hallaron un incremento significativo de la producción de ATP de las células suplementadas con glutamina tanto en condiciones de hiperoxia como a concentraciones normales de oxígeno.
- Folic acid: Se trata de una revisión bibliográfica de Ormazabal *et al*³⁷ en la que revisan el papel que juega el ácido fólico en el contexto de las enfermedades mitocondriales. Se comentan algunos de los resultados publicados de pacientes con enfermedad mitocondrial a los que se les trató con ácido fólico, siendo los resultados variables: en algunos casos hubo mejoría clínica, mientras que en otros el tratamiento con ácido fólico no mejoró el curso de la enfermedad.
- Beta-Alanine: En un estudio de Schnuck *et al*³⁸ investigan el efecto de la beta-alanina en células musculares (mioblastos procedentes de ratón). Como resultados más destacables encuentran un aumento de la biogénesis mitocondrial (vía aumento en la expresión de TFAM y PPAR β/δ) y un aumento de la tasa de consumo de oxígeno en las células tratadas con beta-alanina respecto a células control.

- Glycine: En el estudio de Heidari *et al*³⁹ investigaron el efecto funcional en mitocondrias procedentes de hígado del suministro de glicina en un modelo animal (ratón) de colestasis hepática. En el estudio encontraron un aumento en el potencial de membrana mitocondrial y en la producción de ATP en ratones tratados con glicina y un descenso en la producción de especies reactivas de oxígeno y peroxidación lipídica en los ratones tratados respecto a controles sin tratamiento.

- Beta-Amino Isobutyrate. Se trata de una revisión de Tanianskii *et al*⁴⁰ en la que repasan los estudios relacionados con el ácido beta-amino isobutírico en el contexto del síndrome metabólico. Respecto a los hallazgos relativos a función mitocondrial los autores concluyen que el ácido beta-amino isobutírico produce un aumento hepático de la beta-oxidación mitocondrial y de la expresión de CPT1, un aumento en la biogénesis mitocondrial vía aumento en la expresión PGC-1 α y una disminución en la producción mitocondrial de especies reactivas de oxígeno.

Por último, y aunque no está relacionado con ninguno de los fármacos priorizados en concreto, se referencia el trabajo de Nisoli *et al*⁴¹ en el que se revisan los beneficios de la suministración de diferentes aminoácidos en la biogénesis y producción de ATP mitocondrial.

Sólo se ha encontrado un artículo que relacione uno de los fármacos *mitoplus* con un efecto mitocondrial negativo. Se trata de el estudio de Armagan *et al*⁴² en el que encuentran que la administración de D-serina en ratas produce un aumento de estrés oxidativo mitocondrial y una disminución en la función mitocondrial en tejido cerebral.

También se revisó, aunque no tan exhaustivamente como se ha hecho con los fármacos *mitoplus*, algunos de los fármacos priorizados sólo por el modelo basado en interacciones (fármacos *Inter*). Se encontraron dos publicaciones relacionadas con estos fármacos:

- Pyridoxine phosphate: En el artículo de Kannan *et al*⁴³ estudian el efecto protector de la vitamina B6 y sus derivados en la función mitocondrial, encontrando una disminución en la producción de especies reactivas de oxígeno y en la peroxidación lipídica en células monocíticas.

- Risedronic acid: En el artículo de Matsumoto *et al*⁴⁴ concluyen que el risedronato induce la apoptosis de osteoclastos mediante una ruta de señalización mitocondrial.

10. Conclusiones

Tras el análisis de los resultados obtenidos las principales conclusiones son:

- Se ha demostrado, siguiendo la estrategia de agrupación jerárquica de estructuras que, a pesar de contar con un número de entradas muy escaso para su entrenamiento, los modelos basados en fármacos han seleccionado

fármacos con estructuras similares a los fármacos del train set con los que se construyeron los modelos (MitoTrain).

