

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **3 071 824**

21 Número de solicitud: 202431096

51 Int. Cl.:

G16B 15/30 (2009.01)

G16B 20/00 (2009.01)

G16B 40/00 (2009.01)

G16H 70/40 (2008.01)

12

SOLICITUD DE PATENTE

A1

22 Fecha de presentación:

24.12.2024

43 Fecha de publicación de la solicitud:

25.06.2026

71 Solicitantes:

UNIVERSIDAD DE MÁLAGA (75,00%)

Avenida Cervantes, 2

29071 Málaga (Málaga) ES y

**CONSORCIO CENTRO DE INVESTIGACIÓN
BIOMÉDICA EN RED M.P. (25,00%)**

72 Inventor/es:

GARCÍA RANEA, Juan Antonio;

PERKINS, James Richard;

DÍAZ SANTIAGO, Elena Dolores;

MOYA GARCÍA, Aurelio y

BERNAL MUÑOZ, Manuel

74 Agente/Representante:

RUBIO ROMERO, Juan Carlos

54 Título: **Método de predicción de los efectos fenotípicos de agentes moleculares con potencial terapéutico**

57 Resumen:

Método de predicción de los efectos fenotípicos de agentes moleculares con potencial terapéutico. La presente invención se refiere a métodos (y a sus usos) para la predicción e identificación de los efectos fenotípicos de agentes moleculares con potencial terapéutico que comprende la obtención de pares de asociaciones agente-fenotipos significativas a partir de al menos una red tripartita fenotipo-dominio-agente obtenida a partir de pares de asociaciones dominio-fenotipo significativas obtenidas a partir de al menos una red tripartita fenotipo-diana-dominio obtenida a partir de pares de asociaciones diana-dominio y de pares de asociaciones diana-fenotipo significativas, y de pares de asociaciones agente-dominio significativas obtenidas a partir de pares de asociaciones diana-dominio y de pares de asociaciones agente-diana; y de al menos una red tripartita fenotipo-diana-agente obtenida a partir de pares de asociaciones diana-fenotipo significativas obtenidas a partir de al menos una red tripartita de relaciones fenotipos-entidades-dianas, y de pares de asociaciones agente-diana.



FIGURA 1

ES 3 071 824 A1

DESCRIPCIÓN**MÉTODO DE PREDICCIÓN DE LOS EFECTOS FENOTÍPICOS DE AGENTES
MOLECULARES CON POTENCIAL TERAPÉUTICO****5 Campo técnico**

La presente invención está enmarcada en el ámbito del descubrimiento, síntesis y reposicionamiento de agentes moleculares con potencial terapéutico (fármacos, pequeños compuestos o moléculas, miRNAs, etc.). En particular se refiere a un método de predicción de los
10 efectos fenotípicos de diferentes agentes moleculares con potencial terapéutico, más particularmente a la predicción de aquellos efectos fenotípicos que se producen a través de la interacción de dichos agentes moleculares con potencial terapéutico con dianas biológicas funcionales (e.g. proteínas, genes, y dominios funcionales en proteínas y genes) relacionadas con entidades biológicas que pueden asociarse a su vez a uno o varios fenotipos (e.g. entidades como
15 son las enfermedades, síndromes, pacientes, líneas celulares, organoides, etc.).

Estado de la técnica

20 Aunque la mayoría de los fármacos interactúan con una diana primaria, también lo hacen frecuentemente con otras dianas. Se considera diana primaria a la responsable de los efectos deseados o terapéuticos del fármaco que interactúa con la misma; sin embargo, tanto las interacciones con la diana primaria como las interacciones con otras dianas (denominadas secundarias u “off-target”) pueden producir consecuencias fenotípicas adicionales, como son los
25 efectos secundarios (ES) y los efectos adversos (EA). En adelante nos referiremos de manera conjunta a los ES y EA como Efectos Secundarios y Adversos (ESA), y a cualquier efecto, tanto los deseados como los ESA, los denominaremos de manera conjunta como efecto fenotípico o fenotipo.

30 Los ESA son un gran problema durante el desarrollo de nuevos medicamentos, y en muchos casos, son la causa principal de la retirada del mercado de una gran variedad de fármacos, lo que puede conllevar implicaciones negativas en la salud de los pacientes, así como, importantes pérdidas económicas, ya que el coste medio estimado para llevar un nuevo fármaco al mercado es de cientos de millones de dólares, y muchos fármacos son retirados tras haberse invertido
35 grandes cantidades de dinero en su desarrollo y validación experimental. Por ello, es muy útil una metodología como la presentada en esta solicitud, basada en el modelado y análisis de datos en redes de asociación, que nos da la posibilidad de predecir con antelación a la validación experimental, y durante todo el desarrollo del fármaco, sus potenciales fenotipos asociados

incluyendo los ESA, así como la diana o conjunto de dianas relacionadas con cada uno de los distintos efectos fenotípicos asociados a la acción del fármaco.

En ocasiones, lo que se registra inicialmente como ES de un fármaco se identifica posteriormente un efecto deseado o terapéutico. El ejemplo más famoso es el Viagra, un medicamento cuyo principio activo es el compuesto denominado sildenafil, desarrollado en principio para el tratamiento de la angina de pecho, pero que finalmente fue comercializado para tratar la impotencia masculina al observarse que este fármaco inducía como efecto secundario la erección del pene. Al procedimiento por el que un fármaco existente es utilizado para el tratamiento de una nueva patología se denomina comúnmente como reposicionamiento. Hay trabajos previos en los que se estudia el reposicionamiento de fármacos para tratar distintas enfermedades basándose en su acción sobre dianas conocidas o predichas mediante diferentes aproximaciones, tales como, métodos que emplean la similitud de los perfiles de expresión génica asociados a la administración de los fármacos, o estudios que han analizado el reposicionamiento de fármacos basándose en las similitudes de sus perfiles de efectos secundarios.

Si bien existen diversos estudios que buscan detectar las interacciones entre fármacos y ESA directamente utilizando la literatura biomédica, pocos buscan entender las interacciones considerando las proteínas diana de los fármacos y los fenotipos que pueden provocar. Un estudio previo ha combinado los datos de interacciones fármaco-diana con los datos sobre cómo estas dianas moleculares pueden dar lugar a fenotipos patológicos, como una forma de identificar nuevas relaciones fármaco-ESA (*Nguyen et al. 2019*). Nguyen y colaboradores recopilaron datos para relacionar 1819 fármacos con 1046 dianas para mostrar cómo los genes que codifican las dianas de los fármacos podrían utilizarse para hacer predicciones sobre los efectos secundarios clínicos, como la afectación del sistema de órganos. Este trabajo fue fundamental para mostrar la validez de combinar datos de fenotipo de diana con datos de diana de fármacos, aunque difiere de manera fundamental de la metodología objeto de la presente invención, que se enfoca en determinar las relaciones fármaco-fenotipo a través de la acción a través de una o varias dianas del fármaco, e incluye la identificación y análisis de los dominios de las proteínas, al construir las redes, para vincular los fármacos, sus dianas y los posibles efectos secundarios mediados por ellas.

Por tanto, ninguna de las metodologías identificadas en el estado de la técnica implementa la identificación de asociaciones de codependencia o asociación fármaco-fenotipo modelando redes heterogéneas de asociaciones significativas para cuatro niveles conformados por (1) **agentes**, término que en el presente documento engloba toda una variedad de agentes moleculares con potencial terapéutico (e.g. fármacos; pequeños compuestos y moléculas; miRNA; etc.), mediante su interacción con (2) **dianas**, término que en el presente documento hace referencia a las

diferentes dianas biológicas (e.g. proteínas, genes y los dominios funcionales presentes dentro de dichas proteínas y genes), cuya disminución o falta de función se asocia a (3) **entidades biológicas**, tales como enfermedades, síndromes, pacientes, líneas celulares, organoides, etc., que a su vez se asocian a la observación en dichas entidades de (4) **fenotipos**, clasificados en el presente documento en deseados, adversos y secundarios.

La mayoría de los estudios anteriores sobre las relaciones fármacos-ESA han tendido a centrarse en las dianas proteicas. Sin embargo, en los últimos años se ha producido un fuerte impulso para considerar dichas dianas proteicas en términos de sus unidades constituyentes, estructural y evolutivamente independientes, denominadas dominios proteicos, lo que permite una asociación más específica de los fármacos con regiones específicas dentro de sus dianas. Otros trabajos han tratado de inferir nuevas interacciones mediante un enfoque basado en el aprendizaje, con inteligencia artificial, utilizando como grupo de entrenamiento un conjunto conocido de interacciones fármaco-dominio.

Los métodos antes descritos se centran en la acción de los fármacos sobre una única diana terapéutica o también denominada en inglés "*magic bullet*". Sin embargo, muchos fármacos producen sus efectos mediante su acción sobre múltiples dianas, dianas moleculares que realizan múltiples funciones en diferentes procesos biológicos, emergiendo así el término polifarmacología, no solo para describir esta capacidad de acción multidiana de los fármacos, si no también, la necesidad de tener en cuenta esta característica fundamental de los fármacos en el desarrollo o reposicionamiento de fármacos.

Hasta donde se conoce, ningún estudio previo ha utilizado una metodología como la que es objeto de la presente invención, basada en el modelado e integración de la red de relaciones fármaco-diana y de la red tripartita de fenotipos-enfermedades-dianas para predecir asociaciones entre fármacos y fenotipos. La presente invención soluciona la problemática indicada identificando asociaciones fármaco-fenotipo, que pueden utilizarse en la predicción de los efectos potenciales de los fármacos, incluidos tanto los ESA como los efectos terapéuticos o deseados, y obtener además información sobre los mecanismos moleculares subyacentes de una o varias dianas a través de las cuales se median estos efectos. Por otra parte, y en línea con lo ya comentado, es conocido que la mayoría de los fármacos fracasan durante su desarrollo, a menudo debido a que son asociados a efectos adversos imprevistos, no logrando pasar los controles de eficacia y seguridad implementados en las distintas fases del ensayo clínico. El ensayo clínico es una evaluación experimental de distintos productos con potencial aplicación terapéutica, entre los que se incluyen los fármacos. La implementación de un ensayo clínico para la evaluación de un nuevo fármaco implica un coste muy alto, sobre todo, en fases avanzadas del mismo. Es por ello por lo que existe una necesidad clara e insatisfecha de nuevos enfoques que permitan abaratar dichos

costes, identificando con antelación, o durante estadios tempranos del desarrollo de un nuevo fármaco, las probables relaciones de dicho fármaco con ESA, así como, orientando a los investigadores sobre los potenciales mecanismos moleculares a través de los cuales los fármacos ejercen sus efectos, tanto los terapéuticos o deseados, como los ESA

5

Explicación de la invención

La metodología objeto de la presente invención se basa en que los fenotipos, asociados con variantes hipo-funcionales o no funcionales de dianas biológicas (tales como, proteínas, genes o dominios funcionales dentro de dichas proteínas y genes), pueden producirse de igual forma cuando la función de dichas dianas se ven afectadas por su interacción (normalmente inhibitoria de la función) con agentes moleculares con potencial terapéutico, tales como, fármacos, pequeños compuestos, miRNA y cualquier otra molécula con capacidad de unirse a una o varias dianas. Por tanto, a partir del modelado de una red cuatripartita de asociaciones, con distinto grado de confianza, entre agentes-dianas-entidades-fenotipos, es posible identificar las relaciones matemáticamente significativas de pares agentes-fenotipos de entre todas las asociaciones espúreas presentes en el modelo de la red.

15

Conforme a lo anterior, para el modelado de la red cuatripartita, necesitamos:

20

- De una parte, de cuatro niveles o categorías de elementos, definidos anteriormente: (1) agentes, (2) dianas, (3) entidades biológicas y (4) fenotipos.
- De otra parte, necesitamos conocer las siguientes asociaciones entre los elementos de los distintos niveles.
 - i. Pares agentes-dianas: Para modelar el nivel de relaciones agentes-dianas que se integran en el modelo de red considerado en la presente invención necesitamos identificar la afinidad de unión entre el agente (compuesto) y su diana molecular. Esta afinidad debe ser suficientemente alta para garantizar una actividad biológica relevante. En términos prácticos, esto significa identificar interacciones donde el compuesto demuestre una potencia significativa sobre su diana, típicamente representada por valores de concentración por debajo de 1 μ M. Este umbral asegura que las interacciones incluidas en el modelo tengan relevancia biológica y potencial terapéutico real, excluyendo aquellas interacciones demasiado débiles que, aunque detectables, no serían prácticamente útiles en un contexto farmacológico.
 - ii. Pares dianas-entidades: Las relaciones entre dianas y entidades proceden de observaciones experimentales y de relación estadística, que asocian mutaciones

25

30

35

o variantes de dichas dianas (e.g. genes y proteínas) con una cohorte de pacientes que definen una enfermedad, con pacientes individuales, tipos celulares, organoides, etc. donde se encuentran presentes dichas mutaciones o variantes.

