



**Universidad
Europea MADRID**

***Implementación de mapas de acción
farmacogenómica a la herramienta de
análisis mecanístico HiPathia***

***Autor: Raúl Parra Garcés
Tutor del trabajo: María Peña Chilet
Tutor académico: María Peña Chilet***

***Facultad de Ciencias Biomédicas y de la Salud
Titulación: Máster Universitario en Bioinformática
Curso 2021-2022***

1 Índice

1	ÍNDICE	2
2	ÍNDICE DE TABLAS.....	4
3	ÍNDICE DE FIGURAS.....	4
4	RESUMEN	6
5	ABSTRACT	7
7	INTRODUCCIÓN	8
7.1	FARMACOGENÉTICA.....	8
7.2	ANÁLISIS MECANÍSTICOS Y BIOLOGÍA DE SISTEMAS	9
7.3	RUTAS DE SEÑALIZACIÓN.....	10
7.4	ANÁLISIS MECANÍSTICO DE RUTAS (MPA).....	11
8	OBJETIVOS	12
8.1	HIPÓTESIS.....	12
8.2	OBJETIVOS.....	12
8.2.1	<i>Objetivos concretos:</i>	13
9	MATERIAL Y MÉTODOS	13
9.1	FUENTE DE DATOS	13
9.2	EL MODELO MECANÍSTICO DE HIPATHIA	13
9.3	VÍAS MODELADAS	15
9.4	DESCOMPOSICIÓN DE LAS VÍAS DE SEÑALIZACIÓN	18
9.4.1	<i>Archivo SIF</i>	19
9.4.2	<i>Archivo ATT</i>	19
9.5	AUTOMATIZACIÓN DEL PROCESO DE LECTURA DE RUTAS	20
9.6	INTEGRACIÓN DIRECTA DE LOS DATOS TRANSCRIPTÓMICOS Y GENÓMICOS	22
9.6.1	<i>Formato de los datos</i>	22
9.7	ANÁLISIS DE LA SEÑAL	24
9.8	IMPLEMENTACIÓN	25
9.9	ANOTACIÓN DE LA FUNCIONALIDAD DE LOS CIRCUITOS	25
10	RESULTADOS.....	25
10.1	PROCESO DE AUTOMATIZACIÓN.....	25
10.1.1	<i>Primera iteración</i>	26
10.1.2	<i>Segunda iteración</i>	27
10.1.3	<i>Tercera iteración</i>	27
10.2	RUTAS FARMACOCINÉTICAS/FARMACODINÁMICAS.....	27
10.2.1	<i>Fluoropirimidinas</i>	28
10.2.2	<i>Metotrexate</i>	30
10.2.3	<i>Tiopurinas</i>	31
10.2.4	<i>Análisis de expresión génica en tejido</i>	35
11	DISCUSIÓN.....	36
11.1	HIPATHIA Y SUS USOS.....	37
11.2	FLUOROPIRIMIDINAS	38
11.2.1	<i>TYMS</i>	39

11.2.2	UPB1	39
11.3	TIOPURINAS	40
11.3.1	PPAT.....	41
11.3.2	RAC1.....	41
11.4	METOTREXATO	42
11.4.1	GART.....	42
12	LIMITACIONES.....	43
13	FUTUROS PASOS	44
14	CONCLUSIONES.....	44
15	REFERENCIAS	46
16	ANEXOS	51
16.1	LISTA DE PAQUETES DE R USADOS	51
16.2	CÓDIGO	51
16.3	VISUALIZADOR DE REPORTES HiPATHIA	52
16.4	ENLACE ARCHIVOS	52

2 Índice de tablas

TABLA 1. CRITERIOS DE PRIORIZACIÓN DEL CPIC, DISPONIBLE EN (PRIORITIZATION OF CPIC GUIDELINES – CPIC, S. F.)	16
TABLA 2. GUÍAS SEGÚN PRIORIZACIÓN DEL CPIC. TOMADO Y ADAPTADO DE (PRIORITIZATION OF CPIC GUIDELINES – CPIC, S. F.)	17
TABLA 3. INFORMACIÓN DE LA MUESTRA DE RNA-SEQ DE TEJIDO. OBTENIDO DE GTEx (GENOTYPE-TISSUE EXPRESSION (GTEx) PROJECT , 2022)	28
TABLA 4. RESUMEN CIRCUITOS HiPATHIA (ORDENADOS DESCENDENTE POR DESVIACIÓN ESTÁNDAR). MEAN: MEDIA, SD: DESVIACIÓN ESTÁNDAR, MIN: VALOR MÍNIMO, MAX: VALOR MÁXIMO.	34
TABLA 5. RELACIÓN CIRCUITO-GENES PRESENTES, PERTENECIENTES A CADA UNA DE LAS RUTAS	36
TABLA ANEXOS 1. RUTA FLUOROPIRIMIDINAS PHARMGKB	53
TABLA ANEXOS 2. TABLA REFERENCIA FLUOROPIRIMIDINAS	54
TABLA ANEXOS 3. TABLA MOLÉCULAS FLUOROPIRIMIDINAS	54
TABLA ANEXOS 4. TABLA GENES FLUOROPIRIMIDINAS	55
TABLA ANEXOS 5. TABLA CODIFICADA FLUOROPIRIMIDINAS	55
TABLA ANEXOS 6. TABLA PASOS INTERMEDIOS (EXPANDED) FLUOROPIRIMIDINAS	56
TABLA ANEXOS 7. ARCHIVO SIF FLUOROPIRIMIDINAS	56
TABLA ANEXOS 8. ARCHIVO ATT FLUOROPIRIMIDINAS	57
TABLA ANEXOS 9. GENES LIST FLUOROPIRIMIDINAS	57