- El modelo RF basado en interacciones obtuvo un buen rendimiento clasificador en el test set y se ha comprobado que es capaz de realizar predicciones con cierto sentido biológico: tomando como ejemplo la familia de genes *mt-aaRS2*, se asociaron aminoácidos o derivados de aminoácidos a dicha familia de genes, cuya función biológica, como aminoacil tRNA sintetasas, es la unión de aminoácidos a sus tRNAs correspondientes.

- Tras una revisión bibliográfica de los 68 fármacos priorizados *mitoplus* y su posible relación con función mitocondrial se han encontrado referencias sobre 7 fármacos del set que apoyan su posible efecto beneficioso en la actividad mitocondrial. Solo se ha hallado una referencia de un fármaco del set *mitoplus* en el que se describe un posible efecto perjudicial sobre la actividad mitocondrial, por lo que globalmente se puede concluir que el set de fármacos priorizados *mitoplus* podría contener fármacos que tengan un potencial efecto beneficioso en EM.

Como principales limitaciones de los modelos se puede mencionar:

- Respecto a los modelos basados en fármacos, un número demasiado escaso de fármacos con los que se construyeron (limitación inherente al escaso número de fármacos disponibles en estudio o para el tratamiento de EM). Al hacer la búsqueda bibliográfica de los 68 fármacos *mitoplus* priorizados se comprobó que no se había incluido un fármaco que cumplía los criterios de inclusión en el modelo: Ácido Fólico³⁷ (compuestos usados actualmente para el tratamiento de enfermedades mitocondriales, los criterios de inclusión se pueden consultar en el capítulo 2 de la memoria).

- Respecto al modelo de interacciones, aparte de la limitación comentada en el capítulo 4 (no tener en cuenta los tipos de interacciones posibles), otro posible defecto es no haber seleccionado más exhaustivamente que proteínas (espacio terapéutico) incluir en el modelo para evitar sesgos. Por ejemplo la existencia de un elevado número de proteínas mitocondriales que interactúan con aminoácidos (pertenecientes, por ejemplo, a la maquinaria de traducción mitocondrial) es posible que haya hecho que el modelo de interacciones seleccione preferentemente aminoácidos sobre otro tipo de estructuras. Es posible, a la vista de los resultados, que un estudio más pormenorizado de las proteínas a incluir en el modelo arrojase una selección de fármacos menos sesgada.

Por último, respecto a la consecución de los objetivos y resultados logrados en el trabajo se ha podido cumplir la mayoría de objetivos planteados en la planificación original (Ver Anexo PEC 1 - Plan Trabajo). Finalmente no se realizó el modelo de interacciones basado en el trabajo de Zhao *et al*⁴⁵, porque se decidió que sería más coste-efectivo, teniendo en cuenta el tiempo disponible, centrarse en la mejora de los modelos de ML convencional y en el análisis de los resultados obtenidos con ellos.

11. Glosario

EM	Enfermedades Mitocondriales
OXPPOS	Fosforilación oxidativa
ML	Machine Learning
Mitop / MitoTrain	Fármacos con un efecto positivo demostrado en enfermedades mitocondriales.
Mito+	Fármacos seleccionados por los modelos basados en fármacos, con un posible efecto positivo en EM.
Mitoplus	Fármacos seleccionados por ambos tipos de modelos (basados en fármacos y basado en interacciones).
Inter	Fármacos seleccionados por el modelo de interacciones no presentes en la selección de los modelos basados en fármacos (no Mito+).

