



Comparación de resultados de dos técnicas de biología de sistemas para la aplicación de un análisis de reposicionamiento de fármacos

Eva Gabarrón Romero

Máster universitario en Bioinformática y bioestadística

Consultor UOC: Brian Jiménez García

Tutor UOC: María Jesús Marco Galindo

Tutor externo: José Manuel Mas Benavente

24 de mayo de 2017

Copyright © (Eva Gabarrón Romero)

Reservados todos los derechos. Está prohibido la reproducción total o parcial de esta obra por cualquier medio o procedimiento, comprendidos la impresión, la reprografía, el microfilme, el tratamiento informático o cualquier otro sistema, así como la distribución de ejemplares mediante alquiler y préstamo, sin la autorización escrita del autor o de los límites que autorice la Ley de Propiedad Intelectual.

FICHA DEL TRABAJO FINAL

Título del trabajo:	<i>Comparación de resultados de dos técnicas de biología de sistemas para la aplicación de un análisis de reposicionamiento de fármacos</i>
Nombre del autor:	<i>Eva Gabarrón Romero</i>
Nombre del consultor/a:	<i>Brian Jiménez García</i>
Nombre del tutor UOC:	<i>María Jesús Marco Galindo</i>
Fecha de entrega (mm/aaaa):	05/2017
Titulación:	<i>Máster universitario en Bioinformática y bioestadística</i>
Área del Trabajo Final:	<i>TFM "Ad-hoc". Tema propuesto por el alumno</i>
Idioma del trabajo:	<i>Castellano</i>
Palabras clave	<i>Biología de sistemas, reposicionamiento.</i>
Resumen del Trabajo:	
<p>Durante los últimos años, han surgido multitud de métodos diferentes que aplican la biología de sistemas para responder preguntas relacionadas con la fisiología humana. El análisis de reposicionamiento de fármacos ejemplifica las aplicaciones posibles que pueden tener este tipo de nuevas tecnologías; el cual consiste en descubrir nuevas indicaciones para fármacos que se utilizan para otras indicaciones. Sin embargo, es sabido que para un mismo análisis biocomputacional pueden generarse resultados distintos al usar diferentes técnicas. Es por ello, que el presente trabajo se ha propuesto, para comparar los resultados de dos de estas metodologías: la TPMS (<i>Therapeutic Performance Mapping System</i>), de Anaxomics Biotech y el método basado en segmentación de estructuras químicas, desarrollado por el Prof. Aloy y su equipo (IRB Barcelona).</p> <p>Para alcanzar dicha finalidad, se ha llevado a cabo un análisis de reposicionamiento para la patología del acné común, aplicando ambos métodos y se ha comparado la concordancia de los resultados. Además, se ha tenido acceso completo a uno de los métodos (TPMS), por lo que este se ha podido estudiar con más profundidad.</p>	

Los resultados muestran cierta concordancia entre los fármacos revelados como relacionados con la patología, a pesar de ello, también una cierta discordancia. Ambos resultados son complementarios y precisan de conocer el funcionamiento metodológico para comprender por qué aparecen resultados comunes y discordantes. Cada método aporta una perspectiva, capaz de detectar resultados verdaderos positivos distintos, fruto de la complejidad de la biología humana y la idiosincrasia de los métodos aplicados sobre ella.

Abstract:

During the last years, it has been appeared multiple different methods that apply systems biology to answer questions related to the human physiology. The reprofiling analysis of drugs exemplifies the possible applications to that kind of new technologies. That analysis consists in discovering new indications for a drug that are being used for other indications. Nevertheless, it is known that the results obtained for the same biocomputational analysis can provide different results by using different techniques. Because of that, it has been proposed for the present work, to compare the results of two different methods: TPMS (Therapeutic Performance Mapping System) of Anaxomics Biotech, and the fragment based chemical method of Professor P. Aloy and his group (IRB Barcelona).

Aim to that, it has been carried out a reprofiling analysis for the acne vulgaris, applying both methods, and it has been compared the concordance of the results. Also, it has been possible the complete access to the TPMS method, therefore, it has been studied more deeply.

The results show some concordance between the drugs revealed as related to the disease, although, some other show discordance. Both results are complementary and it is needed to know how the method works to understand them. Each technique provides a different sight, able to detect distinct true positive results, because of the complexity of human biology and the features of the methods applied.

Índice

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.....	4
1.5 Breve resumen de productos obtenidos.....	6
1.6 Breve descripción de los otros capítulos de la memoria	6
2. Contextualización	7
3. Aplicación práctica en farmacología: reposicionamiento para fármacos	8
3.1. Motivos científicos y clínicos del interés del reposicionamiento de fármacos	8
3.2. Motivos económicos para el reposicionamiento de fármacos.....	9
3.3. Reposicionamiento de fármacos desde el punto de vista de seguridad.....	9
3.4. Reposicionamiento de fármacos desde el punto de vista de eficacia	10
3.5. Estrategia de reposicionamiento para crear una nueva patente con una nueva entidad química.....	10
3.6. Fármacos asociados a biomarcadores: segmentación de pacientes	11
3.7. Análisis de reposicionamiento (reprofiling) y perfil de seguridad (safety profile):.....	11
4. Acné vulgaris: Caracterización de la patología.....	12
4.1. Epidemiología.....	12
4.2. Etiopatogenia.....	12
4.3. Proteínas implicadas en la etiopatogenia del acné: Efectores del acné.....	14
5. Métodos.....	14
5.1. Identificación de efectores para reposicionamiento	14
5.2. Descripción de la tecnología basada en segmentación de estructuras químicas.....	15
5.3. Descripción de la tecnología de Anaxomics, TPMS y Sampling Models	17
5.4. Elección de fármacos para estudiar su mecanismo de acción	26
6. Resultados	27
6.1. Resultados del método basado en subestructuras de compuestos químicos	27
6.2. Resultados del método TPMS	27
6.2.1. Reprofiling: Detección de fármacos candidatos.....	27
6.2.2. Identificación del mecanismo de acción	30
6.2.3. Resultados procedentes del estudio estadístico:.....	35
6.3. Comparación de ambos métodos	38
7. Análisis de resultados y conclusiones	39
7.1. Análisis de reposicionamiento	39
7.1.1. Método basado en segmentación de compuestos	39
7.1.2. Método TPMS.....	41
7.1.3. Resultados comunes y no comunes entre ambos métodos. .	42
7.2. Mecanismos de acción	43

7.3. Análisis estadístico	43
8. Conclusiones globales.....	44
9. Glosario	46
9.1. Listado de todo el software utilizado:.....	46
10. Bibliografía	47
11. Anexos	49
11.1. Front-End del software usado para el reposicionamiento y la subida de restricciones.....	50
11.2. Script en lenguaje Matlab utilizado con las llamadas a las funciones necesarias para la generación de los MoA con comentarios.....	52
11.3. Script en lenguaje Matlab utilizado para la extracción de los valores de activación.....	55
11.4. Script en lenguaje Matlab utilizado con las llamadas a las funciones para realizar el análisis estadístico.....	59
11.5. Tabla que contiene los nodos del MoA para el fármaco Somatropina recombinante.....	61
11.6. Tabla que contiene los nodos del MoA para el fármaco ácido Meclofenámico.....	62

Lista de figuras

Figura 1. Diagrama de Gantt.....	6
Figura 2. Esquema general de la patofisiología del acné (16).....	14
Figura 3. Representación de las distintas redes que pueden diferenciar y correlacionar con las distintas aplicaciones (18).	17
Figura 4. Representación esquemática de una topología.	20
Figura 5. Representación gráfica de los diferentes tipos de conocimiento incluidos como prior restrictions.	21
Figura 6. Diagrama de un perceptrón (31).	23
Figura 7. Representación esquema del proceso de elaboración de un MoA (32).	25
Figura 8. Equivalencias para valores de TPMS y valores de probabilidad ANN.	29
Figura 9. Función de densidad de probabilidad acumulada para ambas categorías en función de su valor de TPMS.....	29
Figura 10. Mecanismo de acción del fármaco Somatropina recombinante para producir acné (pre-validación).....	30
Figura 11. Leyenda para la lectura de links y nodos de los mecanismos de acción.	31
Figura 12. Mecanismo de acción del fármaco Somatropina recombinante para producir acné (post-validación).	31
Figura 13. Mecanismo de acción del fármaco ácido Meclofenámico para producir acné (pre-validación).....	32
Figura 14. Mecanismo de acción del fármaco ácido Meclofenámico para producir acné (post-validación).	33
Figura 15. Gráfico Volcano plot que representa el Log ₂ del Fold Change contra el Log ₂ del p-valor ajustado para el test de Wilcoxon.	37
Figura 16. Gráfico resultado del Análisis de Componentes Principales (ACP) para los componentes uno y dos.....	37
Figura 17. Gráficos de 360° correspondientes al porcentaje de fármacos que han sido detectados y por ambos métodos y por un único método (TPMS, a la izquierda y método quemo-céntrico desarrollado por el equipo del IRB Barcelona, a la derecha).	38

Lista de tablas

Tabla 1. Calendario final del proyecto.	5
Tabla 2. Tabla con información procedente de DrugBank los fármacos a estudiar. Las dianas para las que el efecto es desconocido no se introducen como parámetros en la generación del modelo.	26
Tabla 3. Tabla de resultados del reposicionamiento con el método TPMS mostrando únicamente los fármacos elegidos para continuar el estudio.	27
Tabla 4. Detalle de los parámetros de la construcción de los mecanismos de acción para ambos fármacos.	34
Tabla 5. Resultados del test de hipótesis de Wilcoxon.	36
Tabla 6. Comparación de resultados de reprofiling para la patología acné vulgaris.	38
Tabla 7. Identificación detallada de los nodos que aparecen en el mecanismo de acción realizado con el fármaco Somatropina recombinante con efector.	61
Tabla 8. Identificación detallada de los nodos que aparecen en el mecanismo de acción realizado con el fármaco ácido Meclofenámico recombinante con efector.	62

1. Introducción

1.1 Contexto y justificación del Trabajo

Existen distintos métodos que aplican biología de sistemas que nos permiten realizar predicciones y entender aspectos del comportamiento bajo determinadas condiciones.

El presente trabajo de final de máster (TFM) parte de la premisa de que distintos métodos, en ocasiones reportan resultados diferentes en respuesta al mismo análisis. Teniendo como propósito el poner de manifiesto las diferencias relacionadas con la aplicación de dos de estos métodos para un análisis de reposicionamiento. Con el interés final de intentar dilucidar la causa y, como resultado esclarecer los puntos donde se podría mejorar el sistema.

Se define como reposicionamiento (también llamado reprogramación o *reprofiling*) para fármacos, aquel análisis o experimento que tiene como objetivo identificar una nueva indicación diferente de aquellas para las que fueron lanzados al mercado.

Un fármaco se puede comercializar para una o más indicaciones cuando el fármaco ha superado las pruebas de eficacia y seguridad que marcan las instituciones sanitarias. Los requerimientos de estas instituciones se formalizan mediante la realización de estudios clínicos, que evalúan la eficacia y seguridad del fármaco. La eficacia está directamente relacionada con la indicación clínica para la que se ha desarrollado el fármaco. Sin embargo, la seguridad del fármaco es independiente de la indicación (si el fármaco va a ser administrado al mismo tipo de población evaluado por el ensayo clínico previo). Bajo estas premisas evidencia el interés del sector farmacéutico para dar un nuevo uso terapéutico a un fármaco que además está ya en el mercado.

Cabe destacar que el lanzamiento de un fármaco se rige por distintos criterios:

- Criterios científicos
- Control de las instituciones sanitarias
- Necesidades del mercado.

Esto puede significar que un fármaco con una conocida actividad terapéutica podría no llegar al mercado si existe otro fármaco para la misma indicación que colme las posibilidades de éxito comercial. Por tanto, existen multitud de posibilidades que van desde el desconocimiento al interés económico, por las cuales un fármaco pueda requerir el descubrimiento de indicaciones diferentes a aquellas que tiene reconocidas.

Existen diversas formas de detectar una nueva indicación para un fármaco. Un ejemplo es la derivada de la práctica clínica por parte de personal médico. Quienes, por conocimiento de la patología y del fármaco o por observación de la respuesta clínica del paciente a combinaciones de fármacos, son capaces de detectar actividad clínica inesperada en ciertos pacientes. Lo que genera que se practique el uso *off-label* (fuera de prospecto) de fármacos por su parte. En algunos casos estos tratamientos acaban siendo referentes de tratamientos, especialmente en enfermedades sin tratamiento conocido.

La industria farmacéutica tiene un interés claro con una motivación económica. En primer lugar, porque el mercado del fármaco puede aumentar, pero también, porque una nueva indicación alarga la patente de un fármaco cosa que les beneficia para todas las indicaciones ya existentes. La consecuencia final es que el reposicionamiento de fármacos es una actividad solicitada en el sector farmacéutico que requiere de herramientas y estrategias efectivas.

Por tanto, multitud de estrategias biocomputacionales se han implementado para abordar la tarea de reposicionamiento de fármaco. En este trabajo se presentan dos estrategias que han mostrado ser eficaces, se compara la concordancia de sus resultados identificando los elementos comunes y específicos de cada una de ellas¹.

1.2 Objetivos del Trabajo

Objetivos generales:

1. Estudio y comprensión de dos métodos biocomputacionales del campo de la biología de sistemas aplicados a un ejercicio de reprogramación.
2. Aplicación de los procedimientos sobre la patología del acné vulgar.
3. Comprensión de la causa de las diferencias en los resultados del análisis de reposicionamiento.

¹ Los resultados de esta tarea de reprogramación no serán mostrados en el proyecto, el proyecto únicamente se centrará en el análisis del efecto de fármacos conocidos que se evalúan de forma diferencial por estos los métodos de ensayo.

Objetivos detallados:

1. Estudio y caracterización del acné como enfermedad y evaluación de si es factible realizar el proyecto en base a ella.
Análisis de la metodología de ambos equipos, Prof. Aloy (1) y Anaxomics (2), (3) sobre los compuestos químicos con capacidad de producir o tratar el acné.
2. Entender el mecanismo biológico por el cual se produce la enfermedad a nivel molecular.
3. Entender las bases de clasificación del método de experimentación del Prof. Aloy.
4. Entender los detalles y funcionalidad de los modelos matemáticos creados por Anaxomics.
5. Identificar la esencia de los motivos de la clasificación diferencial producida con ambos métodos para las mismas predicciones.
6. Extraer conclusiones que permitan mejorar los métodos o la aplicación de ambos métodos para la toma de decisión final.

1.3 Enfoque y método seguido

Estrategia escogida

1. Caracterización de los procesos moleculares que producen Acné vulgar tanto a nivel molecular como a nivel fisiológico
2. Identificar y comprender los procedimientos metodológicos de dos métodos en relación al comportamiento de fármacos cuando se testean como posibles candidatos al tratamiento del acné como patología o condición biológica.
3. Comparación de resultados y estrategias de ambos métodos.

Posibles estrategias

La estrategia escogida corre el riesgo de conformar lo que se denomina una falacia de evidencia completa (cherry-picking en inglés), en el sentido de que podemos estar seleccionando lo mejor de los métodos implementados. Esto es debido a que el estudio no se puede llevar a cabo si el sistema no contiene suficiente información para el análisis de una determinada patología y, por tanto, la estrategia escogida no hubiera podido aplicarse tal y como se ha hecho. Consecuentemente, otra posible estrategia podría consistir en el estudio de las

causas por las que uno de los sistemas o ambos no es capaz de reportar resultados.

1.4 Planificación del Trabajo

La planificación inicial del trabajo se conformaba con los siguientes puntos inicialmente:

- Se realizará una búsqueda de información sobre la patología a estudiar, con el objetivo doble de conocer la enfermedad y conocer los inputs apropiados para la realización del modelo.
- Seguidamente se realizará una documentación sobre el primer método a aplicar (TPMS) y aplicación del mismo. Se dispone de información interna de la compañía no publicada pero no confidencial para la consecución de los objetivos del master.
- Posteriormente se llevará a cabo una documentación sobre el segundo método a aplicar (basado en segmentación). En este caso es posible que la realización del experimento sea realizada por parte del IRB de Barcelona, fundadora del método.
- Finalmente se realizará el análisis de los resultados y se extraerán si es posible una serie de conclusiones.

Estos puntos, concretados como las tareas que se muestran en la Tabla 1, fueron definidas al inicio del proyecto y reportadas en el informe PEC 1. Todos los puntos se corresponden con el trabajo que finalmente se ha realizado, con excepción de la tarea número 16, la cual comprende la realización de un estudio estadístico. Este análisis no se había planificado desde el comienzo, sino que fue añadida durante la realización de la PEC 3, por lo que ha sido añadida en el informe PEC 3 y la presente memoria. Las mayores modificaciones se han realizado sobre el tiempo destinado a cada tarea, que han tenido que ir siendo ajustados conforme se ha llevado a cabo el proyecto.

Nº Tarea	Nombre de tarea	Duración	Comienzo	Fin	Realizada
1	PEC 0: Presentación y lectura del plan docente	5 días	mié 22/02/17	mar 28/02/17	✓
2	Búsqueda y lectura de publicaciones sobre la patología del acné	6 días	mié 01/03/17	mié 08/03/17	✓
3	Planificación del trabajo y redacción de la PEC 1	5 días	jue 09/03/17	mié 15/03/17	✓
4	PEC 1	11 días	mié 01/03/17	mié 15/03/17	✓
5	Lectura de información sobre métodos de inteligencia artificial	6 días	jue 16/03/17	jue 23/03/17	✓
6	Método 1: Estudio del software a utilizar	6 días	vie 24/03/17	vie 31/03/17	✓
7	Redacción de la PEC2	3 días	lun 03/04/17	mié 05/04/17	✓
8	PEC 2	15 días	jue 16/03/17	mié 05/04/17	✓
9	Método 1: Obtención de los primeros resultados	4 días	jue 06/04/17	mar 11/04/17	✓
10	Lectura de información y comprensión del funcionamiento del método 2	2 días	mié 12/04/17	jue 13/04/17	✓
10	Lectura de información y comprensión del funcionamiento del método 2	2 días	mar 18/04/17	mié 19/04/17	✓
11	Método 2: Estudio del software a utilizar; Obtención de los primeros resultados	9 días	jue 20/04/17	mar 02/05/17	✓
12	Evaluación y comparación de los resultados	2 días	mié 03/05/17	jue 04/05/17	✓
13	Interpretación de los resultados	1 día	vie 05/05/17	vie 05/05/17	✓
14	Redacción de PEC 3	3 días	lun 08/05/17	mié 10/05/17	✓
15	PEC 3	25 días	jue 06/04/17	mié 10/05/17	✓
16	Estudio estadístico; Evaluación e interpretación de los resultados	4 días	jue 11/05/17	mar 16/05/17	✓
17	Redacción de la memoria final	7 días	mar 16/05/17	mié 24/05/17	✓
18	Preparación de la presentación en PowerPoint	1 día	mié 24/05/17	mié 24/05/17	✓
19	Entrega de la memoria final (documento y presentación)	10 días	jue 11/05/17	mié 24/05/17	✓
20	Preparación de la defensa	9 días	jue 25/05/17	mar 06/06/17	Próximo hito
21	Entrega de la defensa pública	11 días	mié 07/06/17	mié 21/06/17	Próximo hito

Tabla 1. Calendario final del proyecto.