- iii. Pares entidades-fenotipos: Las entidades se pueden asociar a su vez con un conjunto de fenotipos que definen dicha entidad como enfermedad o síndrome, así como fenotipos observados en pacientes, células, organoides, etc., que presentan en su acervo genético distintas variantes o mutaciones de las dianas estudiadas.

Integrando los tres tipos de asociaciones (i, ii y iii) es posible modelar una red cuatripartita uniendo los cuatro niveles de elementos (1, 2, 3 y 4) en una sola red con la siguiente topología: agentes-dianas-entidades-fenotipos. No obstante, hay que tener en cuenta que las asociaciones tienen dos propiedades derivadas de (A) su propia naturaleza y de (B) su integración en una sola red.

(A) Error en las asociaciones. Todas las asociaciones antes descritas (i, ii, y iii) proceden directamente de observaciones experimentales y/o de tratamientos estadísticos sobre los datos obtenidos en dichas observaciones experimentales. Por tanto, por su origen dichas asociaciones tienen distinto grado de confianza que en muchos casos no está ponderado. Esta falta de ponderación puede explicarse por varios factores: los fenotipos asociados a una enfermedad no tienen la misma penetrancia en la cohorte de pacientes afectados, y esta penetrancia raramente está documentada; los fármacos se unen con distinta intensidad a sus dianas, lo que afecta su capacidad de inhibición funcional, y las medidas experimentales de esta interacción conllevan una varianza y margen de error; además, muchas de las relaciones entre mutaciones o variantes de genes y proteínas observadas en pacientes, y su asociación con una enfermedad, carecen de validación funcional. Por tanto, muchas de las asociaciones en los tres niveles del modelo expuesto en este documento pueden no estar correctamente ponderadas o incluso algunas pueden ser falsas.

(B) Asociaciones emergentes. En segundo lugar, existen asociaciones emergentes cuya ponderación estadística sólo pueden ser identificadas integrando y computando todos los pares de asociaciones en una sola red. Por ejemplo, el efecto polifarmacológico, antes descrito, de los agentes sobre distintas dianas que a su vez pueden estar asociadas a distintas entidades (es decir, por ejemplo, enfermedades), las cuales pueden converger en el mismo o mismos fenotipos. O distintas dianas, que, al formar parte de un mismo sistema funcional, por ejemplo el mismo "pathway" metabólico, convergen asociándose a una misma enfermedad

o conjunto de enfermedades. En estos casos, al integrar todas las asociaciones puede incrementar la señal de codependencia entre un agente y uno o un conjunto de fenotipos permitiendo identificar con un mayor nivel de significancia relaciones que de otro modo no serían observables en el análisis no integrado de los pares de asociación.

Con la intención de computar correctamente las dos propiedades (A y B) de las asociaciones antes descritas, se han diseñado e implementado cálculos matemáticos de codependencia en sucesivas subredes tripartitas que permitan ponderar y seleccionar a distintos niveles de significancia las asociaciones entre elementos no directamente relacionados en la red, como son las asociaciones agentes-dominios funcionales, y dianas-fenotipos. El método comprende dos pasos principales:

- (I). Calcular la significancia estadística de las asociaciones diana-fenotipo basándose en la subred diana-entidad-fenotipo. Se seleccionan las asociaciones indirectas diana-fenotipo más significativas estadísticamente.
- (II). Integrar las asociaciones seleccionadas en (I) con las asociaciones directas agentes-dianas para crear una red tripartita agentes-dianas-fenotipos. En esta red se calcula la significancia de las asociaciones indirectas agentes-fenotipos, utilizando el mismo método de cálculo de codependencia empleado en el primer paso.

De conformidad con lo anterior, es objeto de la presente invención un método de predicción de los efectos fenotípicos de diferentes agentes con potencial terapéutico (e.g. fármacos; pequeños compuestos y moléculas; miRNA; etc.) a través de su interacción con dianas biológicas (e.g. proteínas, genes, y dominios funcionales en proteínas y genes) relacionadas con entidades biológicas y/o patológicas (e.g. entidades como son las enfermedades, síndromes, pacientes, líneas celulares, organoides, etc.) que pueden asociarse a su vez a uno o varios fenotipos, con la finalidad de identificar asociaciones significativas entre dichos agentes y los fenotipos, tanto deseados, secundarios como adversos que comprende:

- a) una etapa de obtención de pares diana-fenotipo para construir una red tripartita de relaciones fenotipos-entidades-dianas sobre la que se calcula un índice, o cualquier otra medida de asociación, para obtener asociaciones diana-fenotipo significativas.
- b) una etapa de obtención de pares de relaciones agente-dominio de una base de datos de dominios funcionales de dichos agentes, pares que se combinan con los pares diana-fenotipo de la etapa anterior (a) para construir la red tripartita fenotipo-diana-dominio, red con la que se calculan los pares de asociaciones dominio-fenotipo significativas de forma similar a como se realiza para calcular las asociaciones diana-fenotipo significativas;
- c) una etapa de obtención de pares agente-diana, en donde los pares agente-diana se

obtienen a partir de la base de datos o repositorio de datos de origen experimental, como por ejemplo ChEMBL; y donde los pares agente-dominio se obtienen integrando los pares diana-dominio con los pares agente-diana del paso mediante cualquier procedimiento de cálculo de relaciones agente-dominio significativas; y

- 5 d) una etapa de obtención de pares diana-fenotipo, donde se utilizan los pares obtenidos en las etapas anteriores para construir redes tripartitas fenotipos-dianas-agentes, sobre las que se calculan de manera independiente o combinada las asociaciones significativas entre agentes y fenotipos utilizando un índice, o cualquier otra medida de asociación.

10 Alternativamente a lo anterior, uno o más de los pares de asociación y/o una o más de las redes tripartitas ya están disponibles y no se precisa su obtención ni, incluso el cálculo de asociaciones o relaciones significativas, reduciéndose el método a identificar y analizar el nivel de asociación o significancia de los pares que impliquen directa (caso de agentes conocidos) o indirectamente (caso de agentes no conocidos) al agente molecular con potencial terapéutico cuyos efectos
15 fenotípicos se quieran prever. Asimismo, es posible aplicar el método objeto de la invención realizando una única vez, o de forma puntual, cada una de las distintas etapas (a) a (d) o todas o solo algunas de ellas (en cuyo caso se utilizarían los pares / redes previamente disponibles), si bien una realización preferente implicaría la realización de todas las etapas (a) a (d) cada vez que se ponga en práctica el método objeto de la invención, para así disponer de toda la información
20 necesaria en su estado más actualizado, completo y vigente posible.

En otra realización preferida de la invención, en la etapa de obtención de pares agente-diana, cada agente (e.g. proteína codificada por cada gen de la lista de pares fármaco-gen en la prueba de concepto) puede comprender uno o más dominios funcionales (e.g. dominios FunFams), dominios
25 estructurales 3D, funcionales y/o de relación filogenética, y, a su vez, un mismo tipo de dominio puede estar presente en múltiples agentes, de tal forma que para cada dominio se evaluaría si está significativamente sobrerrepresentado entre las dianas de un determinado agente, utilizando métodos de cálculo o estimación de la significancia de la relación agente-dominio, como por ejemplo, la probabilidad esperada y la prueba binomial utilizadas, y si este es el caso, se asocia
30 dicho agente a ese dominio, creando un par agente-dominio.

En otra realización preferida de la invención, en la etapa de obtención de pares agente-fenotipo, se utilizan diferentes umbrales del índice, o cualquier otra medida de asociación, para evaluar el rendimiento del método basado en el número de asociaciones agente-fenotipo verdaderas
35 encontradas a diferentes umbrales, comparando con una aproximación aleatoria, y utilizando para dicha validación grupos control de asociaciones agente-fenotipo conocidas, tales como, aunque no excluyentes, grupos de datos procedentes de la base de datos de medicamentos en

comercialización en las que se registran sus correspondientes ESA (por ejemplo, base de datos SIDER) y o de la bibliografía científica.

En otra realización preferida de la invención, la evaluación del solapamiento entre las asociaciones agente-fenotipo obtenidas, y los efectos secundarios conocidos del agente comprende comparar los datos de todos los efectos (fenotipos) registrados en una base de datos con las asociaciones agente-fenotipo obtenidas mediante las aproximaciones basadas en dominios y dianas, implementadas de manera independiente o combinada.

En otra realización preferida de la invención, el método de predicción puede incluir una etapa de validación que comprende la confrontación de los resultados obtenidos por el método de predicción con otros predictores, más preferentemente con un predictor o modelo aleatorio.

En la medida en que el método de predicción objeto de la presente invención constituye un método implementado / ejecutado en un computador o en una combinación de recursos electrónicos e informáticos (los recursos informáticos pueden ser un conjunto de instrucciones, por ejemplo, un programa de ordenador; y los recursos electrónicos pueden ser cualquier circuito electrónico capaz de implementar las correspondientes etapas del método), constituyen aspectos adicionales de la invención tanto (1) aparatos, dispositivos o sistemas informáticos adaptados/configurados o que comprenden medios para llevar a cabo las etapas del método de predicción que constituye un primer objeto de la invención; como (2) programas informáticos o instrucciones de programa, más particularmente a programas informáticos en o sobre unos medios portadores, adaptados para poner en práctica, al ejecutarse en aparatos, dispositivos o sistemas informáticos anteriores, el método de predicción que constituye un primer objeto de la invención; como (3) combinaciones de recursos electrónicos (1) e informáticos (2).

El programa informático puede estar en forma de código fuente, de código objeto o en un código intermedio entre código fuente y código objeto, tal como en forma parcialmente compilada, o en cualquier otra forma adecuada para usar en la implementación del método de predicción que constituye un primer objeto de la invención. El medio portador o soporte del programa informático puede ser cualquier entidad o dispositivo capaz de portar el programa. Por ejemplo, el medio portador puede comprender un medio de almacenamiento legible por un computador, tal como, entre otros, una ROM, por ejemplo un CD ROM o una ROM semiconductora, o un medio de grabación magnético, por ejemplo un "floppy disc" o un disco duro, o un DVD, o una unidad USB. Además, el medio portador puede ser un medio portador transmisible tal como una señal eléctrica u óptica que puede transmitirse vía cable eléctrico u óptico o mediante radio u otros medios. Cuando el programa de ordenador esté contenido en una señal que puede transmitirse directamente mediante un cable u otro dispositivo o medio, el medio portador puede estar

constituido por dicho cable u otro dispositivo o medio. Alternativamente, el medio portador puede ser un circuito integrado en el que esté encapsulado ("embedded") el programa de ordenador, estando adaptado dicho circuito integrado para realizar, o para usarse en la realización de aspectos de la invención.

5

Es también objeto de la presente invención el uso del método de predicción que constituye el primer aspecto de la invención sustituyendo en las redes tripartitas los nodos correspondientes a agentes, correspondientes a los pares de interacción agente-diana (e.g. fármaco-gen y fármaco-dominio), por cualquier otro tipo de compuesto, mecanismo o molécula, pre-existentes o nuevos, con conocido o predecible efecto inhibitorio o activador, de manera directa o indirecta, sobre dichas dianas, tales como: efecto inhibitorio o activador de la función de las proteínas o cualquiera de sus dominios; efecto inhibitorio o activador de la transcripción, del nivel de expresión o de la traducción a proteína de los genes diana o de sus dominios. Entre dichos compuestos, mecanismos o moléculas se encontrarían a modo de ejemplo, pero no de manera excluyente a otros ejemplos similares, las moléculas de ARN de cualquier tamaño o función, factores con efecto en la transcripción, nanopartículas, proteínas o sus dominios estructurales y/o funcionales.

10

15

Es también objeto de la presente invención el uso del método de predicción que constituye el primer aspecto de la invención para predecir los efectos de un agente o cualquier otro compuesto o mecanismo, así como, su uso para ayudar a identificar potenciales reposicionamientos para nuevas terapias de agentes, fármacos, compuestos o mecanismos preexistentes o versiones modificadas de los mismos; o para la identificación de nuevas terapias basadas en la identificación del efecto de nuevos agentes, fármacos, compuestos o mecanismos con conocida o con predecible acción sobre las dianas.