3 Índice de figuras

FIGURA 1. EJEMPLO DE UN CIRCUITO DE SEÑALIZACIÓN. A) PROPAGACIÓN DE SEÑAL A TRAVÉS DE UN CIRCUITO COMPUESTO POR UNA ÚNICA PROTEÍNA RECEPTORA. B) MATRIZ DE EXPRESIÓN GÉNICA. C) ALGORITMO DE ANÁLISIS DE GRAFOS RECURSIVO DE PROPAGACIÓN DE LA SEÑAL.	14
FIGURA 2. ALGORITMO DE ANÁLISIS DE GRAFOS RECURSIVO DE PROPAGACIÓN DE LA SEÑAL (HIDALGO ET AL., 2017; SALAVERT ET AL., 2016B).	15
FIGURA 3. PROCESO DE SELECCIÓN DE RUTAS. 5 FU: FLUOROPIRIMIDINAS, MTPU: TIOPURINAS, MTX: METOTREXATO. PK: FARMACOCINÉTICA. PKPD: FARMACOCINÉTICA-FARMACODINÁMICA.	18
FIGURA 4. PASOS INTERMEDIOS PARA INTERACCIONES "MOLÉCULA-GEN-MOLÉCULA". DPYD: DIHIDROPYRIMIDINA DESHIDROGENASA	21
FIGURA 5. PROCESO DE PARSEO Y CREACIÓN DE ARCHIVOS DE IMPLEMENTACIÓN	22
FIGURA 6. RUTA FLUOROPIRIMIDINAS HiPATHIA	29
FIGURA 7. RUTA METOTREXATE HiPATHIA	30
FIGURA 8. RUTA TIOPURINAS HiPATHIA	32
FIGURA 9. HEATMAP. AGRUPACIÓN JERÁRQUICA DE ACTIVACIÓN DIFERENCIAL DE RUTAS	33
FIGURA 10. VIOLIN PLOT. ANÁLISIS DE EXPRESIÓN DE GENES RELACIONADOS CON LA RUTA DE LAS FLUOROPIRIMIDINAS EN TEJIDO HEPÁTICO SANO	35
FIGURA ANEXOS 1. RUTA FLUOROPIRIMIDINAS PHARMGKB	58
FIGURA ANEXOS 2. FLUOROPIRIMIDINAS, CIRCUITO FLUORODEOXIURIDINA TRIFOSFATO	58
FIGURA ANEXOS 3. FLUOROPIRIMIDINAS, CIRCUITO FLUOROURIDINA TRIFOSFATO	59
FIGURA ANEXOS 4. FLUOROPIRIMIDINAS, CIRCUITO TYMS	59
FIGURA ANEXOS 5. FLUOROPIRIMIDINAS, CIRCUITO UPB1	60
FIGURA ANEXOS 6. FIGURA ANEXOS 1. RUTA METOTREXATE PHARMGKB	60
FIGURA ANEXOS 7. METOTREXATE, CIRCUITO ADENOSINA TRIFOSFATO	61
FIGURA ANEXOS 8. METOTREXATE, CIRCUITO APOPTOSIS	61
FIGURA ANEXOS 9. METOTREXATE, CIRCUITO CISTATIONINA	62

FIGURA ANEXOS 10. METOTREXATE, CIRCUITO DEOXITIMIDINA MONOFOSFATO.....	62
FIGURA ANEXOS 11. METOTREXATE, CIRCUITO HIPOXANTINA.....	63
FIGURA ANEXOS 12. METOTREXATE, CIRCUITO I.SERINA.....	63
FIGURA ANEXOS 13. METOTREXATE, CIRCUITO GART.....	64
FIGURA ANEXOS 14. METOTREXATE, CIRCUITO PPAT.....	64
FIGURA ANEXOS 15. METOTREXATE, CIRCUITO DNA METILADO.....	65
FIGURA ANEXOS 16. METOTREXATE, CIRCUITO PROTEÍNAS METILADAS.....	65
FIGURA ANEXOS 17. RUTA TIOPURINAS PHARMGKB.....	66
FIGURA ANEXOS 18. TIOPURINAS, CIRCUITO 8-HIDROXITIOGUANINA.....	66
FIGURA ANEXOS 19. TIOPURINAS, CIRCUITO APOPTOSIS.....	67
FIGURA ANEXOS 20. TIOPURINAS, CIRCUITO DEOXI-TIOGUANOSINA MONOFOSFATO.....	67
FIGURA ANEXOS 21. TIOPURINAS, CIRCUITO METIL-TIOGUANOSINA MONOFOSFATO.....	68
FIGURA ANEXOS 22. TIOPURINAS, CIRCUITO METIL-TIOINOSINA MONOFOSFATO.....	68
FIGURA ANEXOS 23. TIOPURINAS, CIRCUITO METILMERCAPTOPYRINA.....	69
FIGURA ANEXOS 24. TIOPURINAS, CIRCUITO FOSFORIBOSILAMINA.....	69
FIGURA ANEXOS 25. TIOPURINAS, CIRCUITO PPAT.....	70
FIGURA ANEXOS 26. TIOPURINAS, CIRCUITO RAC1.....	70
FIGURA ANEXOS 27. TIOPURINAS, CIRCUITO RIBOSILTIOGUANINA.....	71
FIGURA ANEXOS 28. TIOPURINAS, CIRCUITO TIOINOSINA TRIFOSFATO.....	71
FIGURA ANEXOS 29. TIOPURINAS, CIRCUITO ÁCIDO TIOÚRICO.....	72
FIGURA ANEXOS 30. TIOPURINAS, CIRCUITO ENFERMEDAD HEPÁTICA TÓXICA.....	72
FIGURA ANEXOS 31. VIOLIN PLOT. ANÁLISIS DE EXPRESIÓN DE GENES RELACIONADOS CON LA RUTA DE METOTREXATE EN TEJIDO HEPÁTICO SANO.....	73
FIGURA ANEXOS 32. VIOLIN PLOT. ANÁLISIS DE EXPRESIÓN DE GENES RELACIONADOS CON LA RUTA DE LAS TIOPURINAS EN TEJIDO HEPÁTICO SANO.....	73

4 Resumen

Los modelos computacionales basados en la biología de sistemas nos permiten estudiar procesos biológicos complejos. El algoritmo HiPathia es un modelo mecanístico de análisis de vías de señalización que proporciona interpretaciones de las consecuencias que tienen los niveles de expresión génica sobre los circuitos de señalización y la funcionalidad celular. Esto permite realizar simulaciones de los mecanismos de enfermedad o la respuesta a fármacos (farmacocinética y farmacodinámica), con la intención de predecir su eficacia y sus posibles efectos adversos a partir de datos de expresión de genes en muestras de tejido.

Para ello se implementaron, por medio de un *parseador* automático, tres vías de metabolismo de fármacos disponibles en las bases de datos PharmGKB, y se adaptó la información de estas rutas a la herramienta HiPathia. Posteriormente datos de hígado sano disponibles en GTEx, fueron utilizados para análisis de expresión diferencial de genes. Aquí se presentan tres casos de uso: fluoropirimidinas, tiopurinas y metotrexato. Se detectaron un total de 206 nodos, y 27 circuitos con activación diferencial.

El enfoque actual centrados en genes individuales y sus variantes ignora la complejidad de los procesos celulares que explican los fenotipos de metabolización. Por tanto, el uso de modelos mecanísticos abren nuevas vías para comprender las complejas relaciones entre los genes que, en última instancia, permiten la inferencia del fenotipo de metabolización.

Por medio de su implementación con 3 casos de uso, se demuestra que los métodos de análisis mecanístico de rutas, por medio de la herramienta HiPathia, pueden ayudar a descubrir mecanicistas biomarcadores predictivos de la eficacia y la toxicidad de fármacos.

Palabras clave: Farmacogenética, farmacocinética, farmacodinámica, análisis mecanístico, expresión diferencial de genes, biología de sistemas, medicina personalizada.