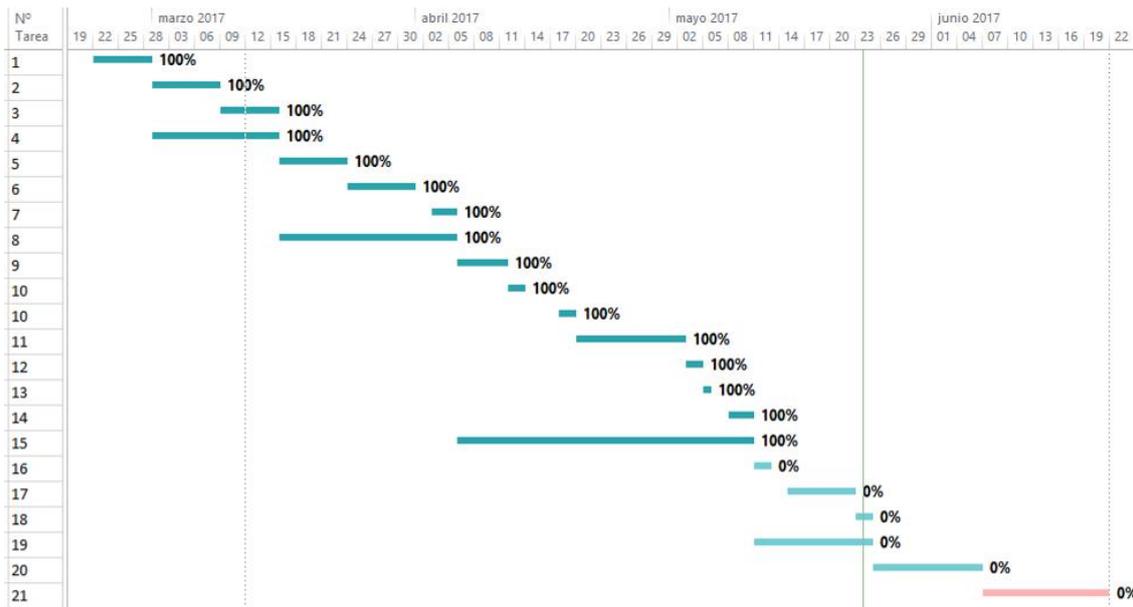


Figura 1. Diagrama de Gantt

1.5 Breve resumen de productos obtenidos

- Comparación de los métodos previamente definidos.
- Mecanismos de acción de dos de los fármacos obtenidos a raíz de los resultados de uno de los métodos.
- Análisis estadístico que pretende comparar el comportamiento de las mismas variables (proteínas) según se aplique un fármaco u otro.

1.6 Breve descripción de los otros capítulos de la memoria

El apartado siguiente (2. Resto de capítulos) se estructura siguiendo el método científico habitual.

- ❖ Primeramente, encontramos una **introducción** en la que se detalla la importancia y el interés por este tipo de análisis y se describe también la patología estudiada, dado que ha sido un proceso necesario para preparar el análisis.
- ❖ El siguiente apartado consiste en presentar los **métodos** utilizados para realizar el estudio.
- ❖ Seguidamente encontramos un apartado de **resultados**, en los que se presentan los resultados obtenidos mediante tablas y figuras.
- ❖ Finalmente encontramos un apartado dedicado a las **conclusiones** e interpretación de los resultados.

2. Contextualización

La biología de sistemas es un campo de investigación que comenzó a desarrollarse en torno a la década de los 60, para establecerse más tarde, como disciplina académica a lo largo del año 2.000. Desde entonces se ha expandido mucho, siendo considerada como multidisciplinar, a consecuencia de que reúne enfoques de distinto origen, como puede ser el biológico, matemático e informático. Constituyendo sus aplicaciones más importantes en los campos de farmacología y la biotecnología (4), (5).

En términos generales, la biología de sistemas consiste en el estudio de un organismo o sistema desde un punto de vista integrador y global que es capaz de interrelacionar sus componentes, los cuales serán predominantemente proteínas y genes.

El comportamiento de un organismo puede ser predicho *in silico* con este sistema gracias a la creación de un modelo matemático, el cual ha de ser entrenado y validado a partir de un conjunto de datos. El modelo permite recrear o simular procesos biológicos con una aplicabilidad práctica como puede ser el predecir una respuesta ante la aplicación de un tratamiento, por ejemplo, un fármaco, o una determinada situación, como una enfermedad, que provoca un cambio en el modelo.

La aplicación biomédica o biotecnológica de este tipo de tecnología se puede abordar a través de diversas aproximaciones o métodos (6), (7), (8).

Centrando el interés en esto, el presente trabajo de final de máster, va a consistir en comparar dos de las aproximaciones posibles, enfocándose en evaluar la causa subyacente a las diferencias o errores que se puedan observar en los resultados.

Se pretende realizar una evaluación de los siguientes métodos:

- TPMS (Therapeutic Performance Mapping System) de ANAXOMICS, basada en la utilización de distintos parámetros que describen desde diferentes puntos de vista el sistema biológico.

- Métodos basados en el potencial químico estadístico, descrito por el Prof. P. Aloy (IRB Barcelona).

3. Aplicación práctica en farmacología: reposicionamiento para fármacos

Una de las aplicaciones de interés de las técnicas computacionales en farmacología es la identificación de nuevas indicaciones para fármacos, análisis denominado como reposicionamiento (en inglés reprofiling o repurposing).

El reposicionamiento de fármacos tiene interés tanto desde una perspectiva clínica como desde una perspectiva económica (6) (7). Por un lado, los clínicos precisan más herramientas químicas que les permitan tratar de modo cada vez más eficaz los procesos patológicos. Por otro lado, el mercado farmacéutico es muy competitivo y la inversión necesaria para conseguir que un fármaco alcance el mercado es muy alta y con una tasa de éxito muy baja. Ambas perspectivas, 3precisan, por tanto, de la identificación de nuevos fármacos o indicaciones que les permita cumplir sus objetivos.

3.1. Motivos científicos y clínicos del interés del reposicionamiento de fármacos

Sobre todo, para enfermedades que carecen de tratamiento, patologías de baja prevalencia o cuando los tratamientos habituales dejan de ser efectivos para un paciente, los clínicos necesitan disponer de herramientas, normalmente tratamientos químicos, que les permitan atender las necesidades de los pacientes. En estos casos, y otros muchos, un prescriptor puede optar por la administración de tratamientos *off-label*, es decir, tratamientos que tienen una intención terapéutica diferente de la indicada y aprobada para el fármaco. Normalmente los médicos apoyan esta decisión en la experiencia, pero sobretodo en publicaciones científicas donde otros facultativos sugieren ese uso *off-label* del fármaco.

A simple vista parece que el uso *off-label* del fármaco es un proceso previo que necesariamente debería conducir a los propietarios de la patente del compuesto a hacer el desarrollo clínico del fármaco para esa nueva indicación. Sin embargo, esto no siempre es así. Los motivos subyacen en el sentido económico del proceso. Por un lado, el coste de desarrollo clínico implica elevados costes de desarrollo. Por otro lado, las enfermedades raras o de menor interés económico no garantizan a los inversores farmacéuticos el retorno de su inversión. Consecuentemente los médicos no disponen de la indicación en los fármacos que posiblemente desde un punto de vista funcional, clínico y práctico es adecuada.

- Existen muchos ejemplos en la literatura de fármacos que se usan de modo *off-label*. Un ejemplo conocido en Anaxomics es el propuesto por el Dr. José Alegre (Coordinador de la unidad de CFS en el Hospital *Vall d'Hebrón Institut de Recerca*) para el tratamiento de la fatiga crónica. Esta enfermedad no dispone de tratamientos indicados para ella, sin embargo, el equipo del Dr. Alegre ha comprobado y demostrado que el fármaco Rituximab es un tratamiento adecuado para muchos de sus pacientes (9).

3.2. Motivos económicos para el reposicionamiento de fármacos

Es evidente que cuantas más indicaciones tiene un fármaco más mercado potencial tiene, por tanto, es más susceptible de generar ingresos. El curso de vida de un fármaco es relativamente breve y va ligado a la patente que protege su explotación, no dura más de 10 años.

Sin embargo, sólo en determinados casos una nueva indicación que extienda la patente del fármaco es realmente de interés para el mercado.

En el caso de nuevos lanzamientos de fármacos o aquellos que están en el punto álgido de su ciclo de vida, el hecho de identificar una nueva indicación podría tener sentido económico. Sin embargo, la indicación debe también acompañarse a ese ciclo de vida del fármaco. Por ejemplo, indicaciones cuyos estudios clínicos requieran el seguimiento de pacientes durante años son especialmente complicadas porque agotan el ciclo de vida del compuesto.

Como puede verse, aunque desde perspectivas altruistas disponer de nuevas indicaciones para fármacos parece obvio cuando se contextualiza todo el proceso aparecen limitaciones que condicionan el proceso de reposicionamiento.

3.3. Reposicionamiento de fármacos desde el punto de vista de seguridad

Una ventaja obvia de reposicionar un fármaco es el hecho de utilizar un compuesto químico cuyo perfil de seguridad es bien conocido, puesto que ya está en el mercado. Por tanto, el desarrollo clínico que se produzca sobre este fármaco será centrado por tanto en su eficacia. No por ello se eliminará la aplicación de los sistemas de control adecuados, pero posiblemente, serán superados más fácilmente. Una prueba de ello es que las fases más tempranas de desarrollo de un fármaco Fase I podría negociarse con las instituciones sanitarias para ser dadas por superadas. Evidentemente, este efecto está directamente relacionado con los tiempos y costes de implementación de un fármaco hasta llegar al mercado, y sobretodo, con el riesgo de la inversión que se centraría en el perfil de eficacia de la nueva indicación.

Una perspectiva interesante es que los fármacos no sólo pueden desarrollarse con la intención de ser más eficaces que otros fármacos ya existentes para una indicación. En algunos casos donde los criterios *Standard of Care* (SoC) de la indicación presenta perfiles de seguridad no confortables existe la posibilidad de que un fármaco, incluso menos eficaz, pueda alcanzar el mercado. En este caso, el fármaco sobre el que se aplicaría el reposicionamiento deberá cumplir un perfil de eficacia quizás no tan exigente pero sí un perfil de seguridad más estricto. Serán las instituciones sanitarias quienes autoricen o no tal estrategia.

3.4. Reposicionamiento de fármacos desde el punto de vista de eficacia

El planteamiento más clásico de diseño de un fármaco es buscar un compuesto eficaz contra una indicación. El reposicionamiento de un fármaco no es ajeno a esto. Sin embargo, aunque parezca poco probable que un fármaco muestre una enorme eficacia para una indicación para la cual no fue diseñado, se han dado casos. Un ejemplo destacado y ampliamente conocido es el caso del fármaco Viagra, cuyo efecto adverso acabó siendo la principal indicación del compuesto.

3.5. Estrategia de reposicionamiento para crear una nueva patente con una nueva entidad química

El desarrollo de un fármaco, sea por reposicionamiento o no, va muy ligado a la patente de explotación económica del mismo. Sin un retorno económico es difícil que se invierta en el coste de desarrollo del mismo. La ampliación de una patente no siempre justifica la inversión de desarrollo clínico del compuesto para la nueva indicación. Por tanto, es necesario definir estrategias que permitan disponer del máximo tiempo posible para su explotación comercial.

Hay 3 estrategias básicas:

- (a) la modificación de la estructura química del compuesto mediante la adición/sustracción de parte de su estructura.
- (b) variar la vía de administración, forma, dosis u otros elementos que permitan desarrollar una nueva patente
- (c) combinar el fármaco con otros compuestos químicos que permitan definir la combinación como una nueva entidad química.

Las estrategias (a) y (b) tienen como principal limitación las propias patentes de los compuestos candidatos. Usualmente las patentes de los compuestos ya contemplan todas las opciones de comercialización y variaciones posibles. La opción (c) por su parte requiere de una nueva patente, y, por tanto, hay que demostrar la innovación. La mejor manera de cumplir estos requerimientos es disponer de una combinación de compuestos, ninguno de los dos con patentada sobre la indicación deseada, y que juntos tienen una actividad sinérgica sobre

esta nueva indicación. Es decir, encontrar 2 compuestos que, combinados muestren una eficacia significativamente superior a la suma de los dos individualmente. Matemáticamente podríamos expresarlo como que $A+B < AB$.

La estrategia (a) y (b) quedan fuera del alcance de este trabajo, sin embargo, la estrategia (c) es abordable mediante uno de los métodos que se van a utilizar, concretamente el TPMS de Anaxomics (1), (2).

3.6. Fármacos asociados a biomarcadores: segmentación de pacientes

Cada vez es más complicado el lanzamiento de un compuesto que tenga un alcance global a una patología. En los últimos años está proliferando la asociación de fármacos con biomarcadores, que permiten identificar aquellos pacientes para los cuales el tratamiento va a resultar especialmente efectivo y seguro.

Estos biomarcadores son más sencillos de descubrir cuando el desarrollo del fármaco se basa en el conocimiento del mecanismo de acción del fármaco y en el conocimiento de los procesos moleculares que subyacen bajo la enfermedad. Este hecho es también accesible a la metodología TPMS de Anaxomics.

3.7. Análisis de reposicionamiento (reprofiling) y perfil de seguridad (safety profile):

El análisis se ha llevado a cabo utilizando un software desarrollado por Anaxomics que se agrupa en la tecnología denominada TPMS (10). Se inputan como entrada (*inset*) una lista de targets, es decir, las proteínas diana de una gran cantidad de fármacos a evaluar. Como salida (*outset*) se indica la patología con la que se desea relacionar, el acné en este caso, pero se inputa como un set de proteínas que la representan. Es por ello que es necesaria la caracterización de la patología, para descubrir qué proteínas son las que están más estrechamente relacionadas con dicha enfermedad.

La misma herramienta es aplicable tanto para descubrir nuevas indicaciones como efectos adversos, en cuyo caso podríamos denominarlo perfil de seguridad.

4. Acné vulgaris: Caracterización de la patología

Se define como acné vulgar a la enfermedad crónica inflamatoria de la piel que involucra lesiones localizadas en las unidades pilosebáceas que derivan a la formación de comedones, pápulas, pústulas, nódulos y cicatrices de distribución facial principalmente (11).

4.1. Epidemiología

Se considera una enfermedad muy común mundialmente, según algunos estudios, casi el 100% de la población lo sufre en algún momento de su vida (12), su prevalencia es notoriamente mayor durante la etapa de la pubertad y la adolescencia.

La gravedad de la enfermedad se considera indirecta, dado que la calidad de vida del paciente puede verse afectada debido a la carga psicológica asociada llegando a incrementar la incidencia de ideales autolíticos (13).

4.2. Etiopatogenia

La piel es un órgano de defensa contra la entrada de potenciales elementos infecciosos, toxinas o distintas sustancias provenientes del exterior, a la vez que evita la pérdida de fluidos internos. Constituye una barrera física gracias al estrato córneo (hidrofóbico) a la par que química (sintetiza proteasas, lisozimas y péptidos antibacterianos) (14).

La forman tres capas, una más externa o epidermis y otra más interna o dermis y por último la llamada subdermis. La dermis presenta distintas subestructuras (o anexos cutáneos), como las unidades pilosebáceas, entre otros.

Cada unidad pilosebácea se compone de un folículo piloso y una glándula sebácea.

El acné constituye una alteración multifactorial que puede ser inducida por una mezcla de factores endógenos y exógenos. Una lesión inicial e invisible para el ojo humano denominada microcomedón puede evolucionar a comedón (no inflamatorio) o pústulas, pápulas o nódulos (inflamatorio). Ambos requieren de una queratinización folicular alterada, hiperplasia de las glándulas sebáceas y hipercolonización de *Propionibacterium acnes* en el folículo.

Hiperqueratinización

Alteraciones en el equilibrio de ácidos grasos contenidos en el sebo o una deficiencia de andrógenos (a través de la enzima 5- α -reductasa), pueden

desencadenar un proceso hiperqueratinización. Este evento puede modificar a su vez la diferenciación de los queratinocitos y composición del sebo, promoviendo altos niveles de interleucina-1 α ; además de derivar en una mayor proliferación de la mencionada bacteria, ambos capaces de estimular inflamación.

Hiperplasia de las glándulas sebáceas y seborrea

El exceso de queratinización combinado con un exceso de sebo puede iniciar el ensanchamiento de los folículos. Esto forma parte del proceso fisiológico natural que se inicia al cumplir los 7-8 años de edad debido a cambios hormonales. Asimismo, los cambios más iniciales en el proceso de evolución del acné también requieren del exceso de ambos, resultando en la formación del llamado microcomedón.

Las lesiones de acné muestran una elevada proliferación de las células epiteliales del folículo, derivada a menudo por un incremento de la sensibilidad a andrógenos y cambios en la composición de los lípidos del sebo (15).

Colonización bacteriana

Las condiciones ambientales que se dan en el folículo con la patología o durante el proceso de formación del acné, son óptimas para la proliferación de especies de propionibacterium, concretamente, de la especie *P. acné* principalmente. Existiendo una clara correlación entre la colonización bacteriana del conducto pilosebáceo y el inicio del micro comedón. Así como la relación inversa entre la colonización y la cura de la patología. Sin embargo, no hay evidencias que correlacionen la cantidad de dicha flora en la superficie de la piel con la severidad del acné.

Seguramente la razón más importante para que la colonización de dicha población bacteriana afecte a la evolución de la patología es la promoción de la inflamación que va asociada con ella. La activación de vías inflamatorias debido a ello tiene lugar mediante la activación de receptores TLR y enzimas de degradación.