12. Bibliografía

1. Craven L, Alston CL, Taylor RW, Turnbull DM. Recent Advances in Mitochondrial Disease. *Annu Rev Genomics Hum Genet.* 2017 Aug 31;18:257-275.
2. Rahman J, Rahman S. Mitochondrial medicine in the omics era. *Lancet.* 2018 Jun 23;391(10139):2560-2574.
3. El-Hattab AW, Zarante AM, Almannai M, Scaglia F. Therapies for mitochondrial diseases and current clinical trials. *Mol Genet Metab.* 2017 Nov;122(3):1-9.
4. Domínguez-González C, Madruga-Garrido M, Mavillard F, Garone C, Aguirre-Rodríguez FJ, Donati MA, Kleinsteuber K, Martí I, Martín-Hernández E, Morealejo-Aycinena JP, Munell F, Nascimento A, Kalko SG, Sardina MD, Álvarez Del Vayo C, Serrano O, Long Y, Tu Y, Levin B, Thompson JLP, Engelstad K, Uddin J, Torres-Torronteras J, Jimenez-Mallebrera C, Martí R, Paradas C, Hirano M. Deoxynucleoside Therapy for Thymidine Kinase 2-Deficient Myopathy. *Ann Neurol.* 2019 Aug;86(2):293-303.
5. Sahdeo S, Tomilov A, Komachi K, Iwahashi C, Datta S, Hughes O, Hagerman P, Cortopassi G. High-throughput screening of FDA-approved drugs using oxygen biosensor plates reveals secondary mitofunctional effects. *Mitochondrion.* 2014 Jul;17:116-25.
6. Vanhaelen Q, Mamoshina P, Aliper AM, Artemov A, Lezhnina K, Ozerov I, Labat I, Zhavoronkov A. Design of efficient computational workflows for in silico drug repurposing. *Drug Discov Today.* 2017 Feb;22(2):210-222.
7. Ekins S, Williams AJ, Krasowski MD, Freundlich JS. In silico repositioning of approved drugs for rare and neglected diseases. *Drug Discov Today.* 2011 Apr;16(7-8):298-310.
8. Calvo SE, Clauser KR, Mootha VK. MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. *Nucleic Acids Res.* 2016 Jan 4;44(D1):D1251-7.
9. Cichonska A, Rousu J, Aittokallio T. Identification of drug candidates and repurposing opportunities through compound-target interaction networks. *Expert Opin Drug Discov.* 2015 Dec;10(12):1333-45.
10. Mestres J, Gregori-Puigjané E. Conciliating binding efficiency and polypharmacology. *Trends Pharmacol Sci.* 2009 Sep;30(9):470-4.
11. Pujol A, Mosca R, Farrés J, Aloy P. Unveiling the role of network and systems biology in drug discovery. *Trends Pharmacol Sci.* 2010 Mar;31(3):115-23.
12. Signes A, Fernandez-Vizarra E. Assembly of mammalian oxidative phosphorylation complexes I-V and supercomplexes. *Essays Biochem.* 2018 Jul 20;62(3):255-270.
13. Viscomi C., Zeviani M. (2019) Experimental Therapies. In: Mancuso M., Klopstock T. (eds) *Diagnosis and Management of Mitochondrial Disorders.* Springer, Cham.