5 Abstract

Computational models based on systems biology allow us to study complex biological processes. The HiPathia algorithm is a mechanistic model of signaling pathway analysis that provides interpretations of the consequences of gene expression levels on signaling circuits and cellular function. This allows us to model disease mechanisms or drug response (pharmacokinetics and pharmacodynamics), with the intention of predicting the efficacy and possible adverse effects from gene expression data in tissue samples.

For this purpose, three drug-related metabolic pathways available in the PharmGKB database were implemented using an automatic parser, and the information from these pathways was adapted to the HiPathia tool. Data collected from healthy liver samples available in GTEx were used for gene expression analysis. Here we discuss three use cases: fluoropyrimidines, thiopurines and methotrexate. A total of 206 nodes, and 27 differentially activated circuits were detected.

The current trend focuses on individual genes and variants, and ignores the complexity of cellular processes explaining metabolization phenotypes. Therefore, the use of mechanistic models opens a new opportunity to understand complex relationships between genes that ultimately are the base for drug-related metabolic phenotype inference.

Through the implementation of 3 use cases, we demonstrate that mechanistic pathway analysis methods, using the HiPathia tool, will help us discover predictive biomarkers for drug efficacy and toxicity.

Keywords: pharmacogenetics, pharmacokinetics, pharmacodynamics, mechanistic pathway analysis, differential gene expression, systems biology, personalized medicine.

6 Introducción

6.1 Farmacogenética

La medicina personalizada se define como una disciplina médica emergente que implica el uso de información genómica de un individuo como parte de atención sanitaria, diagnóstico o tratamiento, así como el manejo de los resultados, las implicaciones éticas y políticas que se derivan de ese uso (Weinshilboum & Wang, 2017).

El exponencial abaratamiento de los costes asociados a la secuenciación de un genoma, gracias a las técnicas de secuenciación de segunda generación (NGS), han potenciado un mayor acceso a información genómica, permitiendo acelerar el descubrimiento de biomarcadores de enfermedad y de dianas terapéuticas.

La farmacogenética es la ciencia que estudia la contribución de factores genéticos a la variación interindividual en la respuesta a fármacos y sus efectos adversos. La farmacogenómica, por su parte, es un concepto más amplio, ya que integra diversas ciencias “ómicas” para explicar dicha respuesta. En el contexto de la medicina personalizada, su uso pretende ayudar a predecir la dosis necesaria para que el tratamiento de un individuo concreto sea lo más eficaz y seguro posible, minimizando los efectos indeseados, que pueden ser incluso potencialmente mortales (Weinshilboum & Wang, 2017).

La variabilidad genética interindividual, entre otros elementos que escapan el objetivo de este trabajo, puede tener un impacto sobre la proteína final, en especial en aquellas proteínas involucradas en la farmacocinética o la farmacodinámica de un fármaco. Los genes que afectan a la farmacocinética suelen referirse a las enzimas que intervienen en la absorción, distribución, metabolismo o eliminación (ADME) e incluyen transportadores y miembros de la familia CYP450. Los genes que afectan a la diana de un fármaco son intrínsecos al fármaco y pueden incluir enzimas, proteínas o receptores. En concreto, la farmacogenética estudia la relación entre los genes y estos efectos (Gonzalez-Covarrubias et al., 2020).

6.2 Análisis mecanísticos y biología de sistemas

Comprender los aspectos de la funcionalidad celular que explican los mecanismos fisiopatológicos o los mecanismos de acción de los fármacos es uno de los principales retos de la medicina de precisión (Auffray & Hood, 2012).

La biología de sistemas es un campo interdisciplinar que combina elementos de la biología molecular y la fisiología con enfoques de modelización cuantitativa integrando elementos de diferentes disciplinas como la ingeniería, la física, la informática y las matemáticas (Rian et al., 2021). El término, acuñado en 1968, se refiere al uso de la teoría de sistemas para entender los sistemas biológicos por medio de circuitos interconectados en una serie de pasos secuenciales (Mesarović, 1968).

Los modelos computacionales mecanísticos se han descrito dentro de la biología de sistemas, y permiten estudiar los mecanismos de regulación implicados en diversos procesos biológicos. Estos modelos proporcionan un medio para analizar la dinámica de los sistemas que describen, así como para estudiar sus propiedades, proporcionando información sobre el comportamiento del sistema en presencia de perturbaciones simples o complejas (Niarakis & Helikar, 2021).

Un rasgo distintivo de la investigación mecanística dentro de la biología de sistemas es intentar comprender un todo complejo descomponiéndolo en sus partes y localizando los fenómenos de interés entre los componentes de dicho sistema (Bechtel & Richardson, 2010). Los organismos vivos y sus componentes: células, tejidos y órganos, son sistemas dinámicos que responden y se adaptan continuamente a diferentes situaciones y a diversos estímulos (Helikar et al., 2008; Puniya et al., 2018). Los modelos computacionales mecanísticos buscan añadir esta dinámica a los métodos de comprensión de los sistemas biológicos complejos.

Los modelos planteados desde la perspectiva de la biología de sistemas se basan igualmente en entidades e interacciones moleculares medidas empíricamente. Dada la abundancia de diferentes moléculas y vías de señalización en cada célula, la representación de todas estas condiciones implica la selección de aquellos componentes considerados como más relevantes para el sistema investigado (Brigandt, 2013). De esta manera, se busca facilitar la interpretación de datos cuantitativos mediante modelos matemáticos y computacionales (Boogerd et al., 2007; Westerhoff & Alberghina, 2005). Actualmente, se incorporan además las nuevas técnicas experimentales que han permitido la secuenciación masiva del ADN y las mediciones automatizadas de las interacciones moleculares a dichos modelos (Ideker et al., 2001; Kitano, 2002).

6.3 Rutas de señalización

Los modelos mecanísticos de las vías de señalización proporcionan una herramienta para explorar las conexiones entre las variaciones del genotipo, en relación con la actividad o la integridad de los genes, y las variaciones del fenotipo a nivel de las células, los tejidos o los organismos (Yu et al., 2016a). Se construyen sobre gráficos que representan el conocimiento biológico de las relaciones funcionales complejas entre las proteínas dentro de la célula. De esta forma, proporcionan un marco conceptual para la interpretación de los datos de expresión génica o de variación genómica y sus consecuencias sobre los procesos biológicos y las respuestas fenotípicas, como la proliferación y la muerte celular, que son especialmente relevantes para estudiar enfermedades o la respuesta a fármacos (Rian et al., 2021).

Uno de los principales inconvenientes de los biomarcadores y las variantes genéticas es que no tienen una interpretación fácil porque a menudo carecen de cualquier vínculo mecanístico con los procesos celulares fundamentales que explican el fenotipo estudiado. En realidad, las diferencias fenotípicas se entienden mejor como alteraciones en la actividad de los módulos funcionales de la célula, que pueden ser: complejos protéicos, vía de señalización u orgánulo celulares (Oti & Brunner, 2007).