20

25

Es también objeto de la presente invención el uso del método de predicción que constituye el primer aspecto de la invención para la identificación de los efectos producidos por la administración simultánea o consecutiva de dos o más agentes (e.g. fármacos y/o compuestos y/o moléculas y/o mecanismos), tanto de efectos adversos, secundarios, o de efectos cruzados producidos entre estos agentes (fármacos y/o compuestos, etc.). Así como, también de efectos deseados ayudando en el diseño de nuevas combinaciones de agentes (e.g. fármacos, compuestos, moléculas y/o mecanismos con fines terapéuticos).

30

Es también objeto de la presente invención un método de identificación de los efectos fenotípicos de diferentes agentes con potencial terapéutico (e.g. fármacos; pequeños compuestos y moléculas; miRNA; etc.) a través de su interacción con dianas biológicas (e.g. proteínas, genes, y dominios funcionales en proteínas y genes) relacionadas con entidades biológicas y/o patológicas

35

(e.g. entidades como son las enfermedades, síndromes, pacientes, líneas celulares, organoides, etc.) que pueden asociarse a su vez a uno o varios fenotipos, con la finalidad de identificar asociaciones significativas entre dichos agentes y los fenotipos, tanto deseados, secundarios como adversos que comprende, además del método de predicción que constituye el primer
 5 aspecto de la invención, una etapa de validación que comprende la evaluación del solapamiento entre las asociaciones fenotipo–agente obtenidas por el método de predicción y las ya conocidas de los agentes en las bases de datos y/o en la bibliografía científica.

En resumen, la metodología objeto de la presente invención es de aplicación en el ámbito de la
 10 I+D farmacéutica, ayudando en la comprensión de los mecanismos que vinculan los medicamentos tanto con sus efectos previstos como con los ESA y contribuyendo a solventar los problemas indicados en el estado de la técnica (desarrollar métodos que identifiquen tempranamente los efectos adversos de los fármacos y sus mecanismos moleculares, dado que la mayoría fracasan en ensayos clínicos con altos costes asociados, abaratando de paso los
 15 costes de producción y desarrollo de la implementación de nuevos fármacos, y potenciando el reposicionamiento de fármacos como alternativa terapéutica y comercial viable) mediante el modelado y explotación de extensas redes de asociación de agentes-dianas-fenotipos. A partir de estas redes, se evalúa la significancia estadística de las coincidencias entre dianas (por ejemplo, genes) y fenotipos en las mismas enfermedades, lo que permite identificar las asociaciones
 20 relevantes entre dichas dianas (genes) y fenotipos; a la par que se obtienen predicciones de relaciones agentes-fenotipos, cada una con un índice matemático que pondera su significancia en términos de co-dependencia. Así, por ejemplo, dichas asociaciones diana (gen)-fenotipo, ponderadas mediante un índice matemático como el índice hipergeométrico, H_{yl} (*Fuxman et al. 2013*), que mide si el número de interacciones compartidas entre una diana (gen) y un fenotipo es
 25 significativamente mayor de lo esperado por azar, indicando así el grado de similitud entre sus perfiles de interacción. Estas asociaciones se integran, en un segundo paso, con datos de interacción agente (por ejemplo fármaco)-diana, lo que permite identificar miles de asociaciones estadísticamente significativas entre agentes (fármacos) y fenotipos a través de dianas compartidas.

30 A lo largo de la descripción y de las reivindicaciones, la palabra “comprende” y sus variantes no pretenden excluir otras características técnicas, aditivos, componentes o pasos. Para los expertos en la materia, otros objetos, ventajas y características de la invención se desprenderán en parte de la invención y en parte de la práctica de la invención. Los siguientes ejemplos y dibujos se
 35 proporcionan a modo de ilustración y no se pretende que restrinjan la presente invención. Además, la invención cubre todas las posibles combinaciones de realizaciones particulares y preferidas aquí indicadas.

Breve descripción de los dibujos

A continuación, se pasa a describir de manera muy breve una serie de dibujos que ayudan a comprender mejor la invención y que se relacionan expresamente con una realización de dicha invención, que se ilustra como un ejemplo no limitativo de ésta.

FIG.1 Muestra el flujo de trabajo de acuerdo con una realización preferente del método objeto de la invención dirigida a la identificación de los efectos fenotípicos de un determinado fármaco.

FIG.2 Muestra unas gráficas con el número interacciones pareadas en las listas de parejas confirmadas procedentes de la base de datos SIDER en comparación con las listas de superposición aleatoria. A la izquierda se muestran los pares para el método proteína-fenotipo y a la derecha los pares para el método dominio-fenotipo.

FIG. 3 Muestra unas gráficas con el número de interacciones en las listas de los pares confirmados, obtenidas de la literatura biomédica, en comparación con las listas de superposición aleatoria. A la izquierda se muestran los pares para el método proteína-fenotipo y a la derecha los pares para el método dominio-fenotipo.

Ejemplos de realización de la invención

A continuación se procede a describir un ejemplo de realización del método objeto de la presente invención dirigido a establecer la asociación de fármacos con posibles fenotipos incluyendo ESA, descritos en términos de fenotipos patológicos de la HPO. Para ello se construyeron y computaron asociaciones significativas en redes de relaciones entre fármacos, dianas, enfermedades y fenotipos. La significación de la asociación se cuantificó mediante *NetAnalyzer* (Rojano et al. 2017, Jabato et al. 2021), un software desarrollado por inventores de la presente invención y que implementa el índice Hyl.

El índice Hyl, que ha demostrado superar a otras métricas a la hora de identificar asociaciones entre genes y fenotipos en estudios desarrollados por inventores de la presente invención (Bueno et al. 2018; Rojano et al. 2017), es la probabilidad transformada logarítmicamente de tener una superposición de interacción igual o mayor que la observada entre A y B:

$$H_{AB} = -\log \sum_{i=|N(A) \cap N(B)|}^{\min(|N(A)|, |N(B)|)} \frac{\binom{|N(A)|}{i} \cdot \binom{n_y - |N(A)|}{|N(B)| - i}}{\binom{n_y}{|N(B)|}}$$

Dependiendo del contexto, “A” puede ser una droga y “B” un efecto fenotípico, conectadas por genes/proteínas o dominios, según la red analizada. También “A” puede ser un gen/proteína o dominio y “B” un fenotipo, conectados por enfermedades.

5

En la presente invención se han implementado dos aproximaciones para identificar asociaciones significativas entre fármacos y fenotipos, una basada en el uso de asociaciones entre los fármacos y sus genes diana (genes que codifican para las proteínas que interaccionan con los fármacos) y otra usando las relaciones entre los fármacos y los dominios de las proteínas diana, tal y como se muestra en la Figura 1. De este modo, el método de predicción resultante comprende los siguientes pasos: A) Obtención de las redes tripartitas fenotipo-enfermedad-gen; B) Obtención de los pares fármaco-gen; C) Obtención de los pares gen-dominio; D) Obtención de los pares gen-fenotipo; E) Obtención de los pares dominio-fenotipo; F) Obtención de los pares fármaco-dominio; G) Obtención de los pares fármaco-fenotipo. A continuación, se describe con mayor detalle cada uno de los pasos.

15

Módulo A. Obtención de las redes tripartitas fenotipo-enfermedad-gen:

Los datos de términos fenotípicos HPO y genes, ambos asociados a enfermedades, se obtuvieron de las bases de datos OMIM (enfermedades mendelianas) y Orphanet (enfermedades raras). En base a dichos datos se construyeron dos redes tripartitas independientes fenotipo-enfermedad-gen para cada una de estas bases de datos (OMIM y Orphanet).

20

Módulo B. Obtención de los pares fármaco-gen:

25

Los pares fármaco-proteína se obtuvieron de la base de datos ChEMBL (versión 29) utilizando los siguientes criterios de selección, según el método desarrollado por inventores de la presente invención y descrito en *Moya-García et al., 2017*:

30

- Pequeña molécula con aplicación terapéutica (Clasificación –“Flag”- terapéutica= 1).
- Interacción de unión directa con una sola proteína (Tipo de ensayo = B; Tipo de relación = D; Tipo de objetivo = Una sola proteína).
- Filtrar las actividades débiles (valor pchembl ≥ 6).

35

Se consideraron fármacos en todas las fases de desarrollo. Para cada par fármaco-proteína se identificó el gen codificante de la proteína diana obteniéndose así los pares fármaco-gen diana.

Módulo C. Obtención de los pares gen-dominio:

Las anotaciones de dominios de proteínas se obtuvieron del repositorio denominado FunFams perteneciente a la base de datos CATH (versión v4.3.0). CATH es una base de datos que identifica y clasifica dominios a partir de las estructuras 3D de proteínas conocidas. FunFams es una recopilación de grupos de secuencias de dominios estructurales de proteínas, filogenética y funcionalmente relacionadas. Utilizando el repositorio FunFam se pudieron obtener las relaciones de anotación proteínas-dominios FunFams, y por extensión las anotaciones de dichos dominios con los genes codificantes de las proteínas, obteniéndose así los pares gen-dominio. Cada proteína de la lista de pares fármaco-proteína puede contener uno o más FunFams, y los dominios del mismo FunFam pueden encontrarse en múltiples proteínas

Módulo D. Obtención de pares gen-fenotipo:

Las redes tripartitas fenotipo-enfermedad-gen (Módulo A) fueron utilizadas para calcular los índices de asociación Hyl para todos los pares gen-fenotipo compartiendo anotación en una o más enfermedades. A partir de esta lista de pares se seleccionaron aquellos con un $Hyl \geq 2.0$, correspondientes a un p-valor de 0,01 de rechazar la hipótesis nula (distribución aleatoria hipergeométrica) siendo verdadera. De esta manera, se identificaron los pares de asociación gen-fenotipo estadísticamente significativos siguiendo los métodos desarrollados por inventores de la presente invención (*Bueno et al. 2018; Rojano et al. 2017*). El subconjunto de asociaciones significativas gen-fenotipo (con valores $Hyl \geq 2.0$; p-valor ≤ 0.01) correspondiente a genes que codifican para dianas (proteínas y dominios de proteínas) de los fármacos registrados en ChEMBL fue utilizado para modelar las redes de asociaciones gen-fenotipo.

Módulo E. Obtención de pares dominio-fenotipo:

Para calcular las asociaciones dominio-fenotipo, en primer lugar, cada gen de la red gen-fenotipo (Módulo D) se emparejó con todas las familias de dominios funcionales de CATH, denominadas FunFams, presentes en la proteína codificada por dicho gen, obteniéndose así una red de pares gen-dominios FunFams. Esta red gen-dominio es combinada con la red gen-fenotipo para construir una red heterogénea tripartita fenotipo-gen-dominio, utilizada para el cómputo de los índices de asociación Hyl entre pares de dominios y fenotipos según ya se ha descrito. Siguiendo la misma metodología, los pares dominio-fenotipo más significativos ($Hyl \geq 2.0$; p-valor ≤ 0.01) fueron seleccionados para los cálculos posteriores.

Módulo F. Obtención de pares fármaco–dominio:

Sobre la base de los pares fármaco-gen (Módulo B) y gen-dominio (genes que presentan el

dominio FunFams; Módulo C), se asociaron los fármacos a los dominios FunFams. Para cada dominio FunFams se evaluó si estaba significativamente presente entre los genes diana de un determinado fármaco. Si era así, se podía asumir que el fármaco se une a sus dianas a través del dominio FunFams que está presente en ellas y que contiene el sitio de unión del fármaco. Se determinó si el dominio FunFams estaba significativamente sobrerrepresentado entre las dianas de un determinado fármaco mediante un test estadístico de sobrerrepresentación según se describe en *Moya-García et al., 2017*. Los pares fármaco-dominio FunFams así determinados se utilizaron para crear la red fármaco-dominio.

Módulo G. Obtención de pares fármaco-fenotipo.

En este módulo se describe la obtención de los pares de asociación estadísticamente significativos entre los fármacos y los fenotipos. Para ello, se integraron los pares fármaco-gen (Módulo B) y los pares fármaco-dominio (Módulo F), con los pares gen-fenotipo (Módulo D) y dominio-fenotipo (Módulo E) respectivamente, tanto para la base de datos de enfermedades OMIM como Orphanet. De esta forma se construyen cuatro redes tripartitas, conteniendo las asociaciones fármaco-gen-fenotipo y fármaco-dominio-fenotipo para OMIM y Orphanet. Cada una de estas cuatro redes tripartitas fueron utilizadas para calcular el Hyl para todos los pares fármaco-fenotipo según se ha descrito anteriormente. Los pares con $Hyl \geq 2,0$ ($p\text{-valor} \leq 0,01$) fueron seleccionados para evaluar el rendimiento del método identificando relaciones fármaco-fenotipo verdaderas en comparación con un predictor aleatorio.