Factores inmunológicos e inflamación

El papel de la inflamación puede verse como primario, dado que en algunos casos se ha demostrado que tiene poder para iniciar la comedogénesis, sobretodo, a través de la interleucina 1. Sin embargo, también puede verse como secundario dado que la inflamación se ve implicada en los mecanismos de promoción de la producción de sebo. Correlacionándose la inflamación prolongada se considera como un marcador de la propensión del acné a dejar marcas o cicatrices permanentes (16).

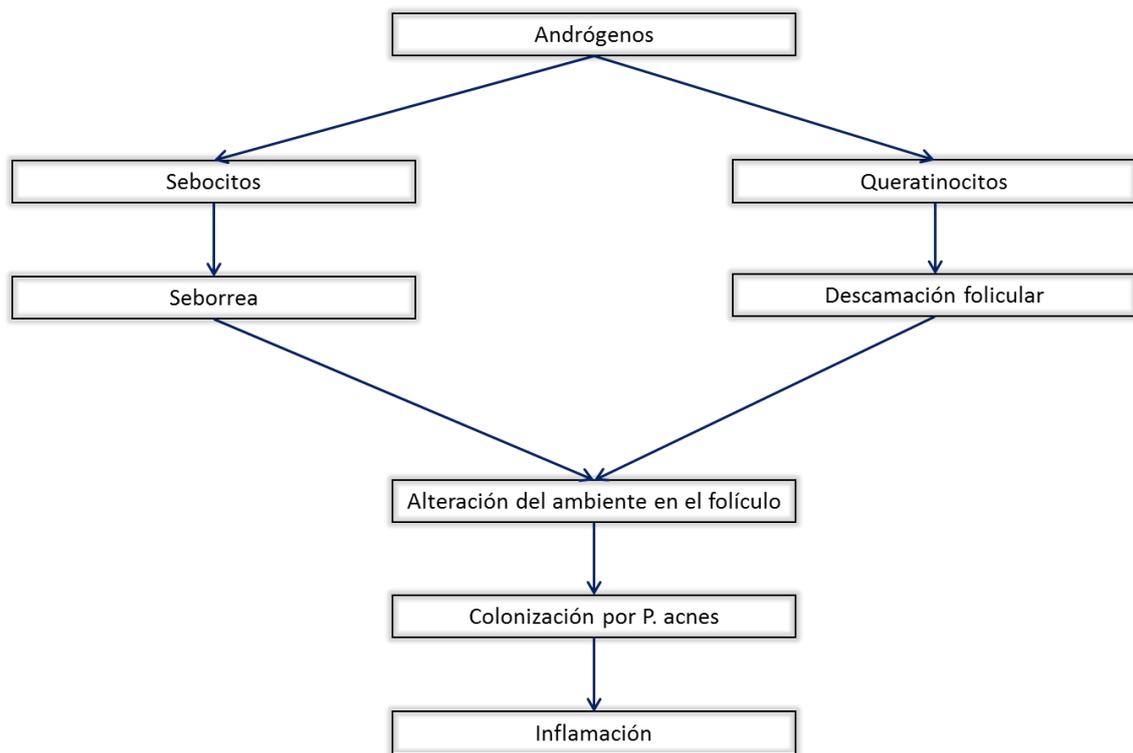


Figura 2. Esquema general de la patofisiología del acné (16).

4.3. Proteínas implicadas en la etiopatogenia del acné: Efectores del acné

Se recogen para el modelo las 79 proteínas que se han demostrado relacionadas directamente con el desarrollo de la patología del acné en publicaciones científicas. Y, por tanto, se han inputado como efectores de la patología.

5. Métodos

5.1. Identificación de efectores para reposicionamiento

Ambas tecnologías precisan de un input o efector para relacionarlo con la información clínica (condiciones/ patologías), que en este caso serán los fármacos a evaluar. Dichos fármacos se encuentran integrados en la base de datos de las compañías y constituyen un total de alrededor 6.500 compuestos procedentes de la base de datos DrugBank (17).

5.2. Descripción de la tecnología basada en segmentación de estructuras químicas

La evidencia demuestra que hasta los fármacos más específicos presentan perfiles de interacción heterogéneos. Lo que ha llevado a pensar que la visión compleja e integrativa dada por la aplicación de la biología de sistemas es necesaria para comprender la farmacología y toxicología (18), (2).

El sistema se basa en el potencial químico estadístico, concretamente, el método inicial es el llamado SEA, *Similarity Ensemble Approach*. Dicho procedimiento fue diseñado para calcular una medida cuantitativa llamada coeficiente de Tanimoto (Tc), que determina la similitud entre ligandos. Los valores de Tc oscilan entre 0 y 1, según muestren menor o mayor similitud, respectivamente. La finalidad del cual es poder relacionar y clasificar conjuntos de proteínas biológicamente en función de la similitud de sus ligandos (19).

La versión adaptada no se restringe solamente a evaluar la región de unión al ligando. Para este análisis son utilizados pequeños fragmentos de toda la proteína, que constituyen una huella química (FP2). Se considera, por tanto, un enfoque quemo-céntrico.

Para aplicar el análisis se utilizan aproximaciones estadísticas y de probabilidad; De un modo simplificado, para comparar dos grupos de moléculas en estudio, es necesario sumar sus valores de Tc (únicamente se consideran aquellos que se encuentren por encima de un punto de corte, como es 0.55). La similitud entre cada grupo o set es comparada estadísticamente mediante el cálculo de Z-score, E-valor y p-valor.

Esta metodología es muy útil a la hora de realizar comparaciones o relaciones entre proteínas, fármacos o enfermedades. En este caso, lo que se ha realizado es un análisis de reprofiling, a partir de datos procedentes de bases de datos como CTD (Comparative Toxicogenomics Database) (18). Es posible relacionar fármacos con una enfermedad (ya sea de forma curativa o como efecto adverso). Además, bases de datos ChEMBL o BindingD, son utilizadas para conocer interacciones entre moléculas (targets y ligandos).

Con toda la información se completa una red de condiciones humanas, a partir de la cual, es posible buscar patrones entre los fármacos, para catalogar determinadas huellas moleculares junto con patologías, indicaciones o efectos tóxicos. Más tarde, es posible aplicar técnicas de *machine learning*, para crear con ello modelos entrenados, que al igual que en el caso de Anaxomics, permitan predecir relaciones desconocidas.

Pasos para crear la red:

1) Identificación de las partes químicas asociadas a un efecto fenotípico. La identificación de las estructuras se realiza mediante bases de datos públicas como Chemical Identifier Resolver (20) y PubChem (21)

Los compuestos químicos se separan en distintos fragmentos (desde 5 hasta 200 fragmentos, siguiendo siempre unas pautas y sin alterar grupos funcionales) (18).

2) Construcción de predictores de fenotipos.

3) Conexión de enfermedades según si se relacionan con estructuras químicas parecidas para la construcción de una red.

4) Creación de redes de patologías con distintas aplicaciones:

- Detección de comorbilidades (relaciones entre patologías).
- Detección de posibles efectos secundarios de fármacos.
- Detección de posibles nuevas indicaciones.

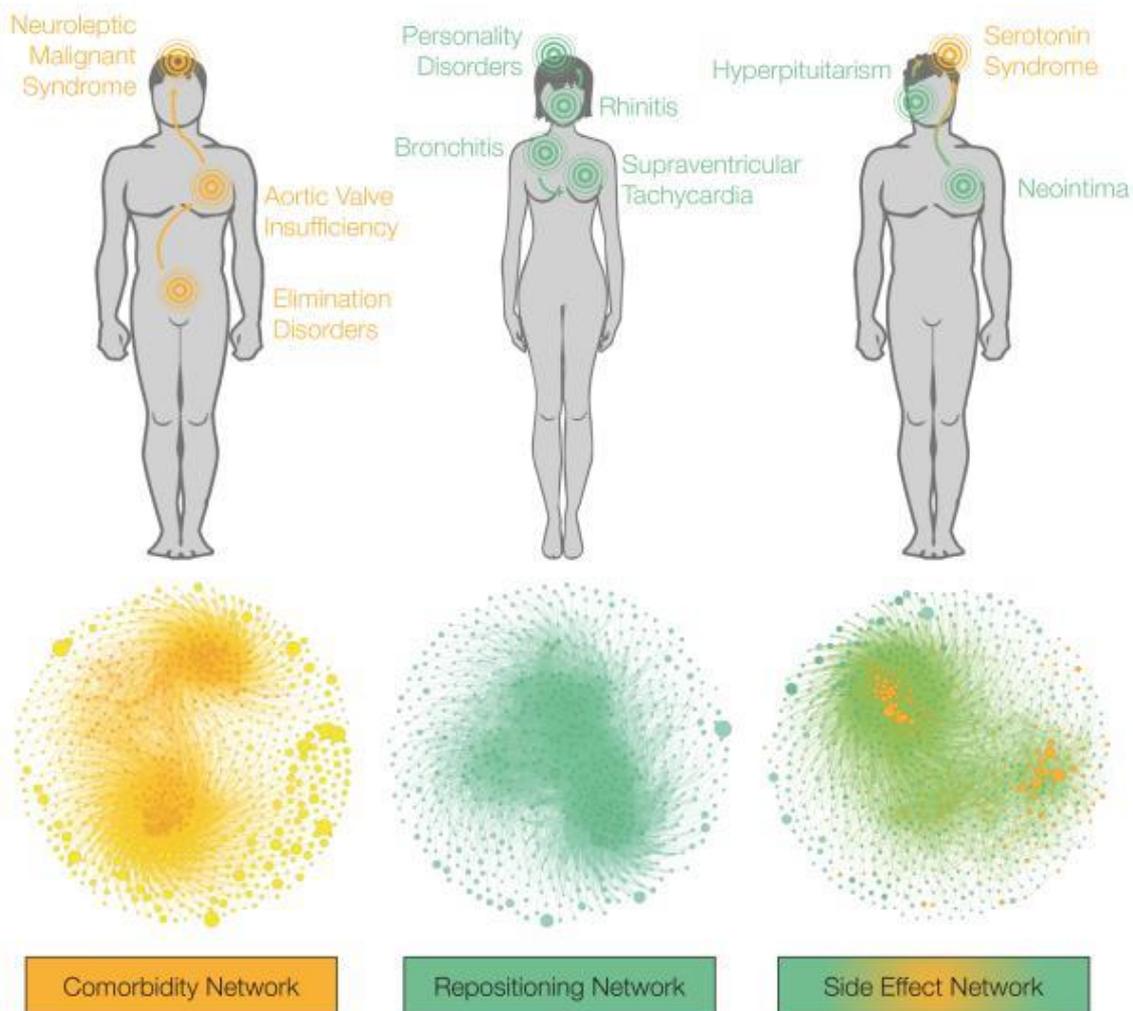


Figura 3. Representación de las distintas redes que pueden diferenciar y correlacionar con las distintas aplicaciones (18).

Modelos predictivos:

Parámetros que tiene en cuenta:

- Phenotypic probability: Probabilidad de generar dicho fenotipo. Se escogen fármacos que sobrepasan un valor de 0.57.
- Target Profile Probability: Probabilidad de interacción con la diana. Se escogen fármacos que sobrepasan un valor de 0.57.
- AUC (Area Under the Curve): Para evaluar el ajuste de los modelos, se miden especificidad y sensibilidad mediante curvas ROC y AUC con los distintos puntos de corte posibles. El valor de mínimo en este caso es de 0.7.

Se utiliza el algoritmo de clasificación RF (Random Forests) para determinar si dos fragmentos interactúan entre sí.

Posibles limitaciones de la tecnología:

- El sistema es ciego al mecanismo molecular, por lo que falta un soporte teórico para discernir por ejemplo entre si el fármaco puede actuar como tratamiento o como toxina.

Los recientes avances en minería de texto (text-mining) son prometedores para paliar esta falta de conexión.

- Es frecuente que la combinación o exclusión de varios fragmentos sea el responsable del efecto farmacológico, por lo que la precisión cuando se analizan fragmentos de forma individual no es suficiente en estos casos.

5.3. Descripción de la tecnología de Anaxomics, TPMS y Sampling Models

1. Construcción de la red biológica

La red biológica o topología consiste en un mapa de interacciones moleculares, generalmente interacciones proteína-proteína, que representan el entorno fisiológico del proceso biológico que se pretende estudiar.

La red molecular se organiza en torno a moléculas semilla, que son las proteínas consideradas como más importantes para el proceso biológico en cuestión. Por ejemplo, las semillas pueden ser una vía molecular, una enfermedad o condición patológica o un fármaco (que pueden ser representados por un conjunto de proteínas).

Una vez se han definido las semillas, se realizará una extensión de la red, añadiendo nuevos nodos (proteínas, aunque podrían ser de otro tipo, como

metabolitos, genes, etc.) y sus relaciones o links (generalmente activación o inhibición, ya sean provenientes de rutas de señalización, asociación funcional, física, etc.).

2. Análisis topológico

Las propiedades funcionales de la red creada son analizadas asegurar que cumple los criterios biológicos establecidos. Para ello, lo primero es transformar la red en un modelo matemático.

Una vez transformado, cada enlace entre nodos es definido como una función con un único parámetro desconocido, de forma que se generan tantas incógnitas como n° de nodos tiene la red. El objetivo final es la resolución de las incógnitas para solucionar el modelo, pero la complejidad del problema en sí exige de computación de alto rendimiento, o en su defecto, de una serie de simplificaciones.

3. Generación del modelo matemático

Los modelos matemáticos permiten la aplicación de la técnica de análisis:

Análisis de presencia (Presence analysis): Se pretende comparar las proteínas de la red con n redes del mismo tamaño escogidas al azar para conocer cuáles de ellas se encuentran más representadas.

Análisis de conglomerados (Clustering analysis): Se trata de una técnica muy utilizada en el campo de la minería de datos. Permite clasificar casos, variables, poblaciones, etc., en función de la similitud existente entre ellos. Existen distintos tipos, como el análisis de conglomerados jerárquico y el de K medias (22), (23). En este caso, el análisis se basa en medidas de densidad de proteínas para poder identificar en el resto de la red regiones con alto n° de proteínas similares. Esta técnica requiere de la aplicación de técnicas de reducción de la dimensionalidad para poder manejar la red problema en un formato 2D.

Técnicas de reducción de dimensionalidad:

Análisis de componentes principales, ACP (en inglés PCA, Principal Component Analysis): Es una técnica estadística que busca la proyección según la cual los datos quedan mejor representados en términos de mínimos cuadrados. A partir de un conjunto de observaciones de numerosas variables (relacionadas o no entre sí), permite reducir la dimensionalidad al escoger las variables llamadas componentes principales, las cuales no muestran correlación lineal y no constituyen una pérdida de información.

En nuestro caso, la transformación por ACP permite convertir la matriz inicial en una nueva versión 2D que representa el mapa o la red de interacciones, localizando juntas aquellas proteínas cuya relación es más parecida.

Análisis de escalado multidimensional (Multidimensional Scaling, MDS):

Se trata de un conjunto de técnicas estadísticas utilizadas para visualizar y explorar datos. En este caso la técnica utilizada en concreto es el algoritmo de Sammon, el cual es utilizado de forma iterativa para reducir el mapa de n dimensiones a un plano bidimensional con una precisión aceptable.

Igual que en el caso del ACP, se mantienen juntas aquellas proteínas que se mostraban cercanas en el modelo inicial.

El modelo matemático pretende describir un sistema biológico con el propósito de generar más información a partir de él con investigaciones indirectas (como es el caso de las predicciones).

En el caso propuesto, los métodos de modelización utilizados son técnicas de Inteligencia Artificial (AI) o aprendizaje automático (machine learning), que incluyen:

- Redes neuronales artificiales (Artificial Neural Networks, ANN) (24).
- Métodos estocásticos como (Simulated Annealing) y el método de Montecarlo.
- Teoría de grafos y reconocimiento de patrones.

Para perfeccionar los resultados se generan múltiples algoritmos matemáticos o modelos (sampling methods). De entre los cuales se escogen los que cumplen un mayor n° de parámetros biológicos, recogidos en la tabla de la verdad (Truth Table), verificado con métodos como la validación cruzada de k interacciones (k -fold cross validation).

4. Mecanismo de acción (MoA):

La forma más habitual y práctica de reportar éstos modelos o resultados es generar una gráfica denominada mecanismo de acción, en la que se representa la vía molecular capaz de producir la respuesta fisiológica correspondiente a la patología o input introducido. Asimismo, es complementario con otros análisis (tales como el reposicionamiento), ya que permite proveer los resultados de una explicación molecular. Lo que permite, por un lado, poder comprender el proceso y por otro corroborar los resultados en caso de tener un sentido biológico.

A continuación, se detalla el proceso de generación de un mecanismo de acción.

Proceso de realización de un MoA:

La composición de un MoA es un proceso complejo que se puede dividir en 3 capas (topológica, priors y restricciones), en función de la información biológica, clínica y topológica que se maneja.

Topología:

Definimos topología como el conjunto de proteínas y/o genes y los links que relacionan unos con otros.

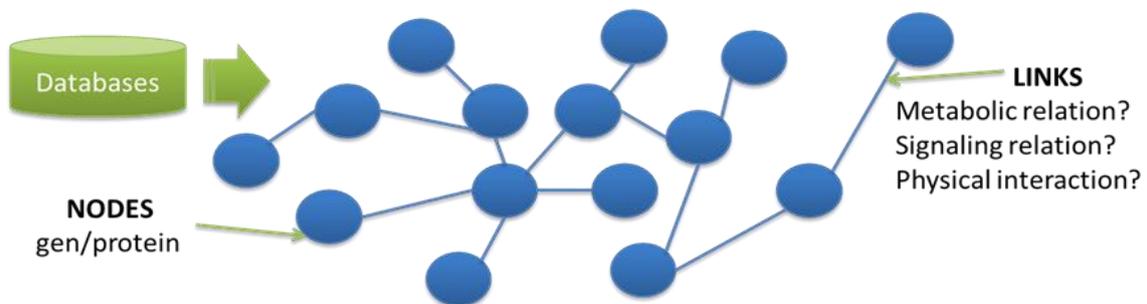


Figura 4. Representación esquemática de una topología.