Los enfoques clásicos basados en los genes ignoran la naturaleza modular de la mayoría de los rasgos humanos, mientras que los enfoques convencionales de enriquecimiento de vías sólo producen resultados ilustrativos de limitada utilidad práctica (Oti & Brunner, 2007).

En realidad, las descripciones de vías más utilizadas son conceptos abstractos contruidos en torno a nociones biológicas completas (por ejemplo, el ciclo celular) y pueden entenderse como compendios de diferentes actividades celulares posibles que, en última instancia, se corresponden con un fenotipo (por ejemplo, mecanismos de acción de los fármacos). Estas actividades celulares son causadas por la acción combinada de diferentes genes que a menudo participan en más de una actividad celular. Esta es la razón por la que la actividad de los genes de forma individual no suele ser descriptiva de la funcionalidad celular, ya que depende críticamente de la actividad de otros genes asociados en la vía para producir funcionalidades celulares específicas, como se describe en el mapa de interacciones que representan cada una de las vías (Amadoz et al., 2019).

Uno de los aspectos más importantes, por tanto, de los modelos mecanísticos es que transmiten la noción de causalidad y pueden utilizarse para predecir las consecuencias posteriores de las perturbaciones de condiciones específicas (Salavert et al., 2016a). El efecto de las mutaciones también puede simularse con modelos mecanísticos, lo que demuestra que estos modelos son excelentes herramientas para interpretar las consecuencias de patrones mutacionales complejos de los procesos biológicos (Razzoli et al., 2016).

6.4 Análisis mecanístico de rutas (MPA)

Los métodos de análisis de actividad mecanística de las vías o rutas (MPA, por sus siglas en inglés), permiten recodificar mediciones genómicas poco informativas en valores cuantitativos de actividad celular y relacionarlos con los fenotipos (Amadoz et al., 2019).

HiPathia es un algoritmo y método de análisis mecanístico de rutas, que integra una red de interacciones de genes dirigida, basada en la evidencia biológica y conocimiento previo, que permite además añadir información de expresión de genes de diferentes tejidos y condiciones. Proporcionando así el contexto funcional de la célula y un modelo mecanístico de la enfermedad.

El método define los circuitos como subgrafos que conectan las proteínas receptoras con las proteínas efectoras en las vías. Estos tipos de circuitos representan la lógica de la transducción de señales en la célula y están relacionados con la mecánica de la funcionalidad celular. HiPathia utiliza los valores de expresión génica normalizados como representaciones de la actividad de las proteínas y, teniendo en cuenta las relaciones entre las proteínas (activación/inhibición) a lo largo de la vía, utiliza un algoritmo de propagación para estimar la cantidad de señal que llegó a la proteína efectora desde la proteína receptora. A continuación, se comparan los perfiles de actividad del circuito mediante una prueba de Wilcoxon convencional (Hidalgo et al., 2017).

HiPathia está disponible en tres formatos diferentes: i) como un paquete R/Bioconductor para usuarios interesados en un uso programático del algoritmo, ii) como un *plugin* para la comunidad Cytoscape (Shannon et al., 2003), que ofrece un entorno gráfico para usuarios finales, y iii) como una herramienta web con una interfaz gráfica intuitiva y dinámica (Peña-Chilet et al., 2019). La interfaz web implementa funcionalidades adicionales que permiten realizar comparaciones de dos clases: i) el análisis del impacto de intervenciones simuladas (inhibiciones tipo *knock-outs* o *knock-downs*,

sobreexpresiones, etc.) sobre la actividad de las vías, y ii) la evaluación de las consecuencias potenciales de las mutaciones sobre la modulación de la señalización (Salavert et al., 2016b).(Hidalgo et al., 2017)(Rian et al., 2021)

7 Objetivos

7.1 Hipótesis

La generación de redes de interacción dirigidas y mapas de procesos celulares son necesarios para modelizar eventos relevantes como mecanismos de enfermedad o respuesta a fármacos, por ello, el uso de información estructurada de rutas de señalización y metabólicas relacionadas con procesos farmacogenéticos (farmacocinética y farmacodinámica) nos permitirá modelizar estos procesos y predecir la respuesta a fármacos a partir de datos de expresión de genes en tejido empleando métodos de análisis mecanístico de actividad de rutas (MPA) como HiPathia. Empleando estos modelos podremos evaluar la variabilidad individual y poblacional relacionada con la respuesta a fármacos, su metabolización y sus posibles efectos secundarios.

7.2 Objetivos

En este trabajo se pretende aplicar una capa adicional al algoritmo ya desarrollado de HiPathia (Rian et al., 2021), de modo que pueda integrar la información procedente de la acción farmacogenética y farmacocinética de fármacos, desde un punto de vista mecanístico. Para ello se implementarán las vías de metabolización disponibles en las bases de datos como PharmGKB (Gong et al., 2021), (Kanehisa et al., 2017)(Wishart et al., 2006)(Frolkis et al., 2010)y se adaptará la información de rutas farmacogenómicas a la herramienta HiPathia, pudiéndose emplear en el paquete de R/Bioconductor (Hidalgo et al., 2017) y a través de la herramienta web. (Hidalgo et al., 2017)

Una vez terminado el proceso de implementación, se analizarán unos datos de tejido sano para mostrar un caso de uso, haciendo hincapié en la interpretación de los resultados obtenidos.

7.2.1 Objetivos concretos:

1. Obtener y *parsear* las rutas de farmacocinética y farmacodinámica disponibles en el repositorio PharmGKB.
2. Adaptar el formato de las rutas de PharmGKB a la herramienta de análisis mecanístico HiPathia.
3. Desarrollar un *parseador* automático de rutas de PharmGKB, que permita una actualización periódica de la información para su implementación en la herramienta HiPathia.
4. Evaluar la variabilidad de la actividad de las rutas farmacogenéticas *parseadas* con datos de expresión de genes en tejido.

8 Material y métodos

8.1 Fuente de datos

La modelización de la actividad de las vías requiere inicialmente de una descripción de las relaciones entre las proteínas dentro de la vía, que puede tomarse de diferentes repositorios de vías. En este marco de modelización se utiliza un total de 3 rutas de un total de 206 disponibles en el repositorio PharmGKB (Gong et al., 2021). Se incluyeron las vías correspondientes a los fármacos: Fluoropirimidinas (PA150653776), Tiopurinas (PA2040) y Metotrexate (PA2039), que incluyen las principales categorías relacionadas con la farmacología, como son: vías de farmacocinética y farmacodinamia.

Para probar el funcionamiento de las rutas, utilizamos muestras de expresión de RNA-Seq provenientes de GTEx (*Genotype-Tissue Expression (GTEx) Project*, 2022). El proyecto Genotipo-Expresión Tisular (GTEx) (*Genotype-Tissue Expression (GTEx) Project*, 2022) es un esfuerzo continuo para construir un recurso público integral para estudiar la expresión y regulación de genes específicos de los tejidos. Este repositorio cuenta muestras de tejidos no enfermos en casi 1000 individuos, principalmente para ensayos moleculares que incluyen WGS, WES y RNA-Seq.