El método de predicción descrito anteriormente, comprendiendo los módulos A a G, puede complementarse con un paso (H) de validación funcional (por ejemplo mediante un sistema HCS “High-Content Screening”), cuya implementación convierte el método de predicción objeto de la invención en un método de identificación, también comprendido en la invención que nos ocupa.

Evaluación del rendimiento en la identificación de asociaciones verdaderas fármaco-fenotipo

Gracias a la invención, y en base a la información depositada actualmente en las distintas bases de datos, fue posible identificar asociaciones estadísticamente significativas para más de 140.000 fármacos de ChEMBL (base de datos de fármacos y moléculas bioactivas con propiedades similares a los fármacos, que reúne datos químicos, bioactivos y genómicos para ayudar a traducir la información genómica en nuevos fármacos eficaces) con más de 3.800 fenotipos, representados como términos de la Ontología de Fenotipos Humanos (HPO, del inglés Human Phenotype Ontology).

Los resultados obtenidos con la metodología objeto de la presente invención fueron comparados con un predictor aleatorio y validados utilizando dos conjuntos de datos de referencia de asociaciones verdaderas fármaco-fenotipo (o conjuntos “gold standard”), como son la base de datos SIDER (SIDER recopila los efectos secundarios conocidos para fármacos actualmente
 5 aprobados y comercializados) y los datos de asociación fármaco-fenotipo obtenidos a partir del minado informático de los textos de la bibliografía biomédica mundial depositada en PubMed (>35 Millones publicaciones en febrero 2023); observándose un enriquecimiento en relaciones fármaco-fenotipo verdaderas, para asociaciones con $Hyl \geq 3,5$, de hasta 10 veces superior al predictor o modelo aleatorio cuando fue validada con SIDER, y de hasta 17,5 veces superior al predictor o
 10 modelo aleatorio al ser validada con los datos derivados de la bibliografía.

Analizando en detalle los resultados con mayor valor del Hyl (estadísticamente más significativos) se observaron asociaciones fármaco-fenotipo ya conocidas o descritas, en particular para cánceres como la leucemia mieloide crónica, que se vinculó con el fármaco Nilotinib. También se
 15 observaron altos valores del índice hipergeométrico de asociación entre el fenotipo “visión borrosa” y el fármaco Tetracaína. Por tanto, el método de la invención representa un importante avance en la identificación de relaciones tanto nuevas como conocidas de fármacos y fenotipos que incluyen los terapéuticos o deseados y los ESA, así como, en la identificación de los potenciales mecanismos moleculares (dianas) que, al ser accionados por el fármaco, producen dichos efectos.

Validación con la base de datos SIDER.

Para evaluar el rendimiento del método objeto de la presente invención en la identificación de asociaciones verdaderas fármaco-fenotipo, se tomó la lista de relaciones conocidas fármacos-
 25 ESAs de la base de datos SIDER (versión 4.1). SIDER se considera una base de datos de referencia -gold standard- de asociaciones verdaderas entre los ESAs y los fármacos disponibles en el mercado; sin embargo, esta base de datos no anota sus fármacos con los identificadores (ID) de la base de datos ChEMBL, ni tiene asociados sus términos ESA a los términos de fenotipo patológico de HPO, sino que utiliza términos UMLS (del inglés “Unified Medical Language System”) en su lugar. Por lo tanto, para poder comparar los pares de asociación en SIDER con los pares
 30 fármaco-fenotipo obtenidos por el método objeto de la presente invención y anotados con HPO y ChEMBL, se realizaron los siguientes pasos: En primer lugar, se utilizó la herramienta *Biothings Client* (versión 0.2.6 <https://pypi.org/project/biothings-client>) para consultar SIDER directamente a través de la base de datos MyChem.Info, lo que permitió conectar los IDs de ChEMBL
 35 directamente con sus ESAs asociados en la base de datos SIDER. En segundo lugar, los términos UMLS que definen los ESAs en SIDER fueron mapeados a fenotipos HPO utilizando la herramienta de mapeo ontológico OXO, a través de la API OxO REST (<https://www.ebi.ac.uk/spot/oxo/>). Estos pasos condujeron a la construcción de una lista derivada

de SIDER de pares fármaco-fenotipo, anotados con los identificadores de ChEMBL y HPO, que podían compararse directamente con las asociaciones fármaco-fenotipo obtenidas mediante el método objeto de la presente invención.

- 5 Conforme a lo anterior, se comparó el solapamiento de las predicciones del método con los datos de SIDER con el solapamiento obtenido por un predictor (modelo) aleatorio. Para ello, se realizaron las asociaciones fenotipo-fármaco de forma aleatoria, manteniendo igual el número de pares y la frecuencia de los identificadores de fármacos y fenotipos en dicha lista de pares obtenidos. Se estimó el rendimiento del método (“real”) con 100 repeticiones de método de
10 aleatorización (“aleatorio”) y comparando el solapamiento de estas listas, real y aleatorias, con los datos de SIDER mediante un índice de relación entre el valor de solapamiento “real” y “aleatorio”, donde “real” se refiere al número de pares fenotipo-fármaco asociados en nuestra lista que también se encuentran en SIDER y “aleatorio” se refiere al solapamiento medio entre las listas de pares aleatorios y SIDER. Este enfoque sigue el método de aleatorización basado en enlaces
15 descrito en trabajos anteriores desarrollados por inventores de la presente invención (*Díaz-Santiago et al. 2020*).

Pares gen-fenotipo						
OMIM				Orphanet		
	Pares	Fenotipos	Genes	Pares	Fenotipos	Genes
Total	103030	7304	4485	134021	6670	3173
Hyl ≥ 2	41224	7279	4351	48928	6646	3153
Hyl ≥ 3	9674	5662	2967	10512	4121	2294
Hyl $\geq 3,5$	3458	2480	1475	4132	1653	650
Pares dominio-fenotipo						
OMIM				Orphanet		
	Pares	Fenotipos	Dominios	Pares	Fenotipos	Dominios
Total	114772	6886	9286	132419	6419	7078
Hyl ≥ 2	112515	6886	9279	127034	6419	7009
Hyl ≥ 3	51394	6252	7807	55630	5870	5447
Hyl $\geq 3,5$	19390	4580	5429	17963	3966	3426

– Tabla 1 –

- 20 Tal y como se ha indicado, se utilizan dos flujos de trabajo para asociar fármacos con fenotipos patológicos. Uno a través de la asociación fármaco-gen (Módulo B) y otro basado en la asociación fármaco-dominio (Módulo F). De la base de datos SIDER se obtuvieron un total de 22.230 pares fármaco-fenotipo que fueron utilizados como conjunto de casos positivos para la validación.

En la Tabla 1, parte superior, se muestra el número total de pares gen-fenotipo (Módulo D), así como, el número de genes y fenotipos diferentes relacionados en el conjunto de estos pares, obtenidos a diferentes umbrales de Hyl con la red OMIM y Orphanet. Mientras la parte inferior de la Tabla 1 muestra los mismos datos, pero para los pares dominio-fenotipo (Módulo E).

En la Tabla 2, parte superior, se muestra el número total de pares fármaco-fenotipo, así como, el número de fármacos y fenotipos diferentes relacionados en el conjunto de estos pares, obtenidos a diferentes umbrales de Hyl con la red OMIM y Orphanet y la aproximación fármaco-gen (Módulo B); mientras la parte inferior de la Tabla 2 muestra los mismos datos, pero con la aproximación fármaco-dominio (Módulo F).

Aproximación fármaco-gen						
OMIM				Orphanet		
	Pares	Fenotipos	Fármacos	Pares	Fenotipos	Fármacos
Total	2012260	3666	141418	2260499	3830	112693
Hyl ≥ 2	1827126	3597	140615	1956625	3711	111809
Hyl ≥ 3	458240	1828	95851	377189	1509	76897
Hyl $\geq 3,5$	111608	728	39397	99566	691	31625
Aproximación fármaco-dominio						
OMIM				Orphanet		
	Pares	Fenotipos	Fármacos	Pares	Fenotipos	Fármacos
Total	3283330	4119	159282	3683928	4257	143879
Hyl ≥ 2	2943821	4039	159036	3383424	4195	143183
Hyl ≥ 3	2395901	3774	143523	2727330	3804	123787
Hyl $\geq 3,5$	2290236	3599	122882	2564173	3618	117703

– Tabla 2 –

Como era de esperar, el número de pares disminuyen con umbrales de asociación más restrictivos. Este patrón es consistente tanto para OMIM como para Orphanet. También se puede observar que el número de predicciones es bastante superior usando las aproximaciones con dominios que con genes para los mismos umbrales de valores Hyl, tanto en los resultados mostrados en la Tabla 1 como en la Tabla 2.

Para validar estos resultados, se identificaron todos los pares fármaco-fenotipo predichos por el método objeto de la presente invención que estuvieran también presentes en la base de datos SIDER y se compararon con el método aleatorio.

En la Tabla 3, en la columna “Método -M-”, se muestran el número de pares que se solapan con fármacos conocidos de la base de datos SIDER para diferentes umbrales de Hyl. La columna “Aleatorio -A-” muestra el número medio de pares aleatorios solapantes con pares presentes en SIDER. La columna Ratio (M/A) muestra la razón entre pares solapantes con SIDER encontrados por el método -M- dividido por los encontrados con el modelo aleatorio -A-. Estas columnas se repiten para los datos del método basado en las relaciones fármaco-gen (parte superior de la Tabla), el método basado en las relaciones fármaco-dominio (parte inferior); y para los pares calculados para la red de OMIM (parte izquierda) y Orphanet (parte derecha).

Aproximación fármaco-gen						
OMIM				Orphanet		
	Método -M- (pares solapantes)	Aleatorio -A- (pares solapantes)	Ratio (M/A)	Método -M- (pares solapantes)	Modelo Aleatorio -A- (pares solapantes)	Ratio (M/A)
Total de pares	280	181.39 ± 12.52	1.54	300	169.14 ± 11.16	1.77
Hyl ≥ 2	116	66.84 ± 6.84	1.74	70	49.56 ± 6.54	1.41
Hyl ≥ 3	15	3.20 ± 1.71	4.69	6	1.53 ± 1.23	3.92
Hyl ≥ 3,5	11	1.10 ± 1.04	10	2	0.36 ± 0.61	5.56
Aproximación fármaco-dominio						
OMIM				Orphanet		
	Método -M- (pares solapantes)	Aleatorio -A- (pares solapantes)	Ratio (M/A)	Método -M- (pares solapantes)	Modelo Aleatorio -A- (pares solapantes)	Ratio (M/A)
Total de pares	255	152.98 ± 12.46	1.67	200	145.07 ± 11.11	1.38
Hyl ≥ 2	160	99.61 ± 9.12	1.61	145	103.44 ± 9.37	1.40
Hyl ≥ 3	108	68.60 ± 7.57	1.57	112	73.74 ± 8.15	1.52
Hyl ≥ 3,5	102	61.72 ± 8.21	1.65	98	65.93 ± 8.21	1.49

– Tabla 3 –

Para los resultados del método basado en las asociaciones fármaco-gen, hay un claro aumento en términos de rendimiento en el enriquecimiento de verdaderos positivos de pares fármaco-fenotipo solapantes con SIDER comparado con el método aleatorio, correlacionando

positivamente con el aumento de los umbrales del Hyl. Esta tendencia aparece más marcada para OMIM que para Orphanet. Por ejemplo, el Ratio (M/A) se incrementa de 1,54 a 10 veces para OMIM, mientras que en Orphanet pasa de 1,77 a 5,56 para el mismo rango de umbrales Hyl. También se observa una clara diferencia en el mayor rendimiento del Ratio (M/A) para mayores umbrales Hyl del método basado en las asociaciones fármaco-gen que en el basado en las asociaciones fármaco-dominio, mostrando este último un enriquecimiento positivo muy bajo y sin variación apreciable a medida que se incrementa el umbral de Hyl (ver Tabla 3).

Validación con la base de datos de bibliografía científica.