Se inicia el proceso a partir de una topología general de grandes dimensiones que servirá de base para la posterior modelización. En Anaxomics existen conformados distintos tipos de topologías que se corresponden con diferentes tipos celulares. Es decir, se contemplan parte de las diferencias que se derivan de cada tipo celular, ya sea por influencias debidas a la fisiología propia de la célula o por su entorno o ubicación física. De algunos de estos tipos de topología, existen dos versiones, denominadas como relevante y no relevante, según tengan mejor o peor calidad (a nivel de referencias), respectivamente. Esto permite analizar a dos niveles de profundidad, siendo recomendado usar el primero por defecto y recorrer al segundo en caso de necesitar más datos para crear el modelo.

Cada una de las topologías se conforma de una gran cantidad de relaciones proteína-proteína con el siguiente formato:

ID proteína 1	ID proteína 2	Interacción
---------------	---------------	-------------

La interacción positiva o activación se representa con un valor +1, mientras que la negativa o inhibición con un valor -1. Contrariamente, una interacción nula o desconocida se representa con un valor igual a 0.

Para el presente trabajo se ha utilizado la topología aplicada usualmente, la llamada *Global Human Cell* (GHC). De forma opcional, hubiera podido usarse una población específica de células epiteliales. Sin embargo, se ha priorizado el hecho de que la variedad GHC es la más universal y versátil de la colección, por tanto, la mejor entrenada y validada por años de uso.

Priors

La segunda capa añade toda la información de las bases de datos públicas implementadas en Anaxomics, tales como:

- KEGG (Kyoto Encyclopedia of Genes and Genomes) (25) y REACTOME (26) → vías metabólicas.
- DrugBank (17) → fármacos.
- GEO (Gene Expression Omnibus) (27) → Expresión de proteínas.
- IntAct (Molecular Interaction Database) (28), BioGRID (29) y MINT Molecular INTeraction database) (30) → Interacciones entre proteínas.

Este conjunto de datos conformaría lo que llamamos como “*prior restrictions*”, el cual consiste en información general de distinta índole que se añade para cada mecanismo de acción de forma automática elevando su robustez.

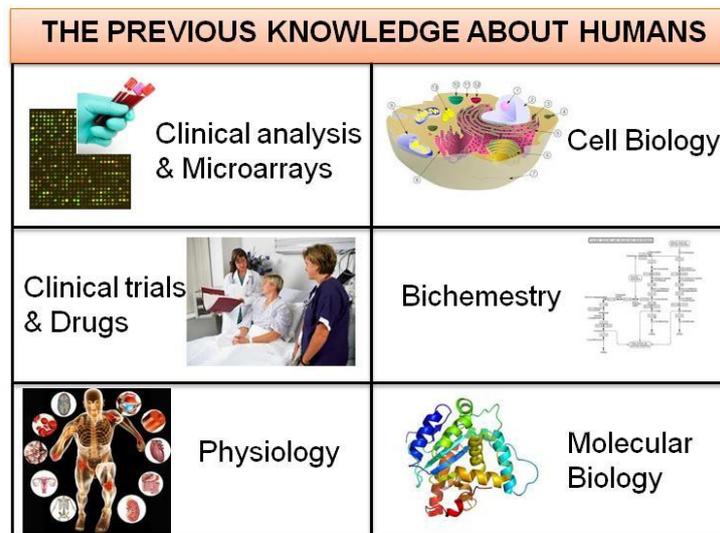


Figura 5. Representación gráfica de los diferentes tipos de conocimiento incluidos como prior restrictions.

Restricciones aplicadas

- Opcionales

Una tercera capa la conforman las modificaciones o restricciones provistas por el usuario. Esta información es opcional y puede conformarse por datos clínicos evaluados, procesos fisiológicos que caracterizan una población a estudiar, datos moleculares experimentales y datos de alto rendimiento (High-Throughput

data o HT), etc. En definitiva, todos los argumentos que puedan ser relevantes para el modelo analizado.

Tales como resultados de:

- Espectrometría de masas
- Microarrays
- Secuenciación de nueva generación (Next Generation Sequencing o NGS)
- Metabolómica
 - Requeridas: Input y output del modelo deseado

Se pueden considerar como restricciones “especiales”, las cuales son indispensables, y definen la entrada y la salida del MoA junto con el tipo de relación que las ha de enlazar. La entrada se ilustrará como la proteína o conjunto de proteínas a raíz del cual se inicia el gráfico y cuyas interacciones emergen hacia la salida o diana de la vía metabólica.

Por ejemplo, tal como se puede ver en este trabajo, la entrada puede ser un fármaco (conjunto de proteínas target), y la salida o respuesta, una patología (conjunto de proteínas efectoras). En caso de que el modelo pretenda simular el mecanismo de acción de un fármaco para curar la enfermedad se indicaría con un signo negativo y en caso contrario, promover la enfermedad, comportándose como un efecto adverso, signo positivo.

Construcción de poblaciones y mecanismo de acción por pasos:

Paso 1: Entrenamiento del modelo GHC

A partir de la topología y junto con parte de la información proveniente de bases de datos públicas (como KEGG y REACTOME), se generan muchísimas poblaciones o posibles soluciones. No hay establecido un número de soluciones fijo, el sistema decide cuándo se ha alcanzado la calidad deseada, y se eligen de entre ellas las 1.000 mejores.

El entrenamiento del modelo consiste en buscar valores que deben tener los links de los cuales no se conoce la relación que une 2 proteínas. Cada nodo se comporta como un perceptrón en el que entran señales (inputs) y salen señales hacia sus proteínas vecinas similar a los perceptrones de las redes neuronales. Las señales de entrada a cada nodo se integran y se someten a una función sigmoidea que normaliza la señal a valores entre -1 y 1 que se multiplica por el valor de cada link. De este modo los valores se determinan mediante un proceso iterativo y supervisado que va optimizándose mediante un proceso basado en el descenso de gradiente.

El proceso es supervisado por la Tabla de la Verdad o *Truth Table* (TT) que contiene todas las restricciones que debe cumplir el modelo, de este modo, la calidad del modelo (*accuracy* o *fitness*) puede medirse contra esa tabla de la verdad.

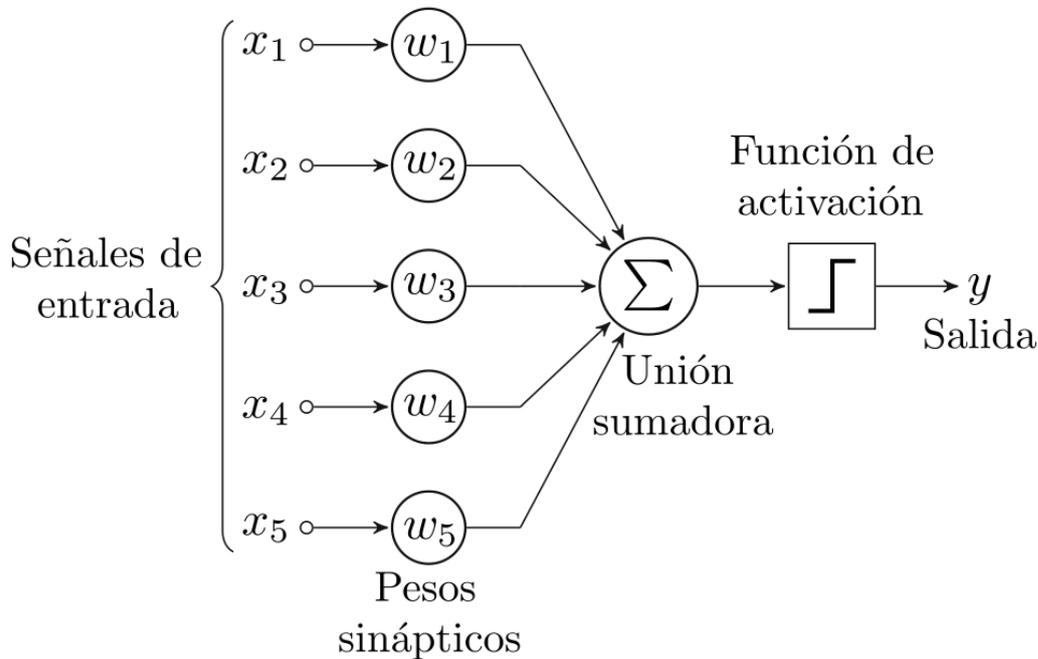


Figura 6. Diagrama de un perceptrón (31).

Generación de modelos matemáticos

Paso 2: Selección del subespacio topológico de interés

Una vez se han definido los puntos anteriores, a partir de las 1000 mejores soluciones de GHC, añadiendo la información proveniente de los priors, las restricciones deseadas, la entrada y la salida (con su signo), se crean 100 soluciones con una topología mucho más reducida. Son escogidas tan solo interacciones que se encuentran a una distancia máxima de 3 nodos. En este punto, el modelo es también contrastado con la Tabla de la Verdad, una base de datos confeccionada a mano en Anaxomics que contiene evidencias biológicas bien contrastadas conocidas que todo humano *a priori* ha de cumplir.

El proceso tiene lugar de la siguiente forma:

Las restricciones deseadas filtran el modelo para eliminar o añadir interacciones que no nos interesan.

Se determina para cada link un valor probabilístico (*probability prior* o PP), el cual, determina la probabilidad de que la interacción ocurra en ese sentido y de esa forma (+/-).

El PP se calcula en base a la información disponible e introducida (restricciones del usuario, restricciones prior y tablas de la verdad), teniendo en cuenta distintos aspectos:

1. Estabilidad inicial definida por expresión de proteínas, valor que penaliza interacciones que producen la disrupción del estado inicial.
2. Concordancia de las interacciones con coexpresiones de HT.
3. Correlación entre signos de proteínas y vías de interacción.
4. Correlación entre signos de dianas de fármacos y la TT.
5. Objetos clínicos adicionales activados como respuesta al estímulo, mide las vías de interacción que van desde el estímulo hasta medidas de objetos clínicos adicionales. Las rutas metabólicas del MoA coherentes con estos objetos clínicos adicionales se consideran más probables.
6. Variaciones de la PP pueden deberse también a las correlaciones entre las vías y la TT extra de respuesta a estímulo indicada por el usuario.
7. Las dimensiones de la ruta metabólica también modifica el valor de probabilidad.

Una vez se han calculado los valores PP son seleccionadas aquellas 3.000 interacciones que presentan mayor probabilidad. Se eliminan de las muestras de GHC aquellas interacciones que NO forman parte de la selección.

Paso 3: Optimización

Finalmente, se aplica un proceso aleatorio de optimización a las 100 funciones de los modelos resultantes. Basándose en el valor de PP se añaden y eliminan interacciones construyendo una versión resumida de las interacciones más frecuentes, para encontrar la combinación o solución óptima para reportar como resultado. La relevancia de las interacciones es determinada basándose en su frecuencia de ocurrencia (votos) sobre la población original. Dicho de otro modo, la solución escoge las interacciones más relevantes y las que mejor se ajustan a la TT, a las entradas definidas (estímulo) y a las restricciones extra introducidas por el usuario. Adicionalmente, el número de nodos y links que se utilizaran para conformar la representación gráfica es ajustado y adaptada para un tamaño N deseado que permita simplificar la ruta metabólica.

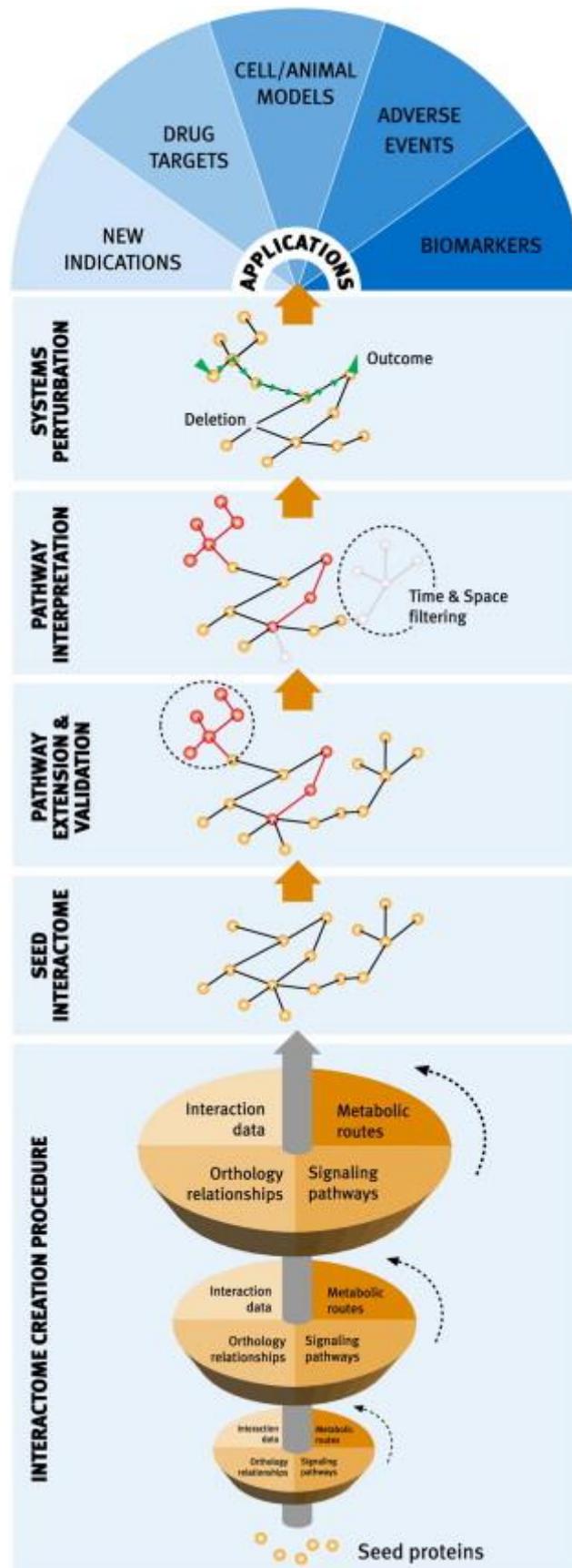


Figura 7. Representación esquema del proceso de elaboración de un MoA (32).

5.4. Elección de fármacos para estudiar su mecanismo de acción

A partir de los fármacos candidatos provenientes del reposicionamiento se han escogido dos compuestos para continuar el análisis mediante el estudio de su mecanismo de acción:

Entre los tipos de fármaco que típicamente tienen como efecto secundario la promoción del acné son las hormonas.

1. Con esto, he escogido el fármaco Somatropina recombinante (hormona del crecimiento). El cual es interesante, dada la relación epidemiológica del acné con la pubertad (33).
2. Además, he escogido el fármaco ácido Meclofenámico, un antiinflamatorio no esteroideo (AINE), utilizado para el tratamiento del dolor leve o moderado. Implicado además en la inhibición de la síntesis de prostaglandinas.

Dado que la inflamación juega un papel proactivo en la fisiopatología del acné, se podría pensar que la relación del fármaco con la patología podría darse en cualquiera de las dos direcciones (34)².

ID de DrugBank	Nombre fármaco	Dianas posibles ID Uniprot Gen (Nombre)	Efecto sobre la diana
DB00052	Somatropina recombinante	P10912 GHR (receptor hormona del crecimiento)	Activa
		P16471 PRLR (receptor de Prolactina)	Activa
DB00939	Ácido Meclofenámico	P23219 PTGS1 (prostaglandina G/H sintasa 1)	Inhibe
		P35354 PTGS2 (prostaglandina G/H sintasa 2)	Inhibe
		P09917 ALOX5 (Araquidonato 5-lipoxigenasa)	Inhibe
		O43526 KCNQ2	Desconocido
		O43525 KCNQ3	Desconocido

Tabla 2. Tabla con información procedente de DrugBank los fármacos a estudiar. Las dianas para las que el efecto es desconocido no se introducen como parámetros en la generación del modelo.

² En el apartado 11. Anexos, en la parte de scripts, puede observarse que se había elegido el fármaco Finasteride para su estudio. Sin embargo, el resultado del MoA era muy complejo y presentaba interacciones muy dudosas, por esto se decidió optar por otro compuesto.

Además, para un estudio más exhaustivo de los resultados para ambos fármacos, se han recuperado los valores de relación de los nodos para los 100 individuos creados a partir del modelo inicial. Obteniendo así dos matrices con x variables y n igual a 100 para comparar estadísticamente. De esta forma se podrá determinar si existen diferencias significativas entre los mecanismos acción de ambos fármacos.

La estadística se ha llevado a cabo con una herramienta que he ayudado a poner a punto. Se trata de un software implementado en Matlab adaptado para distintos tipos de datos y realiza de forma automática distintos test estadísticos generando un informe con los resultados en PDF. Para utilizarlo el usuario ha de introducir los datos mediante un Excel, archivo de texto o matriz de datos y definir el tipo de análisis que se ha de aplicar según los datos, en función si hay 1 o n variables distintas, si son datos pareados o no, etc. Finalmente se deben escoger los resultados más adecuados a reportar dado que se presentan distintos tipos de test y gráficos.

6. Resultados

6.1. Resultados del método basado en subestructuras de compuestos químicos

No dispongo de estos resultados de modo desagregado, por lo tanto no son presentados aquí, si no, en el siguiente apartado (6.3. Comparación de ambos métodos).

6.2. Resultados del método TMPS

6.2.1. Reprofiling: Detección de fármacos candidatos

Nombre fármaco	Valores de predicción ANN (Artificial Neural Network)				
	Global	1	2	3	4
Somatropina recombinante	76,66%	15,79%	75,6%	5,92%	82,06%
Ácido Meclofenámico	76,31%	5,39%	93,4%	10,06%	36,2%

Tabla 3. Tabla de resultados del reposicionamiento con el método TPMS mostrando únicamente los fármacos elegidos para continuar el estudio.

Se muestran en color verde los valores superiores al 75% que se corresponden con elevada correlación. Los motivos del 1-4 se corresponden con: hiperqueratinización folicular, alteración de la producción y composición del sebo, infección por P. acnes e inflamación.

Para una interpretación de los valores reportados se muestran las figuras siguientes (Figura 9 y Figura 8). En primer lugar los valores reportados por el análisis de reposicionamiento se expresa como porcentaje de ANN. Dicho porcentaje se correlaciona con un valor de TPMS tal y como se observa en la figura. En la siguiente gráfica se muestra el gráfico de densidad de Kernel que representa la probabilidad de ser candidato tanto si la relación es positiva como si es negativa. Los cuales se representan en rosa como indicación y azul como no indicación.