8.2 El modelo mecanístico de HiPathia

El modelo mecanístico de HiPathia permite modelar el comportamiento de las vías de señalización, descritas como grafos dirigidos que conectan proteínas receptoras con proteínas efectoras a través de una cadena de activaciones e inhibiciones ejercidas por

proteínas intermedias. Para modelar en detalle las funcionalidades celulares específicas, cualquier vía se descompone en circuitos. Los circuitos de señalización constituyen las entidades funcionales elementales que conectan uno o más receptores con una proteína efectora, que en última instancia desencadena funciones específicas en la célula (Rian et al., 2021).

Sin embargo, las vías disponibles en los diferentes repositorios (KEGG (Kanehisa et al., 2017), PharmGKB (Gong et al., 2021), etc.) describen entidades multifuncionales con actividad a menudo opuestas (por ejemplo, la muerte y la supervivencia celular son desencadenada por la vía de la apoptosis, dependiendo de los circuitos de señalización que transducen la señal). Por lo tanto, para modelar en detalle las funcionalidades celulares específicas, cualquier vía se descompone primero en circuitos. El algoritmo HiPathia simula la propagación de la señal considerando el nivel de actividad de las proteínas que componen el circuito. De este modo, diferentes medidas de la actividad de las proteínas pueden convertirse en actividades del circuito de señalización y, en consecuencia, en perfiles funcionales de la actividad celular (Hidalgo et al., 2017). Para que un circuito esté activo y transduzca la señal para desencadenar una función, requiere de la presencia simultánea de la cadena de proteínas que conectan el receptor con el efector. Del mismo modo, la ausencia de proteínas inhibitoras puede comprometer la transducción de la señal a lo largo del circuito.

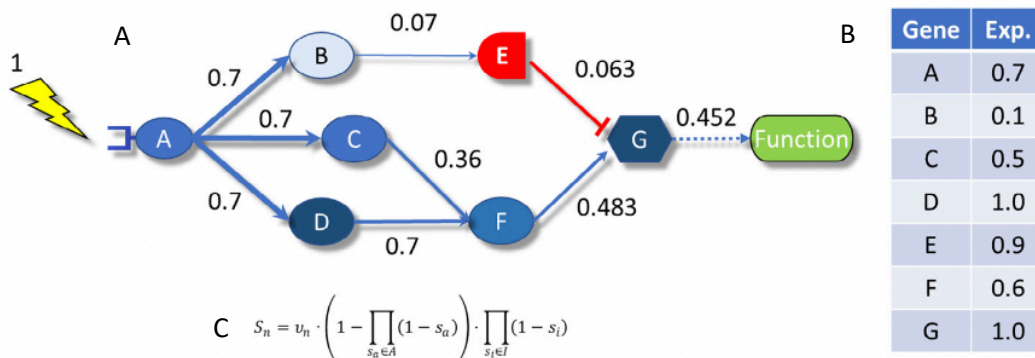


Figura 1. Ejemplo de un circuito de señalización. A) Propagación de señal a través de un circuito compuesto por una única proteína receptora. B) Matriz de expresión génica. C) Algoritmo de análisis de grafos recursivo de propagación de la señal.

La figura 1, tomada con permiso de Rian, K. et al (Rian et al., 2021) ejemplifica cómo se transduce la señal a lo largo de un circuito. En respuesta a estímulos específicos el nodo de entrada genera una señal (con un valor arbitrario de 1), que se transducen a lo largo de la vía siguiendo la dirección de las interacciones hasta llegar a un nodo de salida. Una vez alcanza el nodo efector, desencadena alguna funcionalidad celular específica

como respuesta. En este ejemplo observamos que el circuito se compone de una proteína receptora A, 5 proteínas que transducen la señal B, C, D, E y F (teniendo una de ellas interacción de tipo inhibición), y una proteína efectora G, que finalmente desencadena una función. La tabla (1.B) representa los valores de expresión génica normalizados. La aplicación de la fórmula (1.C) partiendo de una señal hipotética de 1 da lugar a una señal de intensidad final de 0,452.

Para cada que compone circuito de transducción de señales, la señal se propaga a lo largo de los nodos según la siguiente regla recursiva que mencionamos previamente:

$$S_n = v_n \cdot \left(1 - \prod_{s_i \in A_n} (1 - s_i)\right) \cdot \prod_{s_j \in I_n} (1 - s_j)$$

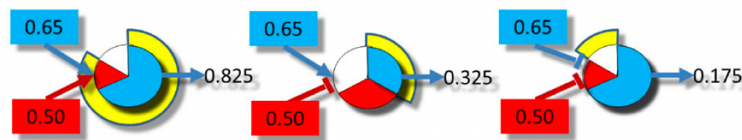


Figura 2. Algoritmo de análisis de grafos recursivo de propagación de la señal (Hidalgo et al., 2017; Salavert et al., 2016b).

donde S_n es la intensidad de la señal para el nodo n actual, V_n es su valor de expresión génica normalizado, A es el conjunto de señales de activación (S_a), que llegan al nodo actual desde los bordes de activación, I es el conjunto de señales de inhibición (S_i) que llegan al nodo desde los bordes de inhibición (Yu et al., 2016b). La causalidad se transmite por el hecho de que los niveles de expresión determinarán las consecuencias funcionales, siguiendo las reglas impuestas por el circuito de transducción. Además, los cambios en la actividad de los nodos se reflejarán (o pasarán desapercibidos) en el último nodo efector, dependiendo de la topología de cada circuito

8.3 Vías modeladas

En la búsqueda, se han identificado un total de 206 vías en PharmGKB (Gong et al., 2021)(Frolkis et al., 2010), correspondientes a la especie humana (*homo sapiens*). Estas rutas están clasificadas en categorías de uso terapéutico de los fármacos que las componen, e incluyen: 21 fármacos antimicrobianos, 48 fármacos anticancerígenos, 34 cardiovasculares y hematológicos, 7 para enfermedades endocrinas y metabólicas, 8 gastrointestinales, 11 musculoesqueléticos, 46 neurológicos, 35 analgésicos, antiinflamatorios e inmunomoduladores, 16 respiratorios, además de 6 rutas de mecanismos fisiológico.

Las vías de PharmGKB van acompañadas de un resumen escrito de la vía de PK y/o PD, así como de otra información importante relacionada con la farmacogenética. Las interacciones dentro de cada vía están respaldadas por elementos curados manualmente a partir de la literatura publicada. La información contenida en cada vía está disponible para su descarga en los formatos TSV, BioPAX y GPML. El gráfico de la vía está disponible para su descarga en formato de archivo PDF.