Sin embargo, hay que tener en cuenta una importante advertencia al utilizar SIDER para evaluar los datos obtenidos por el método objeto de la presente invención: su contenido se limita únicamente a los fármacos comercializados. Como tal, sólo un pequeño porcentaje de fármacos de las listas de pares fármaco-fenotipo obtenidas por el método objeto de la presente invención solapan con los fármacos de la base de datos ChEMBL, puesto que la selección en ChEMBL empleada incluía “fármacos” y “compuestos” en todas las fases de desarrollo, además de los ya comercializados (Módulo B).

Para valorar el rendimiento del método objeto de la presente invención con todo el espectro de fármacos (comercializados y en desarrollo), como segunda validación, también se analizó el solapamiento entre las listas de asociaciones fenotipo-fármaco obtenidas mediante el método objeto de la presente invención y las listas de asociaciones fenotipo-fármaco obtenidas a partir de la co-ocurrencia significativa en la bibliografía científica universal depositada en PubMed. Para ello, se utilizó la metodología desarrollada en colaboración con Pazos *et al.*, 2022. En resumen, la lista de entradas de PubMed que mencionan un determinado fármaco se recupera consultando la *API Entrez* del NCBI en busca de artículos que mencionen el nombre de cada fármaco o cualquiera de sus sinónimos ChEMBL en cualquier campo. Lo mismo se hace para cada nombre de término HPO y sus sinónimos. A continuación, el conjunto de entradas de PubMed que mencionan un término HPO junto con un fármaco determinado se infiere como la intersección de los conjuntos que mencionan cada uno individualmente. Teniendo en cuenta el número de artículos que mencionan el fármaco, el número que menciona el HPO y el número que menciona ambos, así como el tamaño total de PubMed, se aplica una prueba estadística ($p < 0,05$) para seleccionar los pares fármaco-fenotipo significativamente co-citados en la bibliografía (Pazos *et al.* 2022).

Una vez obtenida esta lista de pares fármaco-fenotipo co-citados, se calculó el solapamiento entre estos pares y nuestras listas de predicciones fármaco-fenotipo generadas mediante el flujo de trabajo de análisis resumido en la figura 1. El rendimiento se calculó comparando el solapamiento del método objeto de la presente invención con 100 repeticiones del mismo modelo de

aleatorización descrito anteriormente para la validación con SIDER.

Del minado de la bibliografía científica se obtuvieron un total de 141.158 pares fármaco-fenotipo que fueron utilizados como conjunto de casos positivos para la validación. Los resultados del solapamiento de pares con el método de predicción y el modelo aleatorio se muestran en la Tabla 4 con la misma estructura y disposición que para los resultados mostrados en la Tabla 3 anterior.

	Aproximación fármaco-gen					
	OMIM			Orphanet		
	Método -M- (pares solapantes)	Aleatorio -A- (pares solapantes)	Ratio (M/A)	Método -M- (pares solapantes)	Aleatorio -A- (pares solapantes)	Ratio (M/A)
Total de pares	2111	570.52± 24.84	3.70	1674	570.93 ± 20.08	2.93
Hyl ≥ 2	1283	273.48 ± 17.55	4.69	919	250.07 ± 16.81	3.67
Hyl ≥ 3	215	17.12 ± 4.35	12.56	99	13.98 ± 3.84	7.08
Hyl ≥ 3,5	87	4.98 ± 2.08	17.47	41	3.87 ± 1.86	10.59
	Aproximación fármaco-dominio					
	OMIM			Orphanet		
	Método -M- (pares solapantes)	Aleatorio -A- (pares solapantes)	Ratio (M/A)	Método -M- (pares solapantes)	Aleatorio -A- (pares solapantes)	Ratio (M/A)
Total de pares	2127	532.94 ± 22.13	3.99	1784	576.63 ± 19.16	3.09
Hyl ≥ 2	1721	398.24 ± 17.78	4.32	1536	469.76 ± 22.57	3.27
Hyl ≥ 3	1342	290.34 ± 15.89	4.62	1178	344.55 ± 18.39	3.42
Hyl ≥ 3,5	1243	262.20 ± 16.49	4.74	1067	312.90 ± 16.31	3.41

– Tabla 4 –

Al igual que con la validación con SIDER mostrada anteriormente, se observó que para los resultados del método basado en las asociaciones fármaco-gen, hay un claro aumento en términos de rendimiento en el enriquecimiento de pares fármaco-fenotipo solapantes con la bibliografía comparado con el método aleatorio, correlacionando positivamente con el aumento de los

umbrales del Hyl. No obstante, los valores de enriquecimiento son superiores en todos los umbrales y categorías a los encontrados con la validación con SIDER. Nuevamente, la tendencia de enriquecimiento ascendente aparece más marcada para OMIM que para Orphanet. Por ejemplo, el Ratio (M/A) se incrementa de 3,70 a 17,47 veces para OMIM, mientras que en Orphanet pasa de 2,93 a 10,59 para el mismo rango de umbrales Hyl. Como en la validación con SIDER, en este caso también se observa una clara diferencia en el mayor rendimiento del Ratio (M/A) con el método basado en las asociaciones fármaco-gen que en el basado en las asociaciones fármaco-dominio. Aunque en este caso, el método basado en las asociaciones fármaco-dominio muestra un mayor enriquecimiento, en todos los umbrales Hyl, utilizando los casos de co-citación que el observado con la validación con SIDER (comparar Tabla 2 y Tabla 3).

Tanto el solapamiento con SIDER como con la bibliografía demuestran que nuestras aproximaciones encuentran más asociaciones fármaco-fenotipo conocidas o verdaderas que el azar, siendo la aproximación fármaco-gen la que muestra el mayor rendimiento asociado a los umbrales de Hyl. Por tanto, los resultados de validación aquí presentados demuestran la utilidad de la metodología aquí presentada para la predicción de relaciones fármaco-fenotipo a través de la interacción con sus dianas en redes heterogéneas de relaciones fármaco-diana-enfermedad-fenotipo.

Ejemplos de utilidad de las predicciones de asociación obtenidas con la aplicación del método.

El método objeto de la presente invención es capaz de predecir asociaciones fármaco-fenotipo muy por encima del azar dependiendo del umbral Hyl que apliquemos (ver Tablas 3 y 4). La naturaleza del método basado en múltiples redes de asociación heterogénea fenotipo-enfermedad-diana-fármaco, permite rescatar las relaciones dentro de dichas redes que han conducido a cada una de las predicciones fármaco-fenotipo, como, por ejemplo, las dianas moleculares (genes, proteínas, y dominios de proteínas) que median en el efecto fenotípico o en la relación de un fármaco con una enfermedad concreta. Para ilustrar este punto se presenta a continuación un estudio detallado de algunos de los pares obtenidos por el método en sus dos aproximaciones o versiones: fármaco-gen y fármaco-dominio. Las 20 principales asociaciones fármaco-fenotipo se muestran en las tablas 5 y 6.

La Tabla 5 muestra los principales pares fármaco-fenotipo utilizando OMIM, ordenados descendientemente según el Hyl y utilizando la metodología fármaco-gen, incluyendo únicamente los fármacos de ChEMBL que también se encuentran en SIDER (fármacos comercializados en Fase 4). De izquierda a derecha, ID del fenotipo en la base de datos HPO; nombre o descripción del fenotipo; ID del fármaco de la base de datos ChEMBL; nombre comercial del fármaco; fase 4

(fármacos comercializados); ID de los genes diana en la red de asociaciones.

HPO	Nombre HPO	Fármaco	Nombre fármaco	Fase	Hyl	Dianas
HP:0005506	Leucemia mieloide crónica	CHEMBL255863	NILOTINIB	4	4.72	BCR, ABL1
HP:0005506	Leucemia mieloide crónica	CHEMBL941	IMATINIB	4	4.65	BCR, ABL1
HP:0000958	Piel Seca	CHEMBL1229517	VEMURAFENIB	4	4.56	BRAF, RAF1
HP:0005506	Leucemia mieloide crónica	CHEMBL288441	BOSUTINIB	4	4.24	BCR, ABL1
HP:0005506	Leucemia mieloide crónica	CHEMBL1171837	PONATINIB	4	4.19	BCR, ABL1
HP:0000746	Delirios	CHEMBL589	ROPINIROLE	4	3.91	DRD3, HTR2A
HP:0100723	Tumor del estroma gastrointestinal	CHEMBL941	IMATINIB	4	3.88	KIT, PDGFRA
HP:0100753	Esquizofrenia	CHEMBL243712	AMISULPRIDE	4	3.84	DRD3, HTR2A
HP:0000746	Delirios	CHEMBL243712	AMISULPRIDE	4	3.69	DRD3, HTR2A
HP:0005506	Leucemia mieloide crónica	CHEMBL1421	DASATINIB	4	3.69	BCR, ABL1

– Tabla 5 –

- 5 La Tabla 6 muestra la misma información por columnas que la Tabla 7 utilizando la metodología fármaco-dominio. La última columna muestra el ID de los dominios FunFams presentes en las proteínas de los genes diana del fármaco en la red.

HPO	Nombre HPO	Fármaco	Nombre fármaco	Fase	Hyl	Dianas
HP:0005506	Leucemia mieloide crónica	CHEMBL255863	NILOTINIB	4	26.30	1.10.510.10-ff-70, 2.30.30.40-ff-10, 1.10.555.1

						0-ff-4, 1.20.900.1 0-ff-14, 2.60.40.15 0-ff-57, 4.10.280.3 0-ff-1, 3.30.200.2 0-ff-37, 1.20.120.3 30-ff-3, 2.30.29.30 -ff-112, 3.30.505.1 0-ff-4,
HP:0000 622	Visión Borrosa	CHEMBL698	TETRACAIN E	4	26.2 2	1.20.120.3 50-ff-2, 1.10.287.7 0-ff-3, 1.20.120.3 50-ff-3, 1.10.238.1 0-ff-2, 1.10.287.7 0-ff-6, 1.20.120.3 50-ff-4, 1.10.287.7 0-ff-4, 1.10.287.7 0-ff-1, 1.20.5.119 0-ff-1, 1.20.120.3 50-ff-5,
HP:0005 506	Leucemia mieloide crónica	CHEMBL941	IMATINIB	4	24.9 7	1.10.510.1 0-ff-70, 2.30.30.40 -ff-10,

						1.10.555.1 0-ff-4, 1.20.900.1 0-ff-14, 2.60.40.15 0-ff-57, 4.10.280.3 0-ff-1, 3.30.200.2 0-ff-37, 1.20.120.3 30-ff-3, 2.30.29.30 -ff-112, 3.30.505.1 0-ff-4,
HP:0006 721	Leucemia linfocítica aguda	CHEMBL941	IMATINIB	4	18.9 3	2.60.40.10 -ff-661, 2.30.29.30 -ff-112, 1.10.555.1 0-ff-4, 1.20.900.1 0-ff-14, 1.10.510.1 0-ff-426, 4.10.280.3 0-ff-1, 2.60.40.15 0-ff-57, 2.60.40.10 -ff-1159, 3.30.200.2 0-ff-366, 3.30.200.2 0-ff-25,
HP:0006 721	Leucemia linfocítica aguda	CHEMBL1171 837	PONATINIB	4	17.5 4	2.60.40.10 -ff-661,

						2.30.29.30 -ff-112, 1.10.555.1 0-ff-4, 1.20.900.1 0-ff-14, 1.10.510.1 0-ff-426, 4.10.280.3 0-ff-1, 2.60.40.15 0-ff-57, 2.60.40.10 -ff-1159, 3.30.200.2 0-ff-366, 3.30.200.2 0-ff-25
HP:0001 269	Hemiparesia	CHEMBL1908 360	EVEROLIMU S	4	17.1 0	1.25.10.10 -ff-94, 1.25.10.10 -ff-60, 3.30.1010. 10-ff-4, 1.25.10.10 -ff-83, 1.10.1070. 11-ff-7, 1.25.10.10 -ff-140, 1.20.120.1 50-ff-1
HP:0004 936	Trombosis venosa	CHEMBL1171 837	PONATINIB	4	16.4 1	2.60.40.10 -ff-832, 2.60.40.10 -ff-725, 1.10.510.1 0-ff-140, 2.60.40.10

						-ff-720, 2.60.40.10 -ff-776, 2.60.40.10 -ff-223, 3.30.200.2 0-ff-25
HP:0100 723	Tumor del estroma gastrointestinal	CHEMBL941	IMATINIB	4	15.2 4	2.60.40.10 -ff-725, 1.10.510.1 0-ff-140, 2.60.40.10 -ff-720, 1.10.510.1 0-ff-177, 2.60.40.10 -ff-776, 2.60.40.10 -ff-223, 2.60.40.10 -ff-832, 3.30.200.2 0-ff-25
HP:0100 723	Tumor del estroma gastrointestinal	CHEMBL4777 72	PAZOPANIB	4	13.4 0	2.60.40.10 -ff-725, 1.10.510.1 0-ff-140, 2.60.40.10 -ff-720, 1.10.510.1 0-ff-177, 2.60.40.10 -ff-776, 2.60.40.10 -ff-223, 2.60.40.10 -ff-832, 3.30.200.2 0-ff-25

HP:0002 170	Hemorragia intracraneal	CHEMBL2317 79	APIXABAN	4	12.7 1	2.10.25.10 -ff-443, 2.40.10.10 -ff-13, 2.10.25.10 -ff-162, 2.40.10.10 -ff-11, 4.10.740.1 0-ff-1
----------------	-------------------------	------------------	----------	---	-----------	--

– Tabla 6 –

Para el método basado en fármaco-gen aplicado a la base de datos OMIM (Tabla 7), el resultado con mayor Hyl es para el fármaco Nilotinib, asociado con el fenotipo Leucemia mielógena crónica.