La línea naranja muestra el punto de corte (valor de TPMS de 0.5) a partir del cual se puede asegurar una elevada correlación entre el fármaco y la indicación.

Por tanto, si traducimos el valor de TPMS a probabilidad de ANN, obtenemos un threshold de casi 0.75 (75%) aproximadamente.

Los resultados de la reprogramación muestran elevada relación con la patología global para el fármaco Somatropina recombinante. Si miramos con más detalle, podemos ver que es debido principalmente al mecanismo de la inflamación, pero también a una alteración de la producción y composición del sebo en gran medida.

De igual manera, se observa elevada correlación del segundo fármaco con la patología estudiada. Sin embargo, tan solo es debido al motivo alteración de la producción y composición del sebo.

Cabe destacar que esta relación no indica si es en calidad de tratamiento o de efecto adverso, dilucidar esto forma parte de la tarea del usuario, quien debe evidenciar mediante referencias bibliográficas u otros métodos qué es lo más probable.

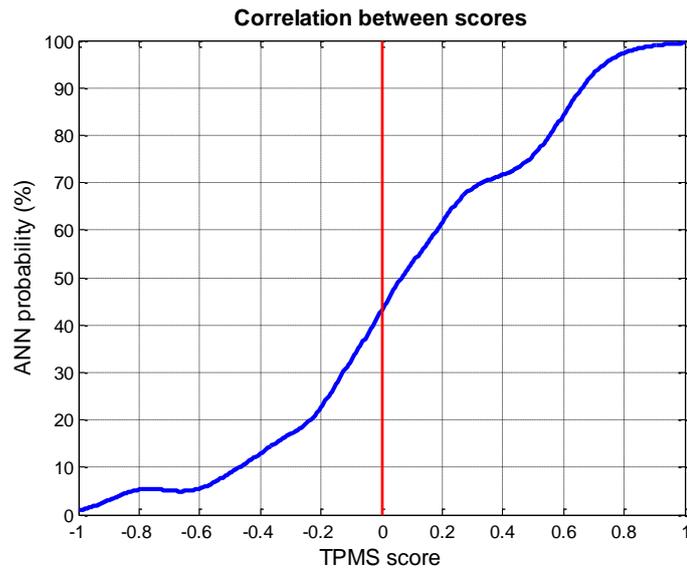


Figura 8. Equivalencias para valores de TPMS y valores de probabilidad ANN.

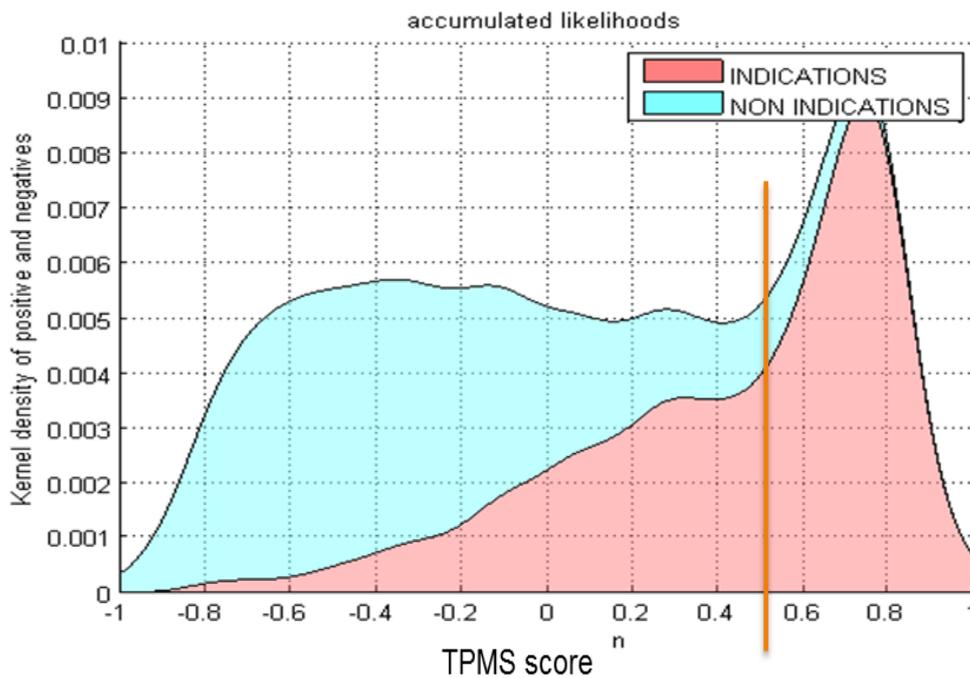


Figura 9. Función de densidad de probabilidad acumulada para ambas categorías en función de su valor de TPMS.

Se observa que a mayor valor de TPMS, menor es la probabilidad de error al considerar un candidato como indicación (35).

6.2.2. Identificación del mecanismo de acción

MoA inicial para el fármaco Somatropina recombinante (pre-validación)

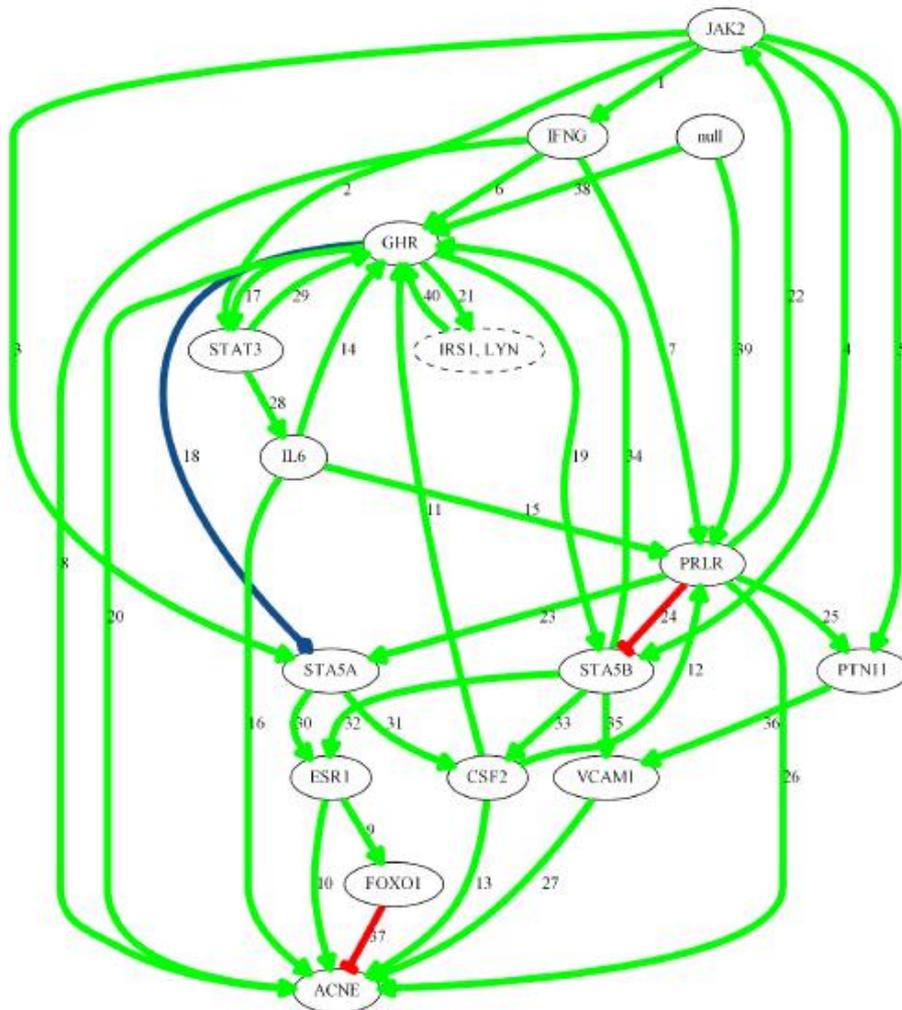


Figura 10. Mecanismo de acción del fármaco Somatropina recombinante para producir acné (pre-validación)

Se muestra en la Figura 11 la leyenda para ayudar a la comprensión del MoA. La lectura se realiza siempre de arriba abajo, considerando interacción por separado. Cabe recordar que el hecho de que dos proteínas se encuentren unidas por un link, no significa que tengan una interacción física o química de forma directa.

Una vez realizados los primeros experimentos no se obtienen resultados de óptima calidad, dado que muchos de los links reportados como solución no habían corroborados por publicaciones científicas en la base de datos de Anaxomics. Por tanto, han sido revisados los links a través de referencias bibliográficas y se han introducido como restricciones para repetir el experimento. Se denotar visualmente en la Figura 10 y la Figura 13, y, por su

parte, la Figura 13 y la Figura 14, que las interacciones entre proteínas de los mecanismos de acción han sido modificadas.



Figura 11. Leyenda para la lectura de links y nodos de los mecanismos de acción.

MoA para el fármaco Somatropina recombinante (post-validación)

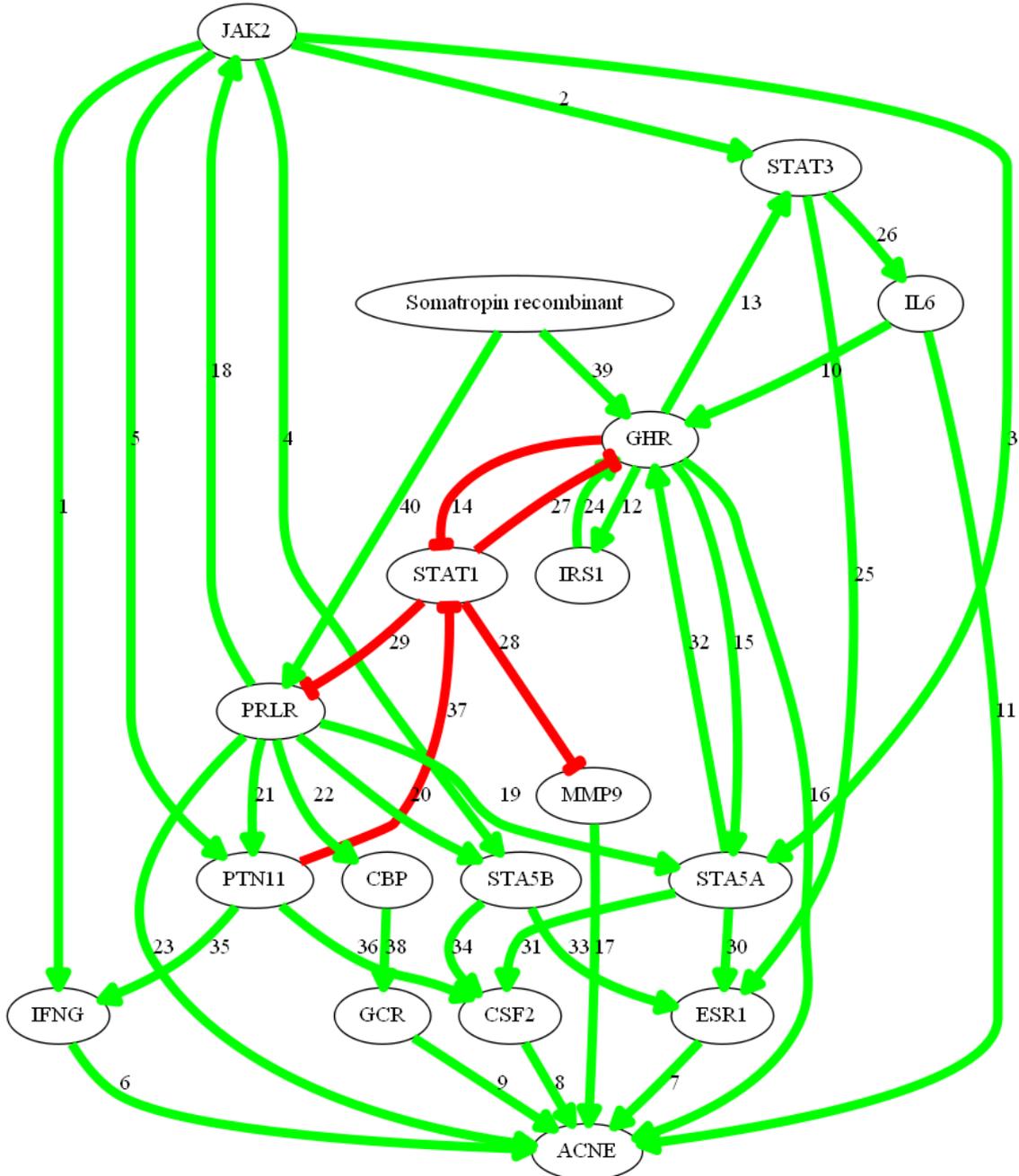


Figura 12. Mecanismo de acción del fármaco Somatropina recombinante para producir acné (post-validación).

MoA inicial para el fármaco ácido Meclofenámico (pre-validación)

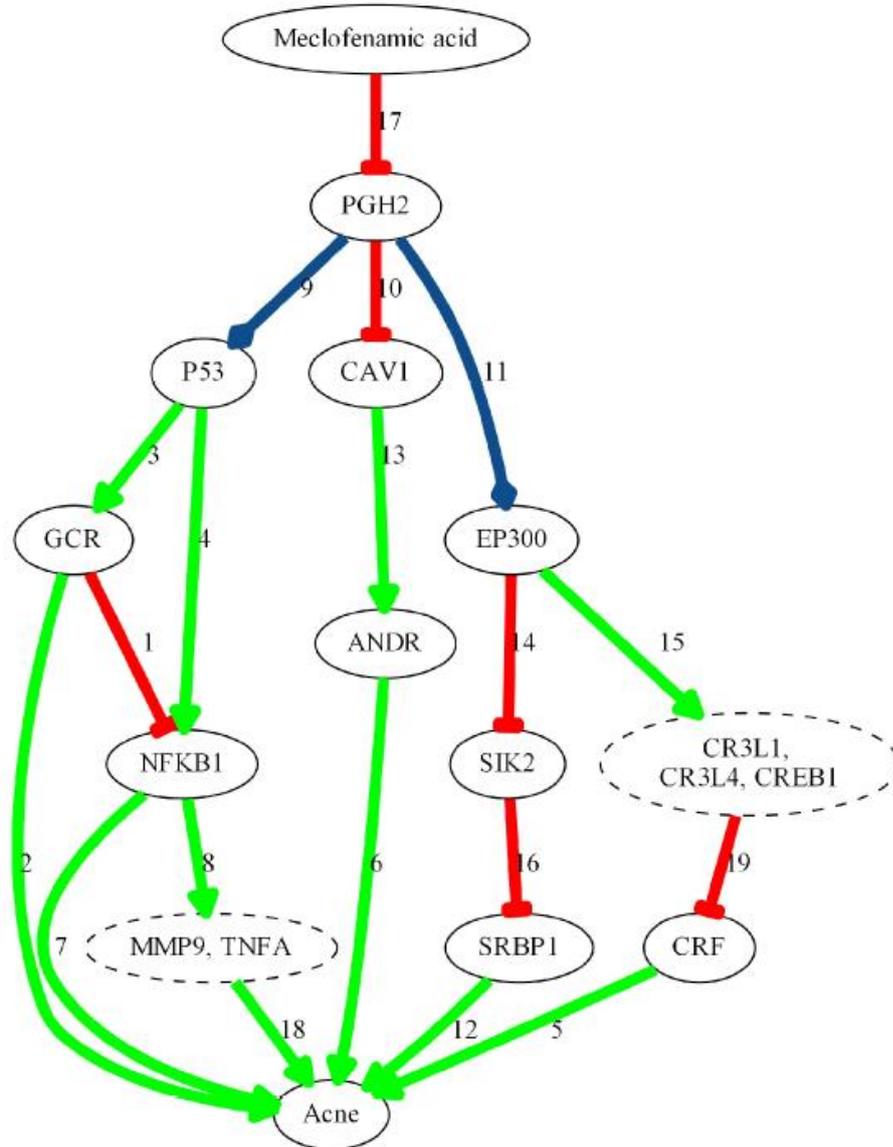


Figura 13. Mecanismo de acción del fármaco ácido Meclofenámico para producir acné (pre-validación).

MoA para el fármaco ácido Meclofenámico (post-validación)

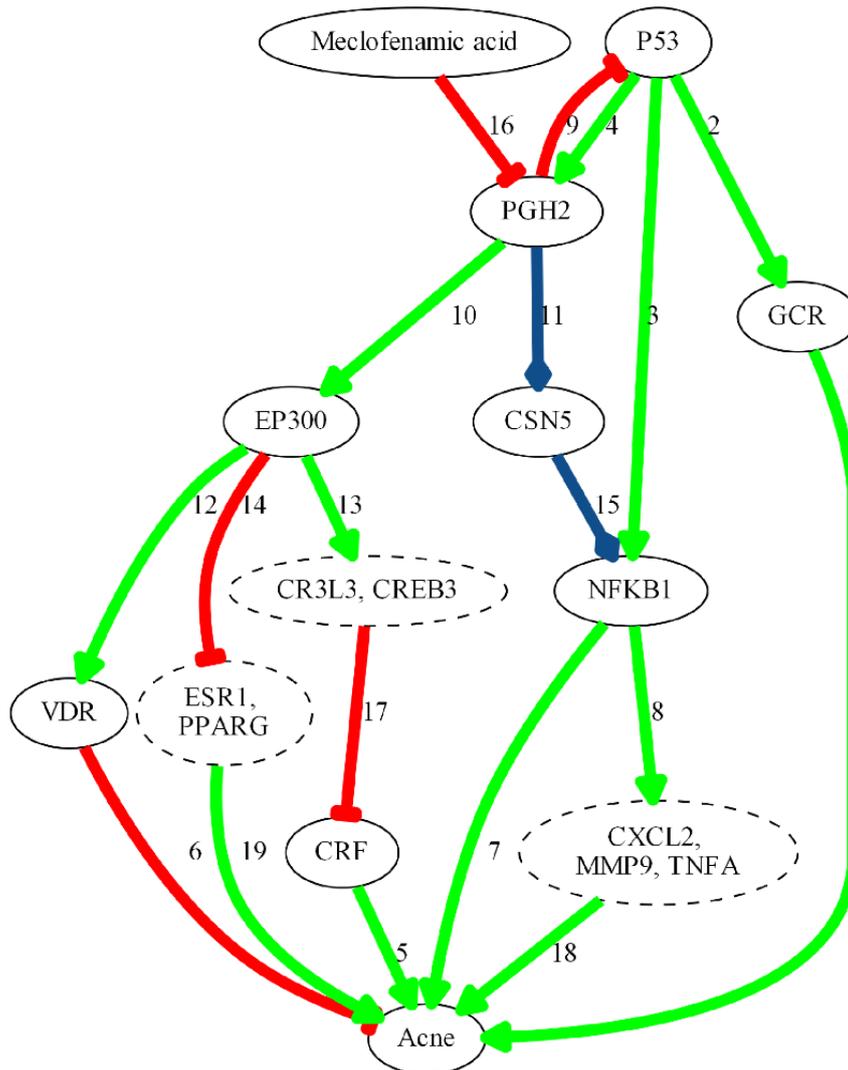


Figura 14. Mecanismo de acción del fármaco ácido Meclofenámico para producir acné (post-validación).