Los elementos que tuve en consideración para la selección de estas vías fueron elegidos a partir del nivel de evidencia disponible en torno a las rutas, y su asociación con una guía de recomendación. El Consorcio para la implementación clínica de la farmacogenética (CPIC, por sus siglas en inglés) asigna niveles de evidencia a los pares de genes-fármacos. Los niveles en orden decreciente se clasifican como: A, B, C y D. Sólo aquellos pares de genes-fármacos cuentan con evidencia suficiente han sido objeto de generación de una guía de recomendaciones terapéuticas. Únicamente los pares de nivel A y B del CPIC tienen evidencia suficiente para que se recomiende al menos una acción de prescripción.

La lista de criterios de priorización propuesta por el CPIC (*Prioritization of CPIC Guidelines – CPIC, s. f.*), puede verse en la tabla 1 a continuación.

La priorización del orden de redacción de las directrices para los genes/fármacos CPIC se basa en los siguientes criterios:

- ¿Existe la posibilidad de prescribir?
- ¿Cuál es la gravedad de las consecuencias clínicas (efectos adversos, falta de respuesta) si la genética no se utiliza para informar la prescripción?
- ¿Está el gen ya sujeto a otras directrices del CPIC?
- ¿Existe una prueba genética disponible para ese gen?
- ¿Qué frecuencia de uso tienen los medicamentos afectados?
- ¿Cómo de comunes son las variantes genéticas de alto riesgo?
- ¿Se mencionan las pruebas genéticas en el etiquetado de los medicamentos?
- ¿Existen recomendaciones de prescripción basadas en la farmacogenética por parte de organizaciones profesionales u otras?

Tabla 1. Criterios de priorización del CPIC, disponible en (Prioritization of CPIC Guidelines – CPIC, s. f.)

Teniendo en cuenta estos elementos, a la fecha de realización de este trabajo, el CPIC recoge un total de 26 asociaciones gen-fármaco en el nivel A con guías publicadas. Esto significa que información genética debe utilizarse para cambiar la prescripción del medicamento afectado. La preponderancia de las evidencias es alta o moderada a favor de cambiar la recomendación terapéutica.

Guías publicadas, en proceso o en planificación por CPIC

Gen-Fármaco	Nivel de evidencia CPIC	Estado
26	A	26 publicadas
10	B	4 en proceso
4	C	4 en planificación

Tabla 2. Guías según priorización del CPIC. Tomado y adaptado de (Prioritization of CPIC Guidelines – CPIC, s.f.)

Adicionalmente me aseguré de contar con rutas pertenecientes a dos áreas relevantes para este estudio: farmacocinética y farmacodinámica. La farmacocinética estudia cómo un organismo metaboliza un fármaco. Los genes que afectan a la farmacocinética de un medicamento pueden alterar su velocidad de metabolización, y por tanto está directamente relacionada con la efectividad y la toxicidad de un fármaco. La farmacodinámica estudia el mecanismo de acción. Los genes involucrados en la farmacodinámica de un compuesto codifican para proteínas que se unen al fármaco, que son necesarias para su transporte o para su acción. Mutaciones en estos genes pueden alterar la estructura final de la proteína, afectando a su capacidad de unión y su efectividad. Se incluyó 1 ruta exclusivamente de farmacocinética (PK), y 2 rutas integradas de farmacocinética-farmacodinámica (PD).

De éstas, se escogieron 2 rutas: Fluoropirimidinas y Tiopurinas. Adicionalmente, se escogió 1 ruta adicional: Metotrexato (nivel C) que, a pesar de no contar con una guía de recomendación disponible, se trata de un fármaco relevante por su amplio especto de aplicación, por el conocimiento profundo de su mecanismo de acción, y por su relación categórica con los dos fármacos anteriores dentro de la clasificación de fármacos antimetabolitos.

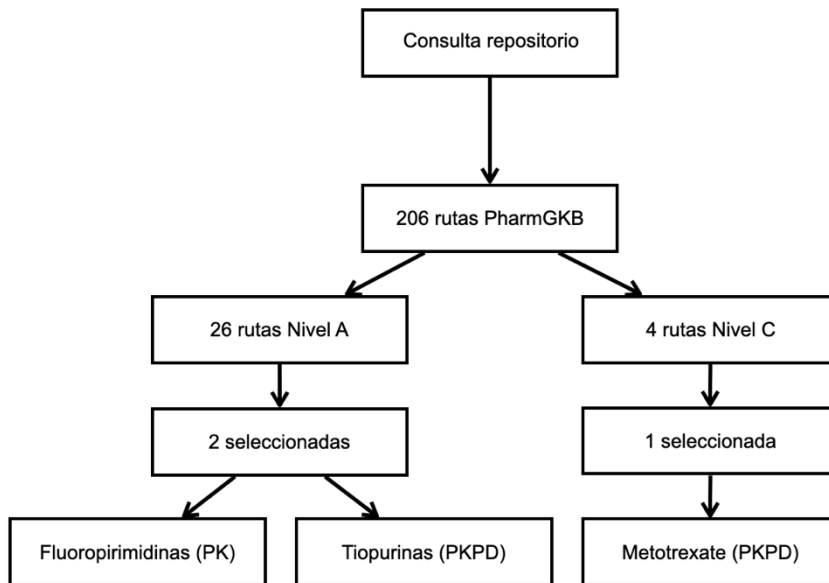


Figura 3. Proceso de selección de rutas. PK: Farmacocinética. PKPD: Farmacocinética-farmacodinámica.

Las vías son redes dirigidas en las que los nodos (compuestos por una o más proteínas) se relacionan entre sí mediante aristas. Sólo se consideran dos tipos de relación entre nodos: las activaciones y las inhibiciones. En las vías PharmGKB (Gong et al., 2021), las aristas definen diferentes tipos de interacciones proteicas que incluyen reacción bioquímica, transporte, etc., a las que se asigna una etiqueta que indica si actúan como activaciones o inhibiciones (Hidalgo et al., 2017).

A partir de estas rutas, se obtendrán i) los archivos de interacción simple (SIF), que describen las interacciones direccionales entre las proteínas de la vía, junto con ii) los archivos de atributos (ATT), que contienen información sobre la disposición de la vía y detalles de formato, que permitirán construir modelos para nuevas vías definidas por el usuario.

8.4 Descomposición de las vías de señalización

Fue necesario iniciar un proceso de automatización de lectura de las rutas disponibles en los repositorios mencionados anteriormente. Para llevarlo a cabo, la aproximación inicial fue descomponer en pasos secuenciales el proceso comprendido desde la lectura de cada ruta individual en formato TVS hasta la obtención de las tablas finales procesadas para su implementación en HiPathia.

HiPathia es capaz de leer e incluir gráficos de archivos con atributos con las siguientes características: i) Cada vía debe guardarse en dos archivos diferentes: .att (archivo ATT)

y .sif (archivo SIF). ii) Los archivos SIF y ATT deben tener el mismo nombre, es decir, "XXX.sif" y "XXX.att" para la vía con ID "XXX". iii) Las funciones no se incluyen en estos archivos, sino que se anotan *a posteriori* siguiendo un archivo de anotaciones de genes a funciones.