- 5 El Nilotinib es un medicamento utilizado para tratar la leucemia mielógena crónica. Evidentemente, no se trata de un ESA, sino de la patología que se pretende tratar con el fármaco, o también denominado fenotipo deseado o terapéutico. El fenotipo y el fármaco están vinculados a través de las proteínas codificadas por los genes *ABL1* y *BCR*. El siguiente fenotipo de la lista es *Piel seca*, asociado al fármaco Vemurafenib, utilizado para tratar el melanoma. En efecto, la piel seca
- 10 aparece como uno de los efectos secundarios; curiosamente, la asociación entre este fenotipo y el fármaco está mediada por las proteínas codificadas por los genes *RAF1* y *BRAF*, ambos relacionados con la dermatitis atópica.

- Para la metodología fármaco-dominio aplicada a OMIM (Tabla 8), el resultado principal es el mismo
- 15 que para el método fármaco-gen (Tabla 7): la predicción del par de asociación entre la Leucemia mielógena crónica y Nilotinib. En el caso de la aproximación fármaco-dominio, el fármaco y el fenotipo están conectados por múltiples dominios proteicos pertenecientes a FunFams relacionados con la actividad de la quinasa o con "breakpoint cluster region". El segundo par, la Visión borrosa asociada a la Tetracaína, no aparece entre los 20 primeros del ranking fármaco-
- 20 gen (Tabla 7). Este par fármaco-fenotipo está asociado a través de dominios de proteínas FunFams relacionados con los canales de sodio.

- Al combinar la información sobre el fármaco y el fenotipo de diferentes fuentes, se genera una lista exhaustiva de los posibles efectos del fármaco, lo que permite identificar los efectos conocidos y
- 25 predecir nuevos efectos putativos de los fármacos, incluyendo tanto los ESA como los efectos deseados o terapéuticos (es decir, el fenotipo patológico para cuyo tratamiento se diseñó el fármaco). Además de estos efectos putativos, nuestro enfoque proporciona información sobre los genes y dominios proteicos responsables (dianas), ofreciendo pistas sobre los mecanismos

moleculares subyacentes.

Estos resultados refuerzan la premisa en la que se basa este estudio: los fenotipos patológicos asociados a las variantes de cada proteína también pueden producirse cuando la proteína funcional está accionada por la unión del fármaco. En el caso de las mutaciones de pérdida de función y de los fármacos que inactivan o inhiben las proteínas, este principio se mantiene. Sin embargo, en otros casos, como el de muchos cánceres, las variantes pueden dar lugar a una ganancia de función. Además, aunque la mayoría de los fármacos se consideran inhibidores, muchos son potenciadores o activadores. Como tal, al combinar los pares procedentes de las aproximaciones fármaco-gen y fenotipo-dominio, también encontramos vínculos entre los fármacos y los fenotipos que son causados por la activación de la proteína, en lugar de su inhibición. El ejemplo más claro de esto lo proporcionan los resultados superiores utilizando tanto los métodos de fármaco-gen como de fármaco-dominio, donde mostramos una conexión entre la Leucemia mieloide crónica y el Nilotinib, que en realidad están vinculados por oncogenes potenciales, lo que encaja con la idea de que es la activación de la proteína la que conduce al fenotipo y, por lo tanto, su inhibición conduce al efecto terapéutico previsto contra este tipo de cáncer.

La metodología, al asociar inicialmente los fármacos con las dianas, y las dianas con los fenotipos, pretende modelar molecularmente cómo los fármacos conducen a sus efectos. Esto tiene aplicaciones potenciales para diseñar validaciones experimentales específicas en los ensayos clínicos, que eventualmente permitan excluir fármacos con potenciales efectos secundarios en etapas tempranas del proceso de diseño, con el consiguiente ahorro de recursos que esto conllevaría. A la inversa, cuando este efecto es la patología para la que estaba destinado el fármaco, la metodología puede proporcionar al usuario información adicional sobre los mecanismos (proteínas/dominios) a través de los cuales actúa su fármaco. Nuestros resultados también tienen implicaciones para la reutilización o reposicionamiento de fármacos, ya que algunos de los efectos predichos podrían ser de interés terapéutico. Igualmente, se pueden extender las redes de asociación con información de dianas adicionales a las contempladas en esta solicitud como, por ejemplo, regiones reguladoras, miRNAs, y otras moléculas y mecanismos accionables por fármacos y otros compuestos que conlleven la acción directa o indirecta sobre la expresión de un gen con efectos fenotípicos asociados.

Aunque las aproximaciones fármaco-gen y fármaco-dominio mostraron cierta similitud en algunos de sus resultados, la mayoría de las veinte principales asociaciones fármaco-fenotipo fueron diferentes. Esto es de esperar: una proteína puede estar anotada con múltiples dominios, y un determinado dominio puede pertenecer a múltiples proteínas influyendo en los resultados obtenidos por cada aproximación, incrementando el número de predicciones finales al combinar

los resultados de ambos métodos.

También es evidente que los resultados de OMIM son sistemáticamente mejores que los de Orphanet. Esto se debe probablemente a que OMIM contiene enfermedades mendelianas, en las que cada enfermedad se considera en gran medida una entidad independiente. Como tal, el vínculo entre la variante, el gen y el fenotipo es posiblemente más directo, mientras que en el caso de Orphanet es más probable que una determinada enfermedad esté vinculada a múltiples genes. De hecho, el número medio de genes por enfermedad en Orphanet es casi el doble que en OMIM (2,19 frente a 1,12). Tanto el incremento de la información existente en las bases de datos utilizadas en el ejemplo de aplicación de este método, como la aparición de nuevas fuentes de información o clasificación de enfermedad, genes, relaciones fenotípicas, moleculares, de interacción con fármacos, o de ontologías biomédicas, pueden todas ser integradas en la red de asociaciones del método, y eventualmente, mejorar su rendimiento.

Aunque existen otras métricas además del índice Hyl, en estudios anteriores se ha demostrado que proporciona los mejores resultados cuando se aplica a datos similares (*Bueno et al. 2018, Rojano et al. 2017*). Es posible que otras métricas de asociación tengan un buen rendimiento para diferentes tipos de redes y esto sigue siendo una posible vía de mejora del método.

Validación funcional de las interacciones significativas fármaco-fenotipo-diana mediante un sistema de HCS.

Las interacciones fármaco-fenotipo-diana encontradas en las diferentes redes y que presentan un grado de interacción significativa podrá ser validadas funcionalmente utilizando para ello un método de adquisición y procesado de imágenes de alta capacidad, HCS (del inglés “High-Content-Screening”). En concreto se empleó el sistema de alto cribado Operetta (PerkinElmer, Inc.), cuyo análisis de imágenes será realizado empleando el software Harmony, versión 4.8 (PerkinElmer, Inc.). Este sistema automatizado permite la utilización de diferentes modelos celulares, desde cultivos de líneas celulares (bien sean líneas inmortalizadas o cultivos primarios) hasta la utilización de modelos más complejos como pueden ser modelos de organoides. La evaluación de la interacción entre el fármaco utilizado y la diana sobre la que actúa vendrá determinada por el empleo de una tinción vital (del inglés “dye”), que ponga de manifiesto el fenotipo que se quiere caracterizar. Los modelos celulares son incubados en presencia de estas tinciones vitales, que ponen de manifiesto diferentes estructuras subcelulares, pudiendo ser monitoreadas a lo largo del tiempo en condiciones ambientales fisiológicas (temperatura y niveles de CO₂) a lo largo del tiempo. Para abordar la reducción o la eliminación de una diana (gen) en concreto, es posible transfectar el modelo celular con un siRNA específico, poniendo de manifiesto el peso que tiene dicha diana en el fenotipo que se quiera evaluar. Este tipo de abordaje, en el

que se utilizan pequeños volúmenes, permite la validación de dianas a gran escala de multitud de dianas; abaratando los tiempos de validación y los costes empleados en dichos procesos.

Referencias

- 5
- Bueno, A., Rodríguez-López, R., Reyes-Palomares, A., Rojano, E., Corpas, M., Nevado, J., Lapunzina, P., Sánchez-Jiménez, F., Ranea, J. A. G. (2018). Phenotype-loci associations in networks of patients with rare disorders: application to assist in the diagnosis of novel clinical cases. *European Journal of Human Genetics*, 26(10), 1451-1461. doi:10.1038/s41431-018-0139-x
- 10 Díaz-Santiago, E., Jabato, F. M., Rojano, E., Seoane, P., Pazos, F., Perkins, J. R., & Ranea, J. A. G. (2020). Phenotype-genotype comorbidity analysis of patients with rare disorders provides insight into their pathological and molecular bases. *PLOS Genetics*. doi:10.1371/journal.pgen.1009054
- Fuxman Bass JI, Diallo A, Nelson J, Soto JM, Myers CL, Walhout AJ. Using networks to measure similarity between genes: association index selection. *Nat Methods*. 2013 Dec;10(12):1169-15 76. doi: 10.1038/nmeth.2728.
- Jabato, F. M., Seoane, P., Perkins, J. R., Rojano, E., García Moreno, A., Chagoyen, M., Pazos, F., Ranea, J. A. G. (2021). Systematic identification of genetic systems associated with phenotypes in patients with rare genomic copy number variations. *Human genetics*, 140(3), 457-475. doi:10.1007/S00439-020-02214-7
- 20 Moya-García, A., Adeyelu, T., Kruger, F. A., Dawson, N. L., Lees, J. G., Overington, J. P., Orengo, C., Ranea, J. A. G. (2017). Structural and Functional View of Polypharmacology. *Scientific Reports*, 7(1), 1-14. doi:10.1038/s41598-017-10012-x
- Nguyen, P. A., Born, D. A., Deaton, A. M., Nioi, P., & Ward, L. D. (2019). Phenotypes associated with genes encoding drug targets are predictive of clinical trial side effects. *Nature communications*, 25 10(1). doi:10.1038/S41467-019-09407-3
- Pazos, F., Chagoyen, M., Seoane, P., & Ranea, J. A. G. (2022). CoMent: Relationships Between Biomedical Concepts Inferred From the Scientific Literature. *Journal of Molecular Biology*, 167568. doi:10.1016/j.jmb.2022.167568
- Rojano, E., Seoane, P., Bueno-Amoros, A., Perkins, J. R., & García-Ranea, J. A. (2017). 30 Revealing the relationship between human genome regions and pathological phenotypes through network analysis. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 197-207.