Para el modelo TPMS se parte aproximadamente de un total de 4.400 proteínas. En la Tabla 4 se indican en número total de proteínas con que se realiza finalmente cada modelo.

De los 79 efectores del acné indicados, 13 de ellos no se han podido tener en cuenta para realizar el modelo por no contener información suficiente, es decir, por falta de interacciones suficientes.

	Somatropina recombinante	Ácido Meclofenámico
Nº de proteínas modeladas por el estímulo (topología pequeña)	1395	853
Precisión (Accuracy)	94.83 %	94.73 %
Cobertura (Coverage)	100.00 %	100.00 %
Incerteza (Uncertainty)	28.29 %	16.54 %
Nº restricciones clínicas computadas	1002552	1002552
Nº de restricciones biológicas computadas	14973	14973
Nº de restricciones introducidas por el usuario	928	928

Tabla 4. Detalle de los parámetros de la construcción de los mecanismos de acción para ambos fármacos.

Para una interpretación biológica de los resultados de los mecanismos de acción se han rastreado las funciones las proteínas incluidas como nodos y se han clasificado de una forma general.

- Fármaco Somatropina recombinante:

En general la mayoría de proteínas integradas en el MoA se relacionan con regulación de dos tipos de procesos, los implicados en proliferación, desarrollo y diferenciación celular; y los relacionados con el sistema inmune e inflamación.

- Fármaco ácido Meclofenámico:

En general las proteínas integradas en el MoA se relacionan con regulación génica a través de factores de transcripción, implicados también en proliferación, desarrollo y diferenciación celular, así como con inflamación.

Esto, a priori, no nos revela gran información si no se estudia en profundidad, dado que se definen funciones muy generales. El punto que sí que llama la atención es que ninguno de los 2 mecanismos de acción comparte ningún nodo, lo que nos lleva a pensar que cada uno de los fármacos está actuando por una vía completamente distinta.

En el apartado 11. Anexos, se incluyen las tablas que contienen las proteínas incluidas como nodos en los mecanismos de acción.

6.2.3. Resultados procedentes del estudio estadístico:

A partir de las 100 poblaciones generadas para cada fármaco se han llevado a cabo un análisis descriptivo (evaluando *kurtosis*, medida general de *outliers* y *skewness*, simetría de los datos) y evaluado distintos test de hipótesis, considerando los datos como pareados.

Ha sido descartada la distribución gaussiana de cada una de las variables mediante 2 test estadísticos diferentes: K-S test (Kolmogórov-Smirnov) y Chi2 test (χ^2 o “chi cuadrado”). Por tanto, se aplica el test no paramétrico de Wilcoxon.

De las 68 proteínas efectoras que finalmente han participado en la generación del MoA, se utilizan 56 como variables, el resto se eliminan dado de no contienen información relevante (todos los valores son igual a cero).

La aplicación de una corrección para test múltiples se recomienda por algunos autores en caso de analizar un número de test de hipótesis mayor a 20; n para la cual la probabilidad de error tipo I (falsos positivo) es de un 60% (36).

Dado que en este caso se realizan 56 test de hipótesis, se considera por tanto, como un análisis de comparaciones múltiples. Se tiene en cuenta el p-valor ajustado por FDR en lugar del p-valor para decidir entre las hipótesis nula o alternativa.

A continuación, se listan los resultados del estudio estadístico. Únicamente se muestran p-valor y p-valor ajustado para el test no paramétrico de Wilcoxon.

Nombre de la variable	p-valor (paired Wilcoxon 2 tail)	FDR Wilcoxon
O00206	0,61408964	0,636833703
O60603	0,10307302	0,137430688
O95864	0,00713119	0,012101408
P01137	0,38824955	0,452957806
P01375	2,3458E-05	5,25467E-05
P01579	5,4956E-09	1,81032E-08
P01583	0,00130107	0,002350312
P01584	0,01663915	0,025883124
P01903	0,08568129	0,117028099
P03372	0,00039274	0,000785486
11	0,41326299	0,472300557
12	1,9738E-13	1,22817E-12
13	1,6625E-09	6,20648E-09
14	5,9718E-13	3,34423E-12
15	1,2917E-09	5,16678E-09
16	9,5916E-10	4,13177E-09
17	8,1003E-08	2,38745E-07

18	2,9173E-15	2,67834E-14
19	0,00048828	0,000942888
20	0,00910587	0,014997909
21	0,07425851	0,103961912
22	0,54808226	0,590242429
23	9,286E-09	2,88897E-08
24	2,0802E-17	8,85392E-16
25	7,3102E-16	1,02342E-14
26	0,25	0,304347826
27	0,36971653	0,440513313
28	3,3479E-15	2,67834E-14
29	4,9045E-06	1,14439E-05
30	0,46772205	0,523848696
31	3,1233E-09	1,09316E-08
32	0,13678453	0,170220743
33	0,00034452	0,000714562
34	0,00526085	0,009206493
35	0,00056261	0,001050198
36	1,1037E-15	1,23619E-14
37	9,0281E-07	2,52788E-06
38	0,90381963	0,903819628
39	4,2169E-17	8,85392E-16
40	0,87935559	0,895343877
41	0,12074515	0,157249497
42	0,50007052	0,549097046
43	1,0298E-06	2,7461E-06
44	4,7432E-17	8,85392E-16
45	1,1673E-10	5,94276E-10
46	1,8862E-13	1,22817E-12
47	0,000284	0,000611683
48	3,2534E-06	8,15495E-06
49	6,9004E-10	3,2202E-09
50	3,3494E-06	8,15495E-06
51	0,0625	0,092105263
52	0,125	0,159090909
53	0,60338994	0,636833703
54	0,07338568	0,103961912
55	0,01953125	0,029560811
56	0,01452637	0,023242188

Tabla 5. Resultados del test de hipótesis de Wilcoxon.

Se muestran en verde los valores de FDR o p-valor ajustado menores al valor de significación alfa 0.05³.

³ Por motivos de confidencialidad no se muestran todos los identificadores de proteína en la tabla y se han sustituido por números.

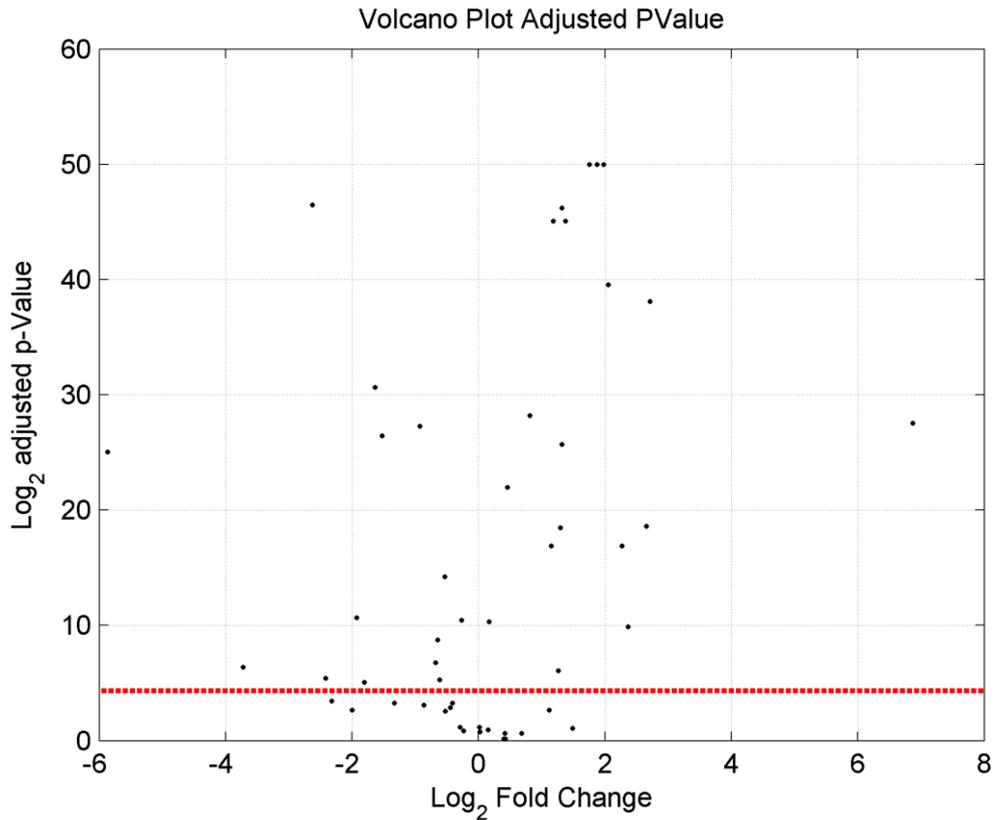


Figura 15. Gráfico Volcano plot que representa el Log₂ del Fold Change contra el Log₂ del p-valor ajustado para el test de Wilcoxon. Se marca en rojo una línea punteada que separa diferencias consideradas como significativas o no (zona superior e inferior respectivamente).

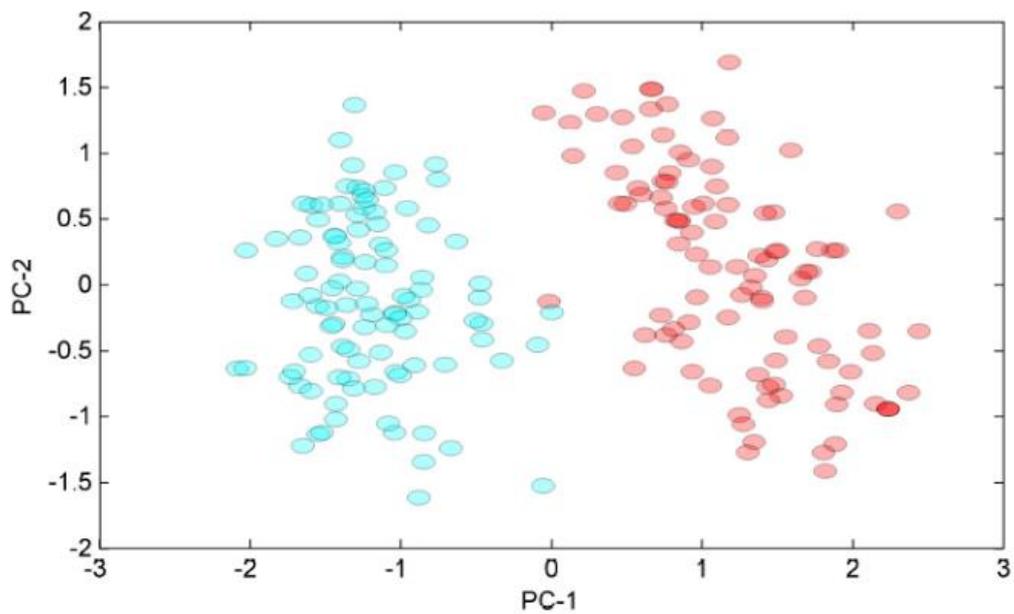


Figura 16. Gráfico resultado del Análisis de Componentes Principales (ACP) para los componentes uno y dos. Se muestra en azul y rojo la población perteneciente al fármaco Somatropina recombinante y ácido Meclofenámico.

Como puede observarse, de un total de 56 proteínas, 37 de ellas muestran diferencias significativas para los valores de activación de las dos poblaciones. Al contrario, para las 19 restantes no podemos descartar que no sean iguales. Se observa en la Figura 15 cómo para la mayor parte de las variables los tratamientos están diferenciados, dado que se encuentran en la parte superior de la gráfica. El valor del eje x en cambio no es muy informativo, debido probablemente a un número de variables insuficiente para observar un patrón claro.

En la Figura 16 se observa cómo se dividen de forma evidente las variables de cada tratamiento en función del componente principal 1.

6.3. Comparación de ambos métodos

Nº de fármacos detectados como relacionados	Método TPMS	Método IRB
Compartidos	33	
Distintos	279	105
Total	316	138

Tabla 6. Comparación de resultados de *reprofiling* para la patología *acné vulgaris*.

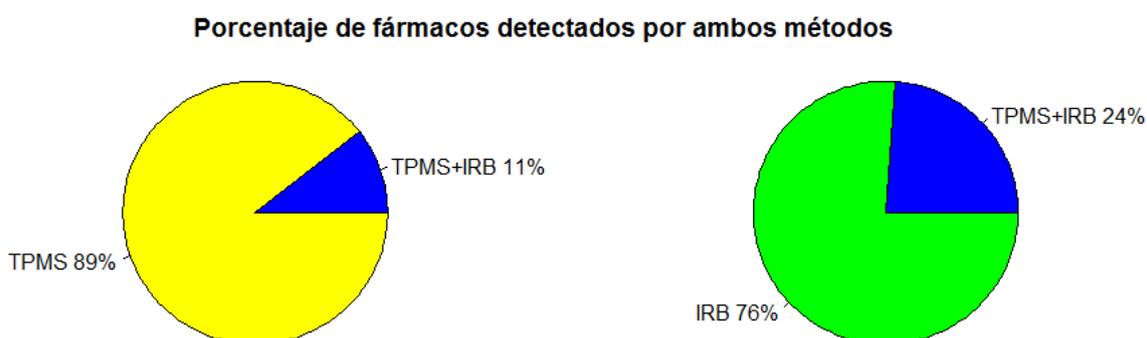


Figura 17. Gráficos de 360º correspondientes al porcentaje de fármacos que han sido detectados y por ambos métodos y por un único método (TPMS, a la izquierda y método quemo-céntrico desarrollado por el equipo del IRB Barcelona, a la derecha).

7. Análisis de resultados y conclusiones

7.1. Análisis de reposicionamiento

7.1.1. Método basado en segmentación de compuestos

- Muchos compuestos químicos con actividad biológica conocida están compuestos por subestructuras químicas comunes con otros compuestos químicos.
- Existen múltiples bases de datos de compuestos, tales como DrugBank (17) o PubChem (21), donde están documentadas las estructuras químicas de compuestos con actividad biológica conocida. Estas bases de datos describen además otra información relevante del compuesto como la indicación farmacológica del compuesto, sus efectos adversos o las proteínas dianas del compuesto.
- El Prof. P. Aloy y su equipo han diseñado una estrategia que permite obtener una división de las subestructuras del compuesto químico de tal manera que se puede establecer una relación entre esas subestructuras químicas y la actividad biológica del compuesto. Además, dado que las proteínas diana de cada compuesto son también conocidas es también posible establecer una relación estadística entre las subestructuras que lo componen y las dianas del compuesto.
- Ambas fuentes de información, la biológica-funcional y la información sobre sus dianas, puede ser agrupada con el fin de establecer potenciales estadísticos que relacionen ambos tipos de datos.
- El número de compuestos distintos que interactúan con una misma proteína diana es limitado, lo que limita la capacidad de sacar conclusiones con potencial estadístico. Por ese motivo es necesario establecer relaciones entre el conjunto de proteínas diana conocida de manera que nos permita realizar agrupaciones (clústeres) de esas proteínas diana.
- Existen multitud de formas de relacionar las proteínas diana entre ellas. Una de las formas convencionales es establecer una relación topológica entre esas proteínas. La relación topológica más directa que puede

establecerse entre dos proteínas es el mínimo número de proteínas necesarias para conectar esas dos proteínas diana.

- No obstante, la forma de medir la relación topológica entre dos proteínas no es trivial. Existen multitud de formas de construir el mapa de relaciones de proteínas y genes. Las bases de datos que tienen la información para hacer esa tarea son diversas pero ruidosas. Existen multitud de relaciones falsas positivas y falsas negativas que año tras año van estableciéndose con más calidad.
- En términos genéricos es posible establecer una relación estadística entre todos los elementos descritos hasta el momento: la actividad biológica, la subestructura química del compuesto y las proteínas diana conocidas de ese compuesto. La explotación de las relaciones de esos potenciales estadísticos mediante técnicas de conglomerados (*clustering*) y clasificación por procedimientos de aprendizaje supervisado permite extraer conclusiones biológicas.
- Dado un compuesto problema es posible descomponerlo en subestructuras químicas. Algunas de esas subestructuras pueden tener estadísticamente una relación funcional que podría heredar el compuesto original.
- Los resultados que describen los artículos del Prof. Aloy y su grupo son muy interesantes y prometedores.
- Sin embargo, esta aproximación presenta también muchas limitaciones:
 - Para poder obtener información de un compuesto problema es necesario que las subestructuras químicas del compuesto hayan sido contenidas en otros compuestos de funcionalidad biológica conocida.
 - Como todo análisis estadístico las conclusiones a las que se pueden llegar dependen del conjunto de datos de aprendizaje.
 - Las conclusiones a las que se puede llegar podrían basarse en premisas no ciertas que no se dan siempre en la naturaleza. Por ejemplo, el fármaco problema podría nunca alcanzar la proteína diana ya sea por barreras físicas, químicas o por una ubicación distinta. Además, lo podemos sumar otros aspectos complejos como es el estado de administración del fármaco, en algunos casos se administra un profármaco que necesita de una bioactivación en el organismo, por tanto, este paso modifica la molécula en sí y no siempre es un proceso conocido.

- La relación topológica entre las proteínas diana se establece como se ha comentado a partir de bases de datos ruidosas. Esta medida por tanto está sujeta a un error que debe ser manejado y controlado.
- El proceso de fragmentación que se lleva a cabo también es determinante en el resultado de la comparativa con los fragmentos referenciados. Además, influye el hecho de que muchos de los efectores de un fármaco pueden ser producidos por congregaciones de distintos fragmentos, lo cual genera falsos negativos.
- Al tratarse de un método de aprendizaje supervisado (al igual que el TPMS), se tiende a ponderar los aspectos más representados de los modelos, influyendo ligeramente así en los resultados.