8.4.1 Archivo SIF

Los archivos con la extensión SIF (*Windows Setup Information Format*) se clasifican como archivos de configuración. Debe tener el siguiente formato: Archivo de texto con tres columnas separadas por tabuladores. Cada fila representa una interacción en la vía. La primera columna es el nodo origen, la tercera el nodo destino y la segunda el tipo de relación entre ellos. Esta interacción debe ser estandarizada, e incluir únicamente las acciones "activación" e "inhibición". El nombre de los nodos en este archivo se almacenará como los ID de los nodos. Los ID de los nodos deben tener la siguiente estructura N (guión) ID de la vía (guión) ID del nodo. HiPathia distingue entre dos tipos de nodos: simples y complejos. Los nodos simples pueden incluir muchos genes, pero sólo uno es necesario para realizar la función del nodo. Los nodos complejos incluyen diferentes nodos simples y representan complejos de proteínas. Cada nodo simple dentro del complejo representa una proteína del complejo. Este nodo requiere la presencia de todos sus nodos simples para realizar su función. Los ID de los nodos complejos son la yuxtaposición de los ID de los nodos simples incluidos, separados por espacios.

8.4.2 Archivo ATT

Los archivos con la extensión ATT se clasifican como archivos de configuración de atributos en el caso de HiPathia. Incluye el siguiente formato: Archivo de texto con doce columnas separadas por tabuladores. Cada fila representa un nodo (simple o complejo). Las columnas incluidas son: ID: El ID del nodo, Etiqueta: Nombre que se mostrará en la imagen de la vía. Generalmente, se utiliza como etiqueta el nombre del primer gen incluido en el Entrez ID. Coordenada X e Y de la posición del nodo en la vía (*layout*). Color: Color por defecto del nodo. Forma: Forma del nodo. Se debe utilizar "rectángulo" para los genes y "círculo" para los metabolitos. Tipo: Tipo del nodo, ya sea "gen" para los genes o "compuesto" para los metabolitos. Para los nodos complejos, el tipo de cada uno de sus nodos simples incluidos se yuxtapone separado por comas. Label.cex: Cantidad en la que se debe escalar la etiqueta de trazado respecto a la predeterminada. Label.color: Color por defecto del nodo. Anchura: Anchura por defecto del nodo. Altura: Altura por defecto del nodo. Genes List: Lista de genes incluidos en cada nodo, con Entrez ID. Nodos simples: Entrez ID de los genes incluidos, separados por comas (" , ")

y sin espacios. Nodos complejos: Lista de genes de los nodos simples incluidos, separados por una barra ("/") y sin espacios, y en el mismo orden que en el ID del nodo.

La disposición de cada nodo en el espacio (*layout*) se obtiene a partir de una serie de coordenadas disponibles en el archivo en formato GPML. Sin embargo, esta opción nos dio problemas por el solapamiento de algunos de los circuitos, por lo que adicionalmente fue modificado de forma manual para facilitar su visualización e interpretación.

8.5 Automatización del proceso de lectura de rutas

El proceso consistió en una primera iteración que buscaba descomponer las rutas en los elementos necesarios para la generación de los archivos de implementación .sif y .att., y se resume en la figura 5 al final de este apartado. Para ello, se realizó la lectura de los archivos correspondientes a las rutas previamente seleccionadas, que habían sido obtenidos desde PharmGKB. Este primer paso con la intención de explorar los archivos y determinar sus características principales. Se fragmentó la cada tabla correspondiente a las rutas en sus dos componentes principales: genes y compuestos. En ambos casos, la idea fue identificar los elementos únicos pertenecientes a cada uno de los grupos, que permitiera la creación de un diccionario de codificación para poder nombrar estos elementos según las características necesarias para su reconocimiento en HiPathia, pero que al mismo tiempo permitiera su consulta posterior en caso de ser necesario.

Este proceso genera tres *dataframes* iniciales: tabla original (fármaco), tabla de moléculas (fármaco_mol), tabla de genes (fármaco_gen). La tabla original se compone de al menos 4 columnas, que incluyen: *from*, *reaction type*, *to*, *gene(1, 2, 3...)* y número filas variables en función de la cantidad de reacciones bioquímicas. La tabla de moléculas tiene 2 columnas: *molecules*, *coded_mol*. La tabla de genes tiene 3 columnas: *symbol*, *EntrezID*, *coded_gene*. La anotación en EntrezID supone un paso adicional, pero imprescindible, para poder ser compatible con HiPathia.

El siguiente paso consiste en la codificación de los compuestos y genes según la nomenclatura estandarizada para la creación de los archivos. Como se explica previamente, consiste en N (guión) ID de la vía (guión) ID del nodo. El ID de la vía se asigna a través del reconocimiento del nombre de los archivos descargados directamente desde PharmGKB (Gong et al., 2021). Para su consulta posterior, se ha creado un archivo que incluye todos los nombres de los archivos y su fármaco correspondiente. El ID del nodo se genera de forma secuencial, utilizándose un código de dos dígitos (00) para moléculas y tres dígitos (000) para genes, para facilitar la

inspección del proceso. Utilicé los paquetes *rebus* (34) y *stringr* (35,36) para la obtención y extracción de caracteres según las reglas establecidas. En nuestro caso, todas las rutas inician con las letras en mayúscula “PA” correspondientes a “*pathway annotation*”, seguidas de un código numérico sin espacios ni caracteres.

Los valores codificados obtenidos en los pasos anteriores son “mapeados” a una única tabla que contiene todos los valores, originales y codificados, para su control y consulta. Este *dataframe* se compone de al menos 7 columnas, que incluyen: *from*, *mapped_mol*, *reaction type*, *to*, *mapped_mol2*, *gene*(1, 2, 3...), *Entrez*(1, 2, 3...), *coded_gene*(1, 2, 3...).

Posteriormente es necesario crear una tabla extendida de pasos intermedios. Esto es necesario porque las rutas disponibles en PharmGKB recogen un formato “molécula-molécula”, que, en caso de ser mantenido en su formato original, dejaría fuera de la interacción al gen o genes relevantes. Por lo que este paso adicional asegura que todos los elementos participantes en cada ruta metabólica se tengan en cuenta, siguiendo un formato “molécula-gen-molécula” (figura 4).