REIVINDICACIONES

1. Método implementado en computador para la predicción de los efectos fenotípicos de agentes moleculares con potencial terapéutico que comprende la obtención de pares de asociaciones
5 **agente-fenotipos** significativas a partir de
 1. al menos una red tripartita **fenotipo-dominio-agente** obtenida a partir de
 - a. pares de asociaciones **dominio-fenotipo** significativas obtenidas a partir de
 - i. al menos una red tripartita **fenotipo-diana-dominio** obtenida a partir de
10 pares de asociaciones **diana-dominio** y de pares de asociaciones **diana-fenotipo** significativas;
y de
 - b. pares de asociaciones **agente-dominio** significativas obtenidas a partir de pares de asociaciones **diana-dominio** y de pares de asociaciones **agente-diana**;
y de
 - 15 2. al menos una red tripartita **fenotipo-diana-agente** obtenida a partir de
 - a. pares de asociaciones **diana-fenotipo** significativas obtenidas a partir de al menos una red tripartita de relaciones **fenotipos-entidades-dianas**;
y de
 - b. pares de asociaciones **agente-diana**.
- 20 2. Método según la reivindicación anterior caracterizado por que comprende una etapa de obtención de pares de asociaciones **diana-dominio** a partir de una base de datos o repositorio de datos.
- 25 3. Método según la reivindicación 1 caracterizado por que comprende una etapa de obtención de pares de asociaciones **diana-fenotipo** a partir de al menos una red tripartita de relaciones **fenotipos-entidades-dianas** sobre la que se calcula un índice, o cualquier otra medida de asociación, para obtener asociaciones diana-fenotipo significativas;
- 30 4. Método según la reivindicación 1 caracterizado por que comprende una etapa de obtención de pares de asociaciones **agente-diana** a partir de una base de datos o repositorio de datos.
5. Método según la reivindicación 1 caracterizado por que comprende una etapa de obtención de pares de asociaciones **dominio-fenotipo** a partir de al menos una red tripartita **fenotipo-diana-dominio**
35 **dominio** obtenida a partir de pares de asociaciones **diana-dominio** y de pares de asociaciones **diana-fenotipo** significativas obtenidas a partir de al menos una red tripartita de relaciones **fenotipos-entidades-diana**.

6. Método según la reivindicación 1 caracterizado por que comprende una etapa de obtención de pares de asociaciones **agente-dominio** significativas a partir de pares de asociaciones **diana-dominio** y de pares de asociaciones **agente-diana**.

7. Método según cualquiera de las reivindicaciones anteriores caracterizado por que comprende:

- a) una etapa de obtención de pares de asociaciones **diana-dominio** a partir de una base de datos o repositorio de datos;
- b) una etapa de obtención de pares de asociaciones **diana-fenotipo** significativas a partir de al menos una red tripartita de relaciones fenotipos-entidades-dianas sobre la que se calcula un índice, o cualquier otra medida de asociación, para obtener asociaciones diana-fenotipo significativas;
- c) una etapa de obtención de pares de asociaciones **agente-diana** a partir de una base de datos o repositorio de datos;
- d) una etapa de obtención de pares de asociaciones **dominio-fenotipo** significativas a partir de al menos una red tripartita fenotipo-diana-dominio obtenida a partir de los pares de asociaciones diana-dominio obtenidos en la etapa (a) y de los pares de asociaciones diana-fenotipo significativas obtenidas en la etapa (b), red tripartita sobre la que se calcula un índice, o cualquier otra medida de asociación, para obtener asociaciones dominio-fenotipo significativas;
- e) una etapa de obtención de pares de asociaciones **agente-dominio** significativas a partir los pares de asociaciones diana-dominio obtenidos en la etapa (a) y de los pares de asociaciones agente-diana obtenidos en la etapa (c) sobre los que se calcula un índice, o cualquier otra medida de asociación, para obtener asociaciones agente-dominio significativas; y
- f) una etapa de obtención de pares de asociaciones **agente-fenotipo** significativas a partir de
 - (i) al menos una red tripartita fenotipo-dominio-agente obtenida a partir de los pares de asociaciones dominio-fenotipo significativas obtenidas en la etapa (d) y de los pares de asociaciones agente-dominio significativas obtenidas en la etapa (e), y de
 - (ii) al menos una red tripartita fenotipo-diana-agente obtenida a partir de los pares de asociaciones diana-fenotipo significativas obtenidas en la etapa (b) y de los pares de asociaciones agente-diana obtenidas en la etapa (c);
 redes tripartitas (i) y (ii) sobre las que se calcula de manera independiente o combinada un índice, o cualquier otra medida de asociación, para obtener pares de asociaciones agente-fenotipos significativas.

8. Método según cualquiera de las reivindicaciones anteriores caracterizado por que el agente es una proteína y en la etapa (c) de obtención de pares de asociaciones agente-diana, cada agente

puede comprender uno o más dominios funcionales (por ejemplo dominios FunFams), dominios estructurales 3D, funcionales y/o de relación filogenética, y, a su vez, un mismo tipo de dominio puede estar presente en múltiples agentes, de tal forma que para cada dominio se evaluaría si está significativamente sobrerrepresentado entre las dianas de un determinado agente, utilizando métodos de cálculo o estimación de la significancia de la relación agente-dominio, como por ejemplo, la probabilidad esperada y la prueba binomial utilizadas, y si este es el caso, se asocia dicho agente a ese dominio, creando un par agente-dominio.

9. Método según cualquiera de las reivindicaciones anteriores caracterizado por que en la etapa (d) de obtención de pares agente-fenotipo, se utilizan diferentes umbrales del índice, o cualquier otra medida de asociación, para evaluar el rendimiento del método basado en el número de asociaciones agente-fenotipo verdaderas encontradas a diferentes umbrales, comparando con una aproximación aleatoria, y utilizando para dicha validación grupos control de asociaciones agente-fenotipo conocidas, tales como, aunque no excluyentes, grupos de datos procedentes de la base de datos de medicamentos en comercialización en las que se registran sus correspondientes ESA (por ejemplo, base de datos SIDER) y o de la bibliografía científica.

10. Método según cualquiera de las reivindicaciones anteriores caracterizado por que la evaluación del solapamiento entre las asociaciones agente-fenotipo obtenidas, y los efectos secundarios conocidos del agente comprende comparar los datos de todos los efectos (fenotipos) registrados en una base de datos con las asociaciones agente-fenotipo obtenidas mediante las aproximaciones basadas en dominios y dianas, implementadas de manera independiente o combinada.

11. Método según cualquiera de las reivindicaciones anteriores caracterizado por que comprende una etapa de validación que comprende la confrontación de otros predictores, más preferentemente con un predictor o modelo aleatorio, y/con los resultados obtenidos por un método conforme cualquiera de las reivindicaciones anteriores.

12. Método según cualquiera de las reivindicaciones anteriores caracterizado por que el índice de asociación utilizado es el índice Hyl, que es la probabilidad transformada logarítmicamente de tener una superposición de interacción igual o mayor que la observada entre A y B:

$$H_{AB} = -\log \sum_{i=|N(A) \cap N(B)|}^{\min(|N(A)|, |N(B)|)} \frac{\binom{|N(A)|}{i} \cdot \binom{n_y - |N(A)|}{|N(B)| - i}}{\binom{n_y}{|N(B)|}}$$

13. Método según la reivindicación anterior caracterizado por que en la etapa (b) de obtención de pares de asociaciones diana-fenotipo, la red tripartita de relaciones fenotipos-entidades-dianas se utilizan para calcular los índices Hyl para todos los pares diana-fenotipo compartiendo anotación en una o más entidades, seleccionándose aquellos pares de asociaciones diana-fenotipo con un índice Hyl ≥ 2.0 y un p-valor $\leq 0,01$, identificándose así los pares diana-fenotipo estadísticamente significativos utilizados para modelar las redes de asociaciones diana-fenotipo.

14. Método según cualquiera de las reivindicaciones anteriores 12 a 13 caracterizado por que en la etapa (d) de obtención de pares de asociaciones dominio-fenotipo, cada diana de los pares de asociaciones dominio-fenotipo se empareja con todas las familias de dominios funcionales presentes en la proteína codificada por dicha diana, seleccionándose aquellos pares de asociaciones dominio-fenotipo con un índice Hyl ≥ 2.0 y un p-valor $\leq 0,01$, identificándose así los pares dominio-fenotipo estadísticamente significativos utilizados para modelar las redes de asociaciones dominio-fenotipo.

15. Método según cualquiera de las reivindicaciones anteriores 12 a 14 caracterizado por que en la etapa (e) de obtención de pares de asociaciones agente-dominio, para cada dominio se evalúa si está significativamente presente en las dianas del agente mediante un test estadístico de sobrerrepresentación.

16. Método según cualquiera de las reivindicaciones anteriores 12 a 15 caracterizado por que en la etapa (d) de obtención de pares de asociaciones agente-fenotipo, para cada red tripartita (i) y (ii) calcular el índice Hyl para todos los pares agente-fenotipo, seleccionándose aquellos pares de asociaciones dominio-fenotipo con un índice Hyl ≥ 2.0 y un p-valor $\leq 0,01$.

17. Uso de un método según cualquiera de las reivindicaciones anteriores para la predicción de los efectos de un agente o cualquier otro compuesto o mecanismo, así como, su uso para ayudar a identificar potenciales reposicionamientos para nuevas terapias de agentes, fármacos, compuestos o mecanismos preexistentes o versiones modificadas de los mismos; o para la identificación de nuevas terapias basadas en la identificación del efecto de nuevos agentes, fármacos, compuestos o mecanismos con conocida o con predecible acción sobre las dianas.

18. Uso de un método según cualquiera de las reivindicaciones anteriores 1 a 16 para la predicción de los efectos producidos por la administración simultánea o consecutiva de dos o más agentes (e.g. fármacos y/o compuestos y/o moléculas y/o mecanismos), tanto de efectos adversos, secundarios, o de efectos cruzados producidos entre estos agentes (fármacos y/o compuestos,

etc.). Así como, también de efectos deseados ayudando en el diseño de nuevas combinaciones de agentes (e.g. fármacos, compuestos, moléculas y/o mecanismos con fines terapéuticos).

19. Método de identificación de los efectos fenotípicos de agentes moleculares con potencial terapéutico que comprende:

- 1) un método conforme cualquiera de las reivindicaciones anteriores 1 a 16; y
- 2) una etapa de validación funcional, por ejemplo mediante un sistema HCS (*"High-Content Screening"*).

20. Uso de un método según la reivindicación anterior para la identificación de los efectos de un agente o cualquier otro compuesto o mecanismo, así como, su uso para ayudar a identificar potenciales reposicionamientos para nuevas terapias de agentes, fármacos, compuestos o mecanismos preexistentes o versiones modificadas de los mismos; o para la identificación de nuevas terapias basadas en la identificación del efecto de nuevos agentes, fármacos, compuestos o mecanismos con conocida o con predecible acción sobre las dianas.

21. Uso de un método según la reivindicación 19 para la identificación de los efectos producidos por la administración simultánea o consecutiva de dos o más agentes (e.g. fármacos y/o compuestos y/o moléculas y/o mecanismos), tanto de efectos adversos, secundarios, o de efectos cruzados producidos entre estos agentes (fármacos y/o compuestos, etc.). Así como, también de efectos deseados ayudando en el diseño de nuevas combinaciones de agentes (e.g. fármacos, compuestos, moléculas y/o mecanismos con fines terapéuticos).

22. Aparato, dispositivo o sistema informático adaptado o configurado o que comprende medios que para llevar a cabo un método conforme cualquiera de las reivindicaciones 1 a 16.

23. Programa de ordenador que comprende instrucciones que, al ser ejecutado en un aparato dispositivo o sistema informático conforme a la reivindicación anterior, hacen que dicho aparato, dispositivo o sistema informático lleve a cabo un método conforme cualquiera de las reivindicaciones 1 a 16.

24. Medio legible por ordenador (memoria) que comprende instrucciones que, al ejecutarse en un aparato dispositivo o sistema informático conforme a la reivindicación 22, hacen que dicho aparato, dispositivo o sistema informático lleve a cabo un método conforme cualquiera de las reivindicaciones 1 a 16.

25. Un soporte de datos legible por ordenador donde se encuentra almacenado el programa de la

reivindicación 23.