7.1.2. Método TPMS

- El procedimiento de Anaxomics, llamado TPMS, parte de premisas completamente diferentes pese a que persigue exactamente los mismos objetivos.
- El método TPMS se construye a partir de una topología, siendo capaz de encontrar las respuestas siempre que lo permita dicha topología, es decir, han de haber interacciones suficientes para que el efector llegue a interactuar con la diana en cuestión. Además, necesita que el modelo se ajuste con la precisión mínima indicada. En caso de no cumplir estos criterios el modelo se diagnosticaría como no válido y no se construiría.
- Existen bases de datos que contienen información topológica donde se encuentran documentadas las interacciones proteína-proteína conocidas. Además, bases de datos que incluyen información sobre interacciones entre fármacos y sus dianas proteicas. De igual forma la propia base de datos de Anaxomics contiene información sobre el comportamiento biológico humano. Se incluyen por ejemplo las relaciones entre patologías y proteínas.
- El método ha demostrado obtener excelentes resultados, siendo capaz de predecir relaciones desconocidas e insospechadas por investigadores con el conocimiento actual.
- Sin embargo, esta aproximación también presenta muchas limitaciones:

- Al igual que ocurre en el otro método, la existencia de tanta información topológica y biológica entrelazada puede generar ruido. Dando lugar a errores como falsos positivos y negativos, que se van dilucidando y esclareciendo mediante el método científico.
- Además de partir de una correcta topología, el sistema ha de proveerse de información suficiente para la asignación del sentido de los links.
- Debido a estos inconvenientes, es posible que el modelo no se ajuste de forma óptima a las restricciones de calidad dado que no se ajustaría al modelo biológico humano de referencia en un porcentaje suficiente.

7.1.3. Resultados comunes y no comunes entre ambos métodos.

¿Qué significa que los resultados sean comunes?

Por un lado, el hecho de que un fármaco resulte seleccionado como candidato mediante ambos métodos, significa que las dos técnicas contienen la información necesaria para relacionar dicho fármaco con la patología estudiada. Lo cual no significa necesariamente que dicho fármaco esté más correlacionado con la enfermedad, sino que es posible detectarlo desde ambas perspectivas. Sin embargo, a la hora de escoger un fármaco, sí que podría ser preferible que el resultado sea común para descartar con una mayor seguridad que se trate de un falso positivo. De la misma forma, se podría considerar que los resultados negativos comunes ofrecen mayor seguridad para descartar falsos negativos.

¿Qué significa que los resultados no sean comunes?

Por otro lado, el hecho de que un fármaco resulte seleccionado como candidato mediante uno solo de los métodos, quiere decir que estamos en un punto ciego de la patología para aquel método que no lo ha detectado. Un resultado diferencial es debido a que cada método analiza tan solo una parte del conjunto, cuando en realidad los sistemas biológicos presentan interacciones altamente complejas. En este sentido, el método de Anaxomics realiza un aporte un extra, ya que intenta encontrar una solución que tenga una explicación que se sustente en el conocimiento biológico. Es decir, no sólo intenta buscar una solución sino una que se pueda comprender desde un punto de vista molecular.

Concluimos finalmente, que no existe o no podemos alcanzar una verdad absoluta, siendo los resultados complementarios, ambas respuestas pueden ser ciertas sin ser excluyentes.

7.2. Mecanismos de acción

- Se evidencia del hecho de que se hayan creado los modelos, que los fármacos estudiados tienen probabilidad de estar implicados en la patología del acné.
- Ninguna de las proteínas presentes en la figura del mecanismo de acción y tampoco de las que han participado coinciden entre ambos tratamientos.
- Se evidencian claras diferencias en las vías metabólicas afectadas por cada una de las drogas.
- Dado que el reposicionamiento ha resultado evidenciar la relación de ambas con la hiperqueratinización, podemos pensar en la posibilidad de que se lleve a cabo por otras vías que deriven en el mismo resultado.
- A la vista de estas diferencias un MoA comparativo se ha descartado por parecer redundante. Aunque, por otro lado, un MoA sinérgico podría haberse realizado para evaluar si el efecto promotor de la patología de ambos fármacos en conjunto es mayor que la suma de los dos por separado.

7.3. Análisis estadístico

- Los test estadísticos realizados han demostrado la existencia de diferencias significativas para ambos tratamientos.
- El análisis de componentes principales demuestra una alta capacidad para separar ambas condiciones.
- El test de Wilcoxon ha mostrado diferencias significativas entre los dos tratamientos para X proteínas de un total de x.

Aunque podrían haberse realizado más experimentos, todos los resultados obtenidos apuntan a que ambos fármacos podrían tener como efecto adverso el acné. Aún más que el efecto biológico, se deriva la posibilidad de que cada fármaco lo haga de forma independiente, es decir, afectando distintas vías metabólicas.

Los resultados que se presentan en este trabajo han sido obtenidos con las herramientas disponibles en Anaxomics. Estas herramientas ofrecían multitud de opciones de ejecución que no han podido ser exploradas por las limitaciones de tiempo y el entrenamiento requerido para su uso. Es necesario también comentar que los resultados obtenidos no siempre han seguido todas las PNTs (Procedimientos Normalizados de Trabajo) de la compañía ya éstas que implican a varios grupos de trabajo además del departamento de IT donde se realizaron las prácticas.

8. Conclusiones globales

- Descripción de las conclusiones generales

Como opinión personal, me gustaría remarcar el hecho de que las últimas técnicas y metodologías tecnológicas incluyendo las basadas teoría de redes y minería de datos indudablemente tienen una gran potencia de análisis y aplicabilidad. Se demuestra dado que se ha instaurado rápidamente y comienza a ser muy importante en el área de las ciencias de la vida, así como en muchos otros campos.

Sin embargo, no se puede obviar la otra cara de la moneda, que es que son técnicas por el momento difícilmente estandarizables. Cosa que ha sido muy apreciada en los trabajos científicos a lo largo de la historia, durante la cual siempre se ha tendido a trabajar imitando otras referencias para poder comparar resultados fácilmente.

Las técnicas de minería de datos son muy variadas y flexibles, por lo que por una parte se requiere del conocimiento necesario para saber cómo aplicarlas. Por otro lado, al igual que cualquier técnica nueva, requieren de un periodo de testeo y validación que permitan conocer más acerca sus posibilidades. Carencia que sin duda se conseguirá en un breve espacio de tiempo.

- Reflexión sobre las lecciones aprendidas durante la elaboración del trabajo

Más allá de las lecciones derivadas de la realización del proyecto en concreto, nos encontramos con otro aprendizaje aún más importante:

- El seguimiento del trabajo a través de las PEC y la planificación en base a unos tiempos me ha permitido mejorar las habilidades

organizativas y darme cuenta de lo importante que es fijarse objetivos y evaluar a posteriori si se han podido cumplir, generando a la vez así una visión de las capacidades personales.

- Se han podido afianzar y ampliar los conocimientos que sirven de base producto de la realización del resto del máster y comprendido la complejidad a la que pueden llegar las técnicas bioinformáticas.
- Se ha facilitado la oportunidad de trabajar de forma independiente y con libertad de decisión. A la vez que se ha prestado apoyo en caso de ser necesario.
- Se ha adquirido además una visión de la compleja infraestructura informática que requiere una empresa de biotecnología como Anaxomics, en la que se utilizan ordenadores cálculo que precisan de conexiones entre los terminales.

- Reflexión crítica sobre el logro de los objetivos

Los objetivos se han cumplido en gran medida para los distintos aspectos predefinidos.

- Reflexión crítica sobre el seguimiento de la planificación y metodología

La planificación se ha seguido de un modo correcto y en el orden que se ha ido planificando a lo largo de las tres PEC. En cambio, los períodos de tiempo se han reajustado en algún momento de la asignatura. Sin embargo, el desajuste no parece haber sido desproporcionado, teniendo en cuenta que al tratarse de conocimiento nuevo las ideas preconcebidas pueden no ajustarse con la realidad del proyecto, por tanto, no es tarea fácil definir los tiempos.

- Líneas de trabajo futuro que no se han podido explorar en este trabajo

En el futuro se debería seguir trabajando en esta línea, de comparación y validación los nuevos métodos que van surgiendo se puedan ir asimilando por la comunidad científica para poder darle un uso práctico de continuidad. Así como, conocer sus bondades y limitaciones para poder maximizar nuestras capacidades de análisis al máximo.

9. Glosario

- HT: High-Throughput.
- PP: Prior Probability.
- MoA: Mechanism of Action.
- PNT: Procedimiento Normalizado de Trabajo.
- GHC: Global Human Cell.
- TT: Truth-Table, base de datos propiedad de Anaxomics que contiene las premisas o normas o fisiológicas que se cumplen en todos los humanos o en determinadas condiciones patofisiológicas.
- BED: Biological Effectors Database, base de datos propiedad de Anaxomics que relaciona las proteínas procesos biológicos clínicos (efectos adversos, indicaciones de drogas, enfermedades, etc.).
- ID: Código de identificación.
- ANN: Artificial Neural Network.

9.1. Listado de todo el software utilizado:

- FEAX: Utilizado para el análisis de reprofiling y para subir las restricciones.
- PL/SQL Developer: utilizado para la gestión de la base de datos ORACLE.
- Matlab: Utilizado para generar los mecanismos de acción y la realización del análisis estadístico.
- GraphViz: software de visualización de grafos (utilizado indirectamente para crear los MoA y directamente para modificar algunos nombres de los nodos).
- Blue Griffon: Editor de HTML (es utilizado para el reporte en PDF de los MoA y la estadística).
- R: Se ha utilizado para realizar los gráficos circulares.

10. Bibliografía

1. *Quantification of pathway crosstalk reveals novel synergistic drug combinations for breast cancer.* S. Jaeger, A. Igea, R. Arroyo, V. Alcalde, B. Canovas, M. Orozco, A. R. Nebreda and P. Aloy. DOI: 10.1158/0008-5472.CAN-16-0097, November 22, 2016, Cancer Res, págs. 459-469.
2. *Novel Neuroprotective Multicomponent Therapy for Amyotrophic Lateral Sclerosis Designed by Networked Systems.* M. Herrando-Grabulosa, R. Mulet, A. Pujol, J. M. Mas, X. Navarro, P. Aloy, M. Coma and C. Casas. DOI:10.1371/journal.pone.0147626 , January 2016, PLOS ONE.
3. *Anaxomics' methodologies: understanding the complexity of biological processes.* R. Valls, A. Pujol, L. Artigas. 2013, white Pap. .
4. Wikipedia. [En línea] 24 de 05 de 2017.
https://es.wikipedia.org/wiki/Biolog%C3%ADa_de_sistemas.
5. *A Decade of Systems Biology.* H. Chuang, M. Hofree and T. Ideker. doi: 10.1146/annurev-cellbio-100109-104122, s.l. : Annu Rev Cell Dev Biol, 2010, Vols. 26: 721–744. .
6. *A New Method for Computational Drug Repositioning Using Drug Pairwise Similarity.* Lu, Jiao Li and Zhiyong. doi: 10.1109/BIBM.2012.6392722, October 2012 , IEEE Int Conf Bioinformatics Biomed.
7. *Filling the gap in CNS drug development: evaluation of the role of drug repurposing.* A. Caban, K. Pisarczyk, K. Kopacz, A. Kapuśniak, M. Toumi, C. Rémuzat, and A. Kornfeld. doi: 10.1080/20016689.2017.1299833, 2017, J Mark Access Health Policy.
8. E. H. Shortliffe, J. J. Cimino. *Biomedical Informatics, Computer Applications in Health Care and Biomedicine, fourth edition.* s.l. : Springer, 2014. 978-1-4471-4473-1.
9. *Benefit from B-Lymphocyte Depletion Using the Anti-CD20 Antibody Rituximab in Chronic Fatigue Syndrome.* Øystein Fluge, Ove Bruland, Kristin Risa, Anette Storstein, Einar K. Kristoffersen, Dipak Sapkota, Halvor Næss, Olav Dahl, Harald Nyland, Olav Mella. 6(10): e26358.
<https://doi.org/10.1371/journal.pone.0026358>, October 19, 2011, PLoS ONE.
10. ANAXOMICS Systems Biology Solutions. [En línea] 24 de 05 de 2017.
<http://www.anaxomics.com/>.
11. Wikipedia. [En línea] 16 de 05 de 2017.
<https://es.wikipedia.org/wiki/Acn%C3%A9>.
12. *An update on the pathogenesis and management of acne vulgaris.* Julie C. Harper, MD. Number 1, 2004, J AM ACAD DERMATOL, Vol. Vol. 51.
13. *New insights into acne pathogenesis: Exploring the role of acne-associated microbial populations.* Bipul Kumar, et al. 34, 2016, DERMATOLOGICA SINICA , págs. 67-73.
14. B. Kumar, R. Pathak, P. B. Mary, D. Jha., K. Sardana, H. K. Gautam. Issue 2, June 2016, Dermatologica Sinica, Vol. Volume 34, págs. 67–73 .

15. *Recent advances in acne pathogenesis: implications for therapy.* Das S, Reynolds RV. 15(6):479-88. doi: 10.1007/s40257-014-0099-z., 2014, Am J Clin Dermatol.
16. *Pathogenesis of acne.* Kubba R., et al. 75 Suppl 1:1-62., 2009, Indian J Dermatol Venereol Leprol.
17. DrugBank. [En línea] 24 de 05 de 2017. <https://www.drugbank.ca/>.
18. *A chemo-centric view of human health and disease.* M. Duran-Frigola, D. Rossell, P. Aloy. 5:5676. doi: 10.1038/ncomms6676, 2014 Dec 1, Nat Commun.
19. *Relating protein pharmacology by ligand chemistry.* M. J. Keiser, et al. 2, 2007, NATURE BIOTECHNOLOGY , Vol. 25.
20. Chemical Identifier Resolver. [En línea] 24 de 05 de 2017. <https://cactus.nci.nih.gov/chemical/structure>.
21. PubChem. [En línea] 24 de 05 de 2017. <https://pubchem.ncbi.nlm.nih.gov/>.
22. *Data Mining Algorithms for Clasification.* Ozer, P. BSc Thesis Artificial Intelligence, Radboud University Nijmegen : s.n., 2008.
23. Reddy, C. C. Aggarwal and C. K. DATA CLUSTERING Algorithms and Applications. s.l. : CRC Press/ Chapman and Hall, 2014. 978-1-4665-5821-2.
24. *Optimizing artificial neural network models for metabolomics and systems biology: an example using HPLC retention index data.* L. M. Hall, D. W. Hill, L. C. Menikarachchi, M. Chen, L. H. Hall and D. F. Grant. doi:10.4155/bio.15.1., s.l. : Bioanalysis, 2015, Vols. 7(8): 939–955.
25. KEGG. [En línea] 24 de 05 de 2017. <http://www.genome.jp/kegg/>.
26. REACTOME A CURATED PATHWAY DATABASE. [En línea] 24 de 05 de 2017. <http://www.reactome.org/>.
27. GEO. [En línea] 24 de 05 de 2017. <https://www.ncbi.nlm.nih.gov/gds/>.
28. IntAct, IntAct Molecular Interaction Database . [En línea] 24 de 05 de 2017. <http://www.ebi.ac.uk/intact/>.
29. BioGRID 3.4. [En línea] 24 de 05 de 2017. <https://thebiogrid.org/>.
30. MINT. [En línea] 24 de 05 de 2017. <http://mint.bio.uniroma2.it/>.
31. Wikipedia. [En línea] 24 de 05 de 2017. <https://es.wikipedia.org/wiki/Perceptr%C3%B3n>.
32. *Towards a molecular characterisation of pathological pathways.* R. A. Pache, A. Zanzoni, J. Naval, J.M. Mas, P. Aloy. DOI: 10.1016/j.febslet.2008.02.014, 2008, FEBS Letters.
33. DrugBank DB00052. [En línea] 15 de 04 de 2017. <https://www.drugbank.ca/drugs/DB00052#targets>.
34. DrugBank DB00939. [En línea] 15 de 04 de 2017. <https://www.drugbank.ca/drugs/DB00939>.
35. *Revealing the molecular relationship between type 2 diabetes and the metabolic changes induced by a very-low-carbohydrate low-fat ketogenic diet.* al., J. Farrés et. 7:88, 2010, Nutrition & Metabolism.
36. Goldman, Megan. *Why is multiple testing a problem?* s.l. : Stat C141/ Bioeng C141 - Statistics for Bioinformatics, 2008.

11. Anexos

Documentación anexa perteneciente a la ejecución de la metodología:

- 1) Front-End del software usado para el reposicionamiento.
- 2) Script en lenguaje Matlab utilizado con las llamadas a las funciones necesarias para la generación de los MoA con comentarios.
- 3) Script en lenguaje Matlab utilizado para la extracción de los valores de activación con comentarios.
- 4) Script en lenguaje Matlab utilizado con las llamadas a las funciones para realizar el análisis estadístico con comentarios.
- 5) Tabla que contiene los nodos del MoA para el fármaco Somatropina recombinante.
- 6) Tabla que contiene los nodos del MoA para el fármaco ácido Meclofenámico.

11.1. Front-End del software usado para el reposicionamiento y la subida de restricciones.

Las siguientes imágenes pertenecen a capturas de pantalla de la interfaz utilizada que ejemplifican el procedimiento de reposicionamiento. Es un software de fácil manejo, básicamente se escoge un fármaco de entrada y un conjunto de proteínas como salida y se determinan las características deseadas para realizar el experimento. El resultado se recibe por correo electrónico.

The screenshot displays the SIMSCells web interface. At the top left is the SIMSCells logo with the tagline "Bringing Systems Biology to Lab Researchers". The top right shows the user information: "SimsCells User: [2727]-eva.gabarron@anaxomics.com" and a link to "Click Here To Change SimsCells User". Navigation icons for "Restrictions", "SimsCells Sets", and "Deletion Tool" are also present.

On the left side, there is a "Shortcuts" menu with icons for: Activity Report, CRM, IT Task Manager, Project Path, Protein Sets, Interactions, and Protein Bait.

The main content area is divided into two sections. On the left is a "REPORT TYPE" list with "Distance" selected. On the right is a "SEARCH SOP:" field above a table of SOPs.

NAME	SOP_CODE	TYPE
Target Discovery Tool	H.00013	Distance
Detailed Protein Distance	H.00013	Distance
Ann Prediction	H.00013	Distance
Hausdorff Distance	H.00013	Distance

At the bottom right of the main content area, there is an "Open Form" button.