14. Cao D, Xiao N, Xu Q, Chen AF (2015). "Rcpi: R/Bioconductor package to generate various descriptors of proteins, compounds and their interactions." *Bioinformatics*, 31(2), 279–281.
15. Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, 28(5), 1 - 26.
16. Aiyar RS, Bohnert M, Duvezin-Caubet S, Voisset C, Gagneur J, Fritsch ES, Couplan E, von der Malsburg K, Funaya C, Soubigou F, Courtin F, Suresh S, Kucharczyk R, Evrard J, Antony C, St Onge RP, Blondel M, di Rago JP, van der Laan M, Steinmetz LM. Mitochondrial protein sorting as a therapeutic target for ATP synthase disorders. *Nat Commun*. 2014 Dec 18;5:5585.
17. Biesemann, N., Ried, J.S., Ding-Pfennigdorff, D. et al. High throughput screening of mitochondrial bioenergetics in human differentiated myotubes identifies novel enhancers of muscle performance in aged mice. *Sci Rep* 8, 9408 (2018).
18. Civileto G, Dogan SA, Cerutti R, Fagiolari G, Moggio M, Lamperti C, Benincá C, Viscomi C, Zeviani M. Rapamycin rescues mitochondrial myopathy via coordinated activation of autophagy and lysosomal biogenesis. *EMBO Mol Med*. 2018 Nov;10(11).
19. Kamatani N, Kushiya A, Toyo-Oka L, Toyo-Oka T. Treatment of two mitochondrial disease patients with a combination of febuxostat and inosine that enhances cellular ATP. *J Hum Genet*. 2019 Apr;64(4):351-353.
20. Kanabus M, Fassone E, Hughes SD, Biloei SF, Rutherford T, Donnell MO, Heales SJR, Rahman S. The pleiotropic effects of decanoic acid treatment on mitochondrial function in fibroblasts from patients with complex I deficient Leigh syndrome. *J Inher Metab Dis*. 2016 May;39(3):415-426.
21. Leipnitz, G., Mohsen, A., Karunanidhi, A. et al. Evaluation of mitochondrial bioenergetics, dynamics, endoplasmic reticulum-mitochondria crosstalk, and reactive oxygen species in fibroblasts from patients with complex I deficiency. *Sci Rep* 8, 1165 (2018)
22. Matsushashi T, Sato T, Kanno SI, Suzuki T, Matsuo A, Oba Y, Kikusato M, Ogasawara E, Kudo T, Suzuki K, Ohara O, Shimbo H, Nanto F, Yamaguchi H, Saigusa D, Mukaiyama Y, Watabe A, Kikuchi K, Shima H, Mishima E, Akiyama Y, Oikawa Y, Hsin-Jung HO, Akiyama Y, Suzuki C, Uematsu M, Ogata M, Kumagai N, Toyomizu M, Hozawa A, Mano N, Owada Y, Aiba S, Yanagisawa T, Tomioka Y, Kure S, Ito S, Nakada K, Hayashi KI, Osaka H, Abe T. Mitochondrial Acid 5 (MA-5) Facilitates ATP Synthase Oligomerization and Cell Survival in Various Mitochondrial Diseases. *EBioMedicine*. 2017 Jun;20:27-38.
23. Orsucci D, Ienco EC, Siciliano G, Mancuso M. Mitochondrial disorders and drugs: what every physician should know. *Drugs Context*. 2019 Jul 4;8:212588.
24. Morén C, Juárez-Flores DL, Cardellach F, Garrabou G. The Role of Therapeutic Drugs on Acquired Mitochondrial Toxicity. *Curr Drug Metab*. 2016;17(7):648-62.
25. Finsterer J, Segall L. Drugs interfering with mitochondrial disorders. *Drug Chem Toxicol*. 2010 Apr;33(2):138-51.