26. Una señal portadora que transporta el programa de ordenador de la reivindicación 23.

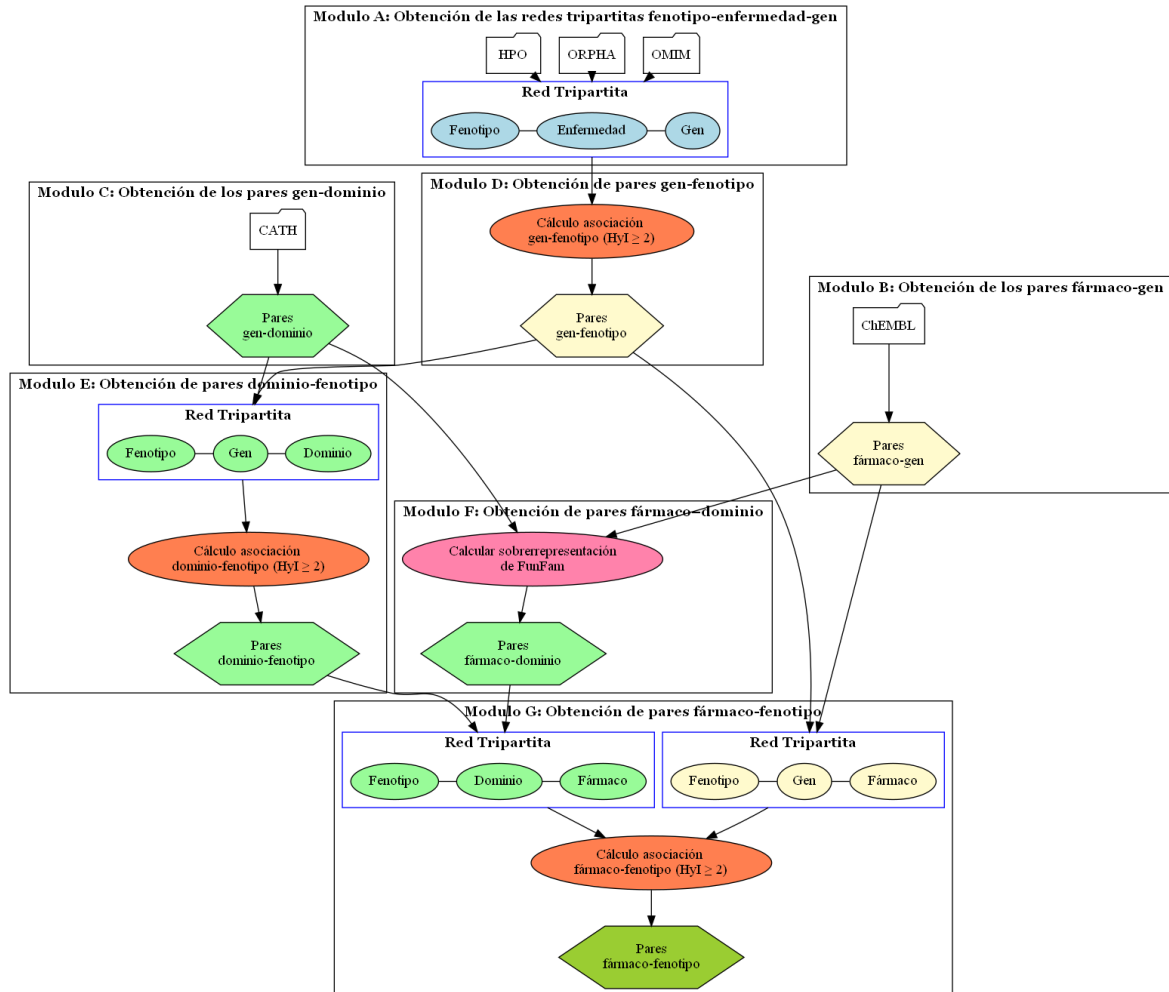


FIGURA 1

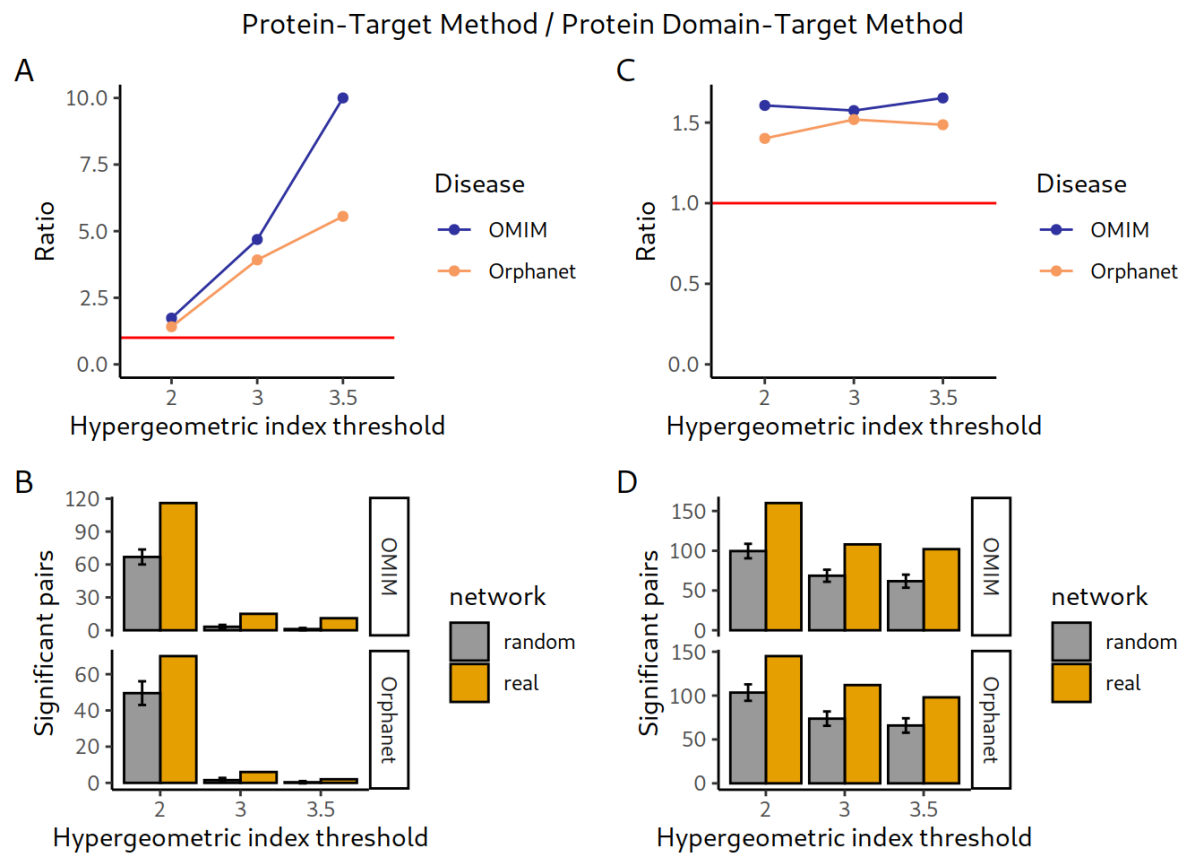


FIGURA 2

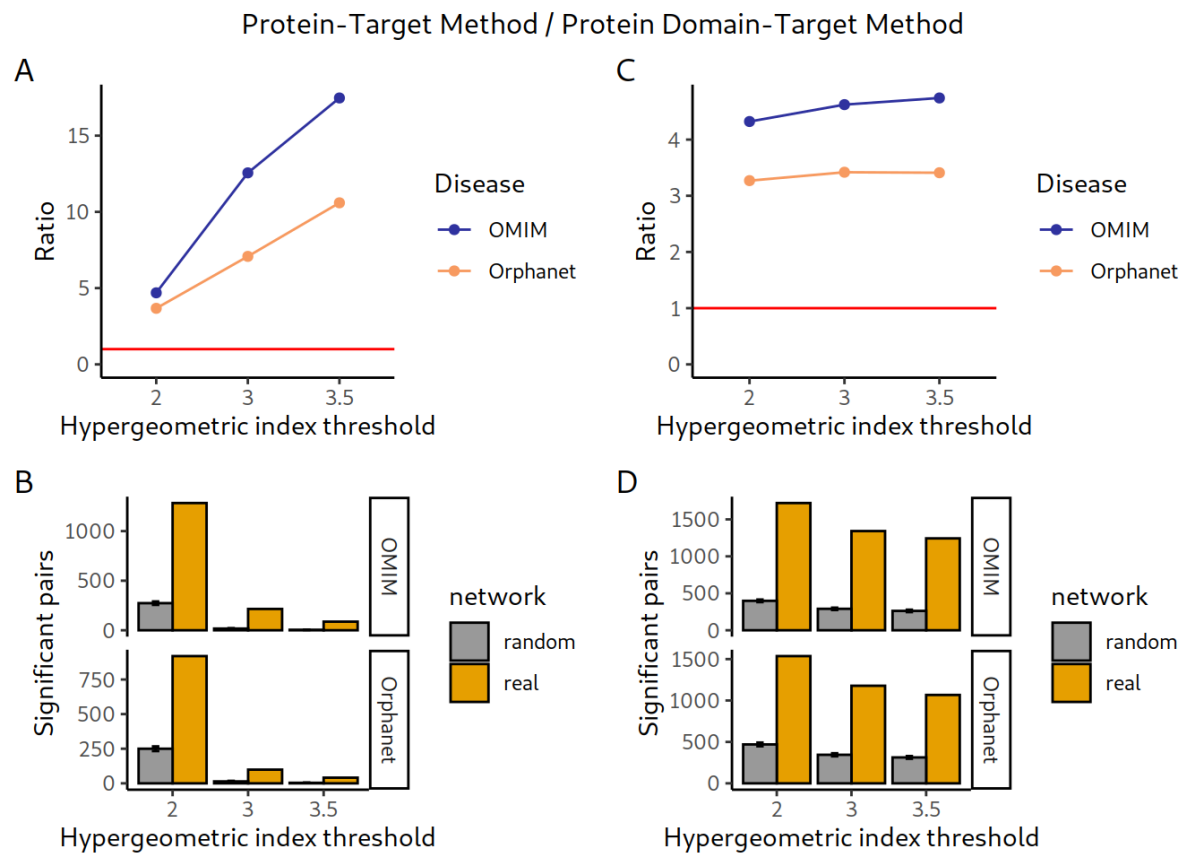


FIGURA 3



- 21 N.º solicitud: 202431096
22 Fecha de presentación de la solicitud: 24.12.2024
32 Fecha de prioridad:

INFORME SOBRE EL ESTADO DE LA TECNICA

51 Int. cl.: Ver Hoja Adicional

DOCUMENTOS RELEVANTES

Categoría	56 Documentos citados	Reivindicaciones afectadas
X	XUE, RUI et al.: "Prediction of Adverse Drug Reactions by Combining Biomedical Tripartite Network and Graph Representation Model". Chemical research in toxicology, 21 01 2020, 13/12/2019, Vol. 33, Nº 1, páginas 202-210, ISSN 1520-5010 (electronic), <DOI: 10.1021/acs.chemrestox.9b00238>.	1-26
X	ZENG, PAN et al.: "Drug repositioning based on tripartite cross-network embedding and graph convolutional network". Expert Systems With Applications. ELSEVIER. Yigitcanlar Tan, 06/05/2024, Vol. 252, ISSN 0957-4174, <DOI: 10.1016/j.eswa.2024.124152>.	1-26
X	KAWICHAJ, THITIPONG et al.: "Meta-Path Based Gene Ontology Profiles for Predicting Drug-Disease Associations". IEEE Access. IEEE, 10/03/2021, Vol. 9, Páginas 41809-41820, ISSN 2169-3536 (electronic), <DOI: 10.1109/ACCESS.2021.3065280>.	1-26
A	CN 115512857 A (UNIV CENTRAL CHINA NORMAL) 23/12/2022; resumen; párrafos [0021-0025].	1-26
A	ROJANO, ELENA et al.: "Revealing the Relationship Between Human Genome Regions and Pathological Phenotypes Through Network Analysis". Bioinformatics and Biomedical Engineering. 2017-04-01. In: Rojas, I., Ortuño, F. (EDS) Bioinformatics and Biomedical Engineering. IWBBIO 2017. Lecture Notes in Computer Science (), Vol. 10208. Springer, Cham. (Electronic), <DOI: 10.1007/978-3-319-56148-6_17>.	1-26

Categoría de los documentos citados X: de particular relevancia Y: de particular relevancia combinado con otro/s de la misma categoría A: refleja el estado de la técnica	O: referido a divulgación no escrita P: publicado entre la fecha de prioridad y la de presentación de la solicitud E: documento anterior, pero publicado después de la fecha de presentación de la solicitud
--	--

El presente informe ha sido realizado ☒ para todas las reivindicaciones ☐ para las reivindicaciones nº:

Fecha de realización del informe 20.10.2025	Examinador S. Sánchez Paradinas	Página 1/2
---	---	----------------------

CLASIFICACIÓN OBJETO DE LA SOLICITUD

G16B15/30 (2019.01)
G16B20/00 (2019.01)
G16B40/00 (2019.01)
G16H70/40 (2018.01)

Documentación mínima buscada (sistema de clasificación seguido de los símbolos de clasificación)

G16B, G16H

Bases de datos electrónicas consultadas durante la búsqueda (nombre de la base de datos y, si es posible, términos de búsqueda utilizados)

INVENES, EPODOC, WPI, NPL, INTERNET