Select your Type of Report Single Drugs Prediction Combined Drugs Prediction

Select your prediction Model
 Perform with Relevant Links Ann Perform with Sampling Population
 Perform with Full Topology Ann
Ann Model: Global Human Cell(Ann Update Database November 2015 (HEALTH 2015-2016 UPDATE))
[Change Ann Model Version](#)

Inset 1:
[Add from Drug Selector](#) [Add from Existing Protein Set](#)
[Add from New Protein Set](#) [Add from File](#) **Num Sets:1**

DB00052

Inset 2:
[Add from Drug Selector](#) [Add from Existing Protein Set](#)
[Add from New Protein Set](#) [Add from File](#) **Num. Sets:**

Outlet: [Add from Existing Protein Set](#) [Add from File](#) [Add from New Protein Set](#)
[Add All Conditions](#) [Add All Motives](#) [Add All Pathways](#) **Num Sets:0**

O00206
O15263
O24006
O60603
O95864
P01137
P01375
P01579
P01583

Report Name

Description

Predict with reverse model(e.g. Biomarker Prediction)

Select your Type of Report REPROFILING

11.2. Script en lenguaje Matlab utilizado con las llamadas a las funciones necesarias para la generación de los MoA con comentarios.

```
cd('C:\Users\eva.gabarron\EvaData\TFM\Lanzar MoA'); %selecciona la carpeta desde donde vamos a trabajar
clear all %hace una limpieza de variables
LITE.LoginRemote('eva.gabarron@anaxomics.com'); %es preciso una identificación con nuestro usuario para utilizar el software
```

```
% descarga de las plantillas de Excel para introducir los datos
SM.DownloadRestrictionsTemplate('templateMoATFMdrugproacne1.xlsx');
SM.DownloadRestrictionsTemplate('templateMoATFMdrugproacne2.xlsx');
SM.DownloadRestrictionsTemplate('templateMoATFMdrugproacne3.xlsx');
```

%NOTA: la mayoría de funciones lleva delante algunas siglas, como en este caso anterior “SM” (Sampling Methods), esto es porque se trata de funciones agrupadas en clases.

```
%MOA PARA DROGA 1 SOMATOTROPINE RECOMBINANT
```

```
%=====
```

```
% FUNCIÓN:
```

```
data=GetBEDAESUniprot(aesNames,outFileName,bAddRelevances)
%para obtener las proteínas que definen los motivos de la enfermedad (de la BED)
```

```
data=DDBB.GetBEDAESUniprot({'acne'},'getFromBedProteinsRelatedToAcne.txt','true')
```

```
%para obtener las proteínas que definen la enfermedad (de la base de datos interna llamada BED (Biological Effectors Database))
%introducimos como parámetros: la patología, el nombre del fichero de salida y un parámetro opcional que añade una columna con el peso de cada proteína sobre la patología.
```

```
% La función anterior genera un documento en formato .txt:
```

```
% Columna 1: AES (Adverse Events) --> Acné
```

```
% Columna 2: código proteína efectora (Uniprot)
```

```
% Columna 3: activaciones --> 1 Activación / -1 inhibición / 9 desconocido
```

```
% (las que tienen un 9 a la hora de hacer el MoA no se tienen en cuenta)
```

% Columna 4: relevancias --> las relevancias son unos valores de peso de cada proteína efectora que permiten filtrar las que tienen mayor importancia.

%Con esta información y las proteínas diana de los fármacos se rellena la plantillas para cada experimento.

% Paso 1: generación de la población:

%=====

% FUNCIÓN:

SMFolder=GeneratePopulation(cellType,stimulusFileName,responseFileName, optionalIDAnalysis,restrictionsFileName,optionalExperimentName)

%tanto el stimulusFileName como responseFileName y restrictionsFileName son parámetros opcionales que en este caso quedarán vacíos, son para introducir restricciones extra a las introducidas con la plantilla Excel.

[inSM] = SM.GeneratePopulation('Global Human Cell', [], [], [], 'templateMoATFMdrugproacne1.xlsx', 'MoAAcneDroga1Somatotropin');

%se definen los parámetros obligatorios, tipo celular, plantilla y nombre del experimento.

% Paso 2: Generar informe

%=====

% FUNCIÓN: id_monitoring=SM.GenerateMoA(inSM,monitoringRestrictions)

%Introducimos como parámetro la población generada con función anterior, definida como en la salida como inSM. Dejamos vacío el parámetro opcional monitoringRestrictions.

[id_monitoring]=SM.GenerateMoA(inSM,[]);

%Anotamos los códigos de identificación del experimento que son proporcionados y esperamos que nos lleguen los resultados por correo electrónico.

%Repetimos el proceso para el resto de fármacos.

%MOA PARA DROGA 2 FINASTERIDE

%=====

% Paso 1: generación de la población:

%=====

```
[inSM] = SM.GeneratePopulation('Global Human  
Cell',[],[],[],'templateMoATFMdrugproacne2.xlsx','MoAAcneDroga2Finasteride');
```

% Paso 2: Generar report

%=====

```
[id_monitoring]=SM.GenerateMoA(inSM,[]);
```

%MOA PARA DROGA 3 Meclofenamic acid

%=====

% Paso 1: generación de la población:

%=====

```
[inSM] = SM.GeneratePopulation('Global Human  
Cell',[],[],[],'templateMoATFMdrugproacne3.xlsx','MoAAcneDroga3Meclofenami  
cAcid');
```

% Paso 2: Generar report

%=====

```
[id_monitoring]=SM.GenerateMoA(inSM,[]);
```

% Este proceso se repite de la misma forma después de la subida de restricciones (depurado de las interacciones).

11.3. Script en lenguaje Matlab utilizado para la extracción de los valores de activación.

```
cd('C:\Users\eva.gabarron\EvaData\TFM\Lanzar Lectura de poblaciones para hacer estadística'); %selecciona la carpeta de trabajo
```

```
% DRUG 1
```

```
%=====
```

```
clear all %limpieza de variables
```

```
idMoa=19555; %definimos la variable con el código identificación del experimento
```

```
LITE.LoginRemote('eva.gabarron@anaxomics.com');
```

```
model = axGeneticMoasReadExperimentModel(idMoa); %la salida de esta función es una estructura con los datos que definen el modelo.
```

```
modelDrug1=model; %renombramos la variable  
save('modelDrug1','modelDrug1','-v7.3'); %la guardamos
```

```
% Paso 1 : traducir proteínas target a tu modelo (de Uniprot a código numérico de Anaxomics)
```

```
%se realiza dos veces, una para cada fármaco
```

```
%Stimulus = DRUGS = TARGETS, las que se pusieron para el MoA, (2 columnas con código Uniprot y signo)
```

```
%(ELIMINANDO TODO LO QUE NO SEA +/-1)
```

```
proteinsListInStimulus=('ListaPRTsInDroga1.xlsx'); %definimos el documento donde guardamos los datos
```

```
[proteinsListStimulus]=axLiteReadUniprotActivations(proteinsListInStimulus);
```

```
%esta función lee el listado de targets que hemos introducido
```

```
targets=proteinsListStimulus(:,1); %definimos las proteínas target con la salida de la función anterior
```

```
targetSigns=cell2mat(proteinsListStimulus(:,2)); %definimos los signos de relación con la salida de la función anterior
```

```
[ia,ib]=ismember(targets,model.nodes(:,1)); %rastreo de los nombres en el campo nodes de la estructura del modelo
```

```
stiUid = [ib(find(ib)),targetSigns(find(ib))]; %selección de la información con las posiciones resultantes "ib"
```

```
% Paso 2 : Traducir proteínas efectoras a tu modelo (de Uniprot a código numérico de Anaxomics)
```

```
%Motivo AES = ACNE= EFECTORES (2 columnas con código Uniprot y signo)
```

```
proteinsListIn=('ListaPRTsInAcne.xlsx'); %definimos el documento donde guardamos los datos
```

```
[proteinsList]=axLiteReadUniprotActivations(proteinsListIn); %leemos el documento
```

```
%repetimos los pasos anteriores
```

```
efectors=proteinsList(:,1);
```

```
signs=cell2mat(proteinsList(:,2));
```

```
[ia,ib]=ismember(efectors(:,1),model.nodes(:,1));
```

```
efs=[ib(find(ib)),signs(find(ib))];
```

```
%EVALUAR LOS 100 MODELOS
```

```
populationPath = axGeneticMoasReadExperimentPupulationPath(idMoa); %la salida de esta función es la ruta de ubicación el modelo.
```

```
noms=dir([populationPath,'*.mat']); %buscamos en la ruta archivos con formato .mat
```

```
% Paso 3: Recoger las activaciones de los efectores
```

```
% queremos generar una matriz de 100 individuos x n valores de activación de los n efectores (después de haber estimulado el modelo con el fármaco).
```

```
cicles = 3;
```

```
%especificamos que busque en la red relaciones de distancia máxima 3, igual que en los MoA.
```

```
efecteAllSamples = []; % inicializamos la variable donde guardaremos los resultados.
```

```

for i=1:size(noms,1); %bucle para los i=100 individuos
    if (mod(i,10) ==0)
        sprintf('%s%04d of %04d',repmat(char(8),1,12),i,size(noms,1)),
        end
    % cada 10 individuos muestra un mensaje de recuento un la pantalla

    coco=axCacheLoadFile([populationPath,'\',noms(i).name]); %nombre del
    modelo individual=coco.

    [activacions2] = sto_stochasticNetDeterministicPropagation(model,
    coco.lkFinal, stiUid , cycles,1); %esta función calcula los valores de propagación
    a distancia 3 introduciendo el modelo poblacional, individual y los targets.

    efecteAllSamples = [efecteAllSamples, activacions2(efs(:,1),3) ]; %se
    almacena cada valor en una misma matriz.
    end

    efecteAllSamplesDrug1=efecteAllSamples;
    save('efecteAllSamplesDrug1','efecteAllSamplesDrug1','-v7.3'); %guardamos la
    variable

    %para obtener los nombres de las proteínas a las que perteneces los valores
    % hacemos el paso contrario al que hemos hecho para obtener los códigos de
    proteína

    variableNamesDrug1=modelDrug1.nodes(efs(:,1),:);

    save('variableNamesDrug1','variableNamesDrug1','-v7.3'); %guardamos el
    vector

    % DRUG 3
    %=====
    %Repetimos los pasos para el fármaco ácido meclonenámico
    clear all
    LITE.LoginRemote('eva.gabarron@anaxomics.com');

    idMoa = 19556; % ID MoA post-depurado

    populationPath = axGeneticMoasReadExperimentPupulationPath(idMoa) ;

    model = axGeneticMoasReadExperimentModel (idMoa) ;
    modelDrug3=model;
    save('modelDrug3','modelDrug3','-v7.3');
    proteinsListInStimulus=('ListaPRTsInDroga3.xlsx');

```

```
[proteinsListStimulus]=axLiteReadUniprotActivations(proteinsListInStimulus);
```

```
targets=proteinsListStimulus(:,1);  
targetSigns=cell2mat(proteinsListStimulus(:,2));  
[ia,ib]=ismember(proteinsListStimulus(:,1),model.nodes(:,1));  
stiUid = [ib(find(ib)),targetSigns(find(ib))];
```

% Paso 2 : traducir proteínas a tu modelo

%Motivo AES = ACNE= EFECTORES

```
proteinsListIn=('ListaPRTsInAcne.xlsx');
```

```
[proteinsList]=axLiteReadUniprotActivations(proteinsListIn);  
efectors=proteinsList(:,1);  
signs=cell2mat(proteinsList(:,2));  
[ia,ib]=ismember(efectors(:,1),model.nodes(:,1));  
efs=[ib(find(ib)),signs(find(ib))];
```

%EVALUAR LOS 100 MODELOS

```
noms=dir([populationPath,'*.mat']);
```

```
cicles = 3;
```

```
efecteAllSamples = []; % initialize
```

```
for i=1:size(noms,1);
```

```
    if (mod(i,10)==0)
```

```
        sprintf('%s%04d of %04d',repmat(char(8),1,12),i,size(noms,1)),
```

```
    end
```

```
    coco=axCacheLoadFile([populationPath,'\',noms(i).name]);
```

```
    %load([populationPath,'\',noms(i).name]);
```

```
    [activacions2]=sto_stochasticNetDeterministicPropagation  
(model,coco.lkFinal, stiUid ,cicles,1);
```

```
    efecteAllSamples = [efecteAllSamples, activacions2(efs(:,1),3) ];
```

```
end
```

```
efecteAllSamplesDrug3=efecteAllSamples;
```

```
save('efecteAllSamplesDrug3','efecteAllSamplesDrug3','-v7.3');
```

%OBTENER LISTA DE VARIABLES

```
variableNamesDrug3=modelDrug3.nodes(efs(:,1),:);
```

```
save('variableNamesDrug3','variableNamesDrug3','-v7.3');
```

11.4. Script en lenguaje Matlab utilizado con las llamadas a las funciones para realizar el análisis estadístico.

```
%Script ESTADISTICA
clear all %limpieza de variables

% definir carpeta de trabajo
cd('C:\Users\eva.gabarron\EvaData\TFM\Lanzar Lectura de poblaciones para
hacer estadistica');

load('efecteAllSamplesDrug1'); % Cargamos las matrices con los datos
load('efecteAllSamplesDrug3');

%transponemos las matrices, para correr las funciones de la clase EDA:
%en columnas → variables = PROTEINAS EFECTORAS
% en filas → muestras = INDIVIDUOS

efecteAllSamplesDrug1=efecteAllSamplesDrug1';
efecteAllSamplesDrug3=efecteAllSamplesDrug3';

%guardamos en un único documento Excel los datos para que sea más fácil
manejarlos (cada tratamiento en una hoja distinta)

xlswrite('efecteAllSamplesBOTHDRUGS', efecteAllSamplesDrug1)
xlsDeleteAllDefaultSheets('efecteAllSamplesBOTHDRUGS.xls') %elimina las
hojas Excel creadas por defecto
xlswrite('efecteAllSamplesBOTHDRUGS', efecteAllSamplesDrug3, 'Hoja2')

load('variableNamesDrug1.mat'); %cargamos los nombres de las variables
load('variableNamesDrug3.mat');

%comprobamos que las variables coinciden para los 2 tratamientos

variableNames1=variableNamesDrug1(:,1);
variableNames3=variableNamesDrug3(:,1);
sum(strcmp(variableNames1, variableNames3))

% en efecto son iguales, dado que el resultado es igual al nº de variables
```

% TEST 1

```
[resultsDrug1]=EDA.DescribeVariables(efecteAllSamplesDrug1,variableNames  
1,'descriptionVariablesDrug1.xlsx');
```

% TEST 2

```
[resultsDrug3]=EDA.DescribeVariables(efecteAllSamplesDrug3,variableNames  
1,'descriptionVariablesDrug3.xlsx');
```

%los resultados contienen un análisis descriptivo típico de los datos

% TEST 3

```
[edaReportPaired]=EDA.NVarsPaired('efecteAllSamplesBOTHDRUGS',[,],variab  
leNames1,[]);
```

% escogemos un test pareado para detectar diferencias en un número n de variables.

% TEST 4

%para esta función los datos de entrada se indican de forma diferente. Es necesario poner los datos en una misma matriz e indicar con una variable categórica, que en este caso es “classLabels”, a qué grupo pertenece cada fila.

```
classLabels= [];
```

```
classLabels=ones(size(efecteAllSamplesDrug1,1),1); %1 =fármaco1.
```

```
classLabels=[classLabels; zeros(size(efecteAllSamplesDrug3,1),1)];
```

```
%0= fármaco 2.
```

```
pcaModel=EDA.NVarsMPopulations([efecteAllSamplesDrug1;efecteAllSamples  
Drug3], classLabels, variableNames1, {'pcaMODEL.xlsx'}, [])
```

%generamos con esta función un análisis de componentes principales.

11.5. Tabla que contiene los nodos del MoA para el fármaco Somatropina recombinante.

Nº Nodo	ID Uniprot	Nombre completo de la proteína	Nombre representado (Uniprot Entry Name)
1	O60674	Tyrosine-protein kinase	JAK2
2	P01579	Interferon gamma	IFNG
3	P03372	Estrogen receptor	ESR1
4	P04141	Granulocyte-macrophage colony-stimulating factor	CSF2
5	P04150	Glucocorticoid receptor	GCR
6	P05231	Interleukin-6	IL6
7	P10912	Growth hormone receptor	GHR
8	P14780	Matrix metalloproteinase-9	MMP9
9	P16471	Prolactin receptor PRLR	PRLR
10	P35568	Insulin receptor substrate 1 IRS1	IRS1
11	P40763	Signal transducer and activator of transcription 3 STAT 3	STAT 3
12	P42224	Signal transducer and activator of transcription 1-alpha/beta	STAT 1
13	P42229	Signal transducer and activator of transcription 5A	STA5A
14	P51692	Signal transducer and activator of transcription 5B	STA5B
15	Q06124	Tyrosine-protein phosphatase non-receptor type 11	PTN11
16	Q92793	CREB-binding protein	CBP

Tabla 7. Identificación detallada de los nodos que aparecen en el mecanismo de acción realizado con el fármaco Somatropina recombinante con efector.

11.6. Tabla que contiene los nodos del MoA para el fármaco ácido Meclofenámico.

Nº Nodo	ID Uniprot	Nombre completo de la proteína	Nombre representado (Uniprot Entry Name)
1	O43889	Cyclic AMP-responsive element-binding protein 3	CREB3
2	Q68CJ9	Cyclic AMP-responsive element-binding protein 3-like protein 3	CR3L3
3	P01375	Tumor necrosis factor	TNFA
4	P14780	Matrix metalloproteinase-9	MMP9
5	P19875	C-X-C motif chemokine 2	CXCL2
6	P03372	Estrogen receptor	ESR1
7	P37231	Peroxisome proliferator-activated receptor gamma	PPARG
8	P04150	Glucocorticoid receptor	GCR
9	P04637	Cellular tumor antigen p53	P53
10	P06850	Corticoliberin	CRF
11	P11473	Vitamin D3 receptor	VDR
12	P19838	Nuclear factor NF-kappa-B p105 subunit	NFKB1
13	P35354	Prostaglandin G/H synthase 2	PGH2
14	Q09472	Histone acetyltransferase p300	EP300
15	Q92905	COP9 signalosome complex subunit 5	CSN5

Tabla 8. Identificación detallada de los nodos que aparecen en el mecanismo de acción realizado con el fármaco ácido Meclofenámico recombinante con efector.