26. Hargreaves IP, Al Shahrani M, Wainwright L, Heales SJ. Drug-Induced Mitochondrial Toxicity. *Drug Saf.* 2016 Jul;39(7):661-74.
27. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, Li Q, Shoemaker BA, Thiessen PA, Yu B, Zaslavsky L, Zhang J, Bolton EE. PubChem 2019 update: improved access to chemical data. *Nucleic Acids Res.* 2019 Jan 8; 47(D1):D1102-1109.
28. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, Sajed T, Johnson D, Li C, Sayeeda Z, Assempour N, Iynkkaran I, Liu Y, Maciejewski A, Gale N, Wilson A, Chin L, Cummings R, Le D, Pon A, Knox C, Wilson M. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* 2017 Nov 8.
29. The UniProt Consortium, UniProt: a worldwide hub of protein knowledge, *Nucleic Acids Research*, Volume 47, Issue D1, 08 January 2019, Pages D506–D515.
30. Yates B, Braschi B, Gray K, Seal R, Tweedie S, Bruford E. Genenames.org: the HGNC and VGNC resources in 2017. *Nucleic Acids Res.* 2017 Jan 4; 45(D1):D619-625.
31. Ong SA, Lin HH, Chen YZ, Li ZR, Cao Z. Efficacy of different protein descriptors in predicting protein functional families. *BMC Bioinformatics.* 2007 Aug 17;8:300.
32. Pedregosa F, Gramfort A, Varoquaux G. [online] "Scikit-Learn: RBF Svm Parameters." URL: https://scikit-learn.org/stable/auto_examples/svm/plot_rbf_parameters.html [Accessed 26 Nov. 2019].
33. Yates, B. HUGO Gene Nomenclature Committee. [online] Gene group: Aminoacyl tRNA synthetases. URL: <https://www.genenames.org/data/genegroup/#!/group/130> [Accessed 4 Jan. 2020].
34. Shibuya N, Koike S, Tanaka M, Ishigami-Yuasa M, Kimura Y, Ogasawara Y, Fukui K, Nagahara N, Kimura H. A novel pathway for the production of hydrogen sulfide from D-cysteine in mammalian cells. *Nat Commun.* 2013;4:1366.
35. Wu JL, Wu QP, Huang JM, Chen R, Cai M, Tan JB. Effects of L-malate on physical stamina and activities of enzymes related to the malate-aspartate shuttle in liver of mice. *Physiol Res.* 2007;56(2):213-20. Epub 2006 Mar 23.
36. Ahmad S, White CW, Chang LY, Schneider BK, Allen CB. Glutamine protects mitochondrial structure and function in oxygen toxicity. *Am J Physiol Lung Cell Mol Physiol.* 2001 Apr;280(4):L779-91.
37. Ormazabal A, Casado M, Molero-Luis M, Montoya J, Rahman S, Aylett SB, Hargreaves I, Heales S, Artuch R. Can folic acid have a role in mitochondrial disorders? *Drug Discov Today.* 2015 Nov;20(11):1349-54.
38. Schnuck JK, Sunderland KL, Kuennen MR, Vaughan RA. Characterization of the metabolic effect of β -alanine on markers of oxidative metabolism and mitochondrial biogenesis in skeletal muscle. *J Exerc Nutrition Biochem.* 2016 Jun;20(2):34-41.
39. Heidari R, Ghanbarinejad V, Mohammadi H, Ahmadi A, Ommati MM, Abdoli N, Aghaei F, Esfandiari A, Azarpira N, Niknahad H. Mitochondria protection as a mechanism underlying the hepatoprotective effects of glycine in cholestatic mice. *Biomed Pharmacother.* 2018 Jan;97:1086-1095.

40. Tanianskii DA, Jarzebska N, Birkenfeld AL, O'Sullivan JF, Rodionov RN. Beta-Aminoisobutyric Acid as a Novel Regulator of Carbohydrate and Lipid Metabolism. *Nutrients*. 2019 Feb 28;11(3). pii: E524.
41. Nisoli E, Cozzi V, Carruba MO. Amino acids and mitochondrial biogenesis. *Am J Cardiol*. 2008 Jun 2;101(11A):22E-25E.
42. Armagan G, Kanit L, Yalcin A. D-serine treatment induces oxidative stress in rat brain. *Drug Chem Toxicol*. 2011 Apr;34(2):129-38.
43. Kannan K, Jain SK. Effect of vitamin B6 on oxygen radicals, mitochondrial membrane potential, and lipid peroxidation in H₂O₂-treated U937 monocytes. *Free Radic Biol Med*. 2004 Feb 15;36(4):423-8.
44. Matsumoto T, Nagase Y, Iwasawa M, Yasui T, Masuda H, Kadono Y, Nakamura K, Tanaka S. Distinguishing the proapoptotic and antiresorptive functions of risedronate in murine osteoclasts: role of the Akt pathway and the ERK/Bim axis. *Arthritis Rheum*. 2011 Dec;63(12):3908-17.
45. Zhao S, Li S. Network-based relating pharmacological and genomic spaces for drug target identification. *PLoS One*. 2010 Jul 26;5(7):e11764.

13. Anexos

Archivos y documentos suplementarios a la memoria:

- PEC 1 - Plan Trabajo
- PEC 2
- PEC 3
- PEC 3 - 2.Rmd
- PEC 3 - 2
- Correlación variables Dataset Fármacos.R
- Dataset Farmacos completo.R
- Dataset completo farmacos_resto de farmacos.R
- Union de los subsets farmacos.R