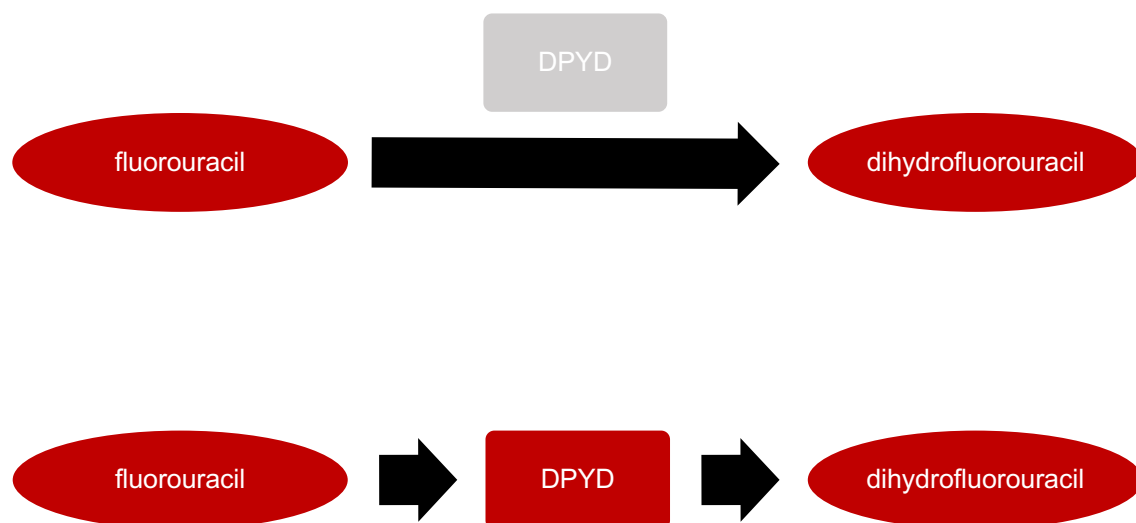


Figura 4. Pasos intermedios para interacciones "molécula-gen-molécula". DPYD: dihidropirimidina deshidrogenasa

Finalmente a partir de las tablas obtenidas, se crean los archivos de implementación .sif (que recoge las interacciones) y .att (que reúne todos los términos a manera de diccionario, junto con sus atributos).

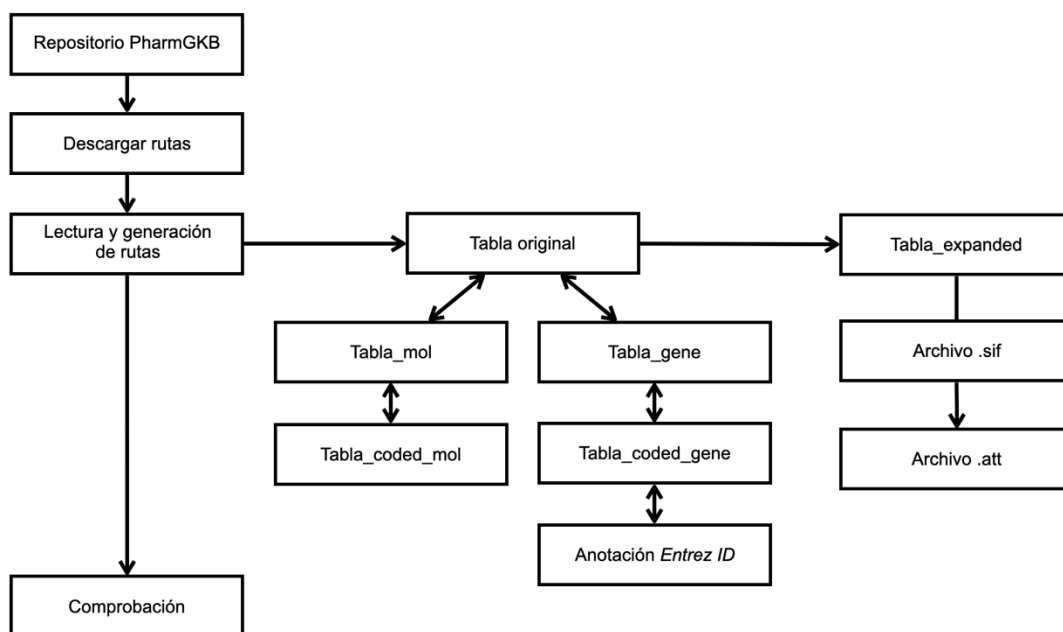


Figura 5. Proceso de parseo y creación de archivos de implementación

8.6 Integración directa de los datos transcriptómicos y genómicos

La integración de los datos genómicos y transcriptómicos en el marco de modelización de las vías de señalización propuesto es sencilla. Para transmitir la señal, una proteína tiene que estar presente (gen expresado) y ser funcional (no albergar mutaciones que la perjudiquen). Preferiblemente, la actividad de las proteínas debe inferirse a partir de experimentos. Sin embargo, la producción de este tipo de datos sigue resultando relativamente compleja. En su lugar, un enfoque muy utilizado es tomar la presencia del ARN mensajero (RNA-seq) correspondiente a la proteína como sustituto de la presencia de la proteína (Hidalgo et al., 2017).

8.6.1 Formato de los datos

En HiPathia se utilizan diferentes tipos de datos. Algunos de estos datos requieren una determinada estructura que se explica en los siguientes enlaces: Las extensiones de archivo recomendadas son '.txt' o '.tsv'.

8.6.1.1 Formato de archivo de la matriz de expresión

El archivo de la matriz de expresión es un archivo de valores separados por tabulaciones. Un archivo de valores separados por tabulaciones (TSV) es un formato de texto simple para almacenar datos en una estructura tabular. Cada registro de la tabla es una línea del archivo de texto. Cada valor de campo de un registro está separado del siguiente por una tabulación.

El archivo de matriz de expresión tiene más de dos columnas. La primera fila es una cabecera y debe contener los nombres de las muestras. La primera columna corresponde a los genes, sondas o proteínas, y se aceptan los siguientes ID: Gen Ensembl, Símbolo HGNC, Entrez id. Las siguientes columnas corresponden a los valores de expresión génica en formato numérico de cada muestra.

En nuestro caso, para probar el funcionamiento de las rutas, utilizamos 227 muestras de expresión de RNA-Seq de hígado sano provenientes de GTEx portal (dbGaP accession phs000424.v9.p2)(*Genotype-Tissue Expression (GTEx) Project*, 2022). El proyecto Genotipo-Expresión Tisular (GTEx) (*Genotype-Tissue Expression (GTEx) Project*, 2022) es un esfuerzo continuo para construir un recurso público integral para estudiar la expresión y regulación de genes específicos de los tejidos. Este repositorio cuenta muestras de tejidos no enfermos en casi 1000 individuos, principalmente para ensayos moleculares que incluyen WGS, WES y RNA-Seq.

Realizamos la lectura de los datos. Para ello, primero vamos a disponer de una serie de información asociada de cada una de las muestras para posteriormente asociar los valores de conteos a cada muestra. La anotación por muestra suele proporcionarse en la plataforma GTEx en formato TSV.

Cada una de las muestras seleccionadas de hígado sano tienen asociado un fichero de conteos en el que se contabilizan el número de lecturas (*reads*) que han alineado a cada uno de los genes. Los ficheros de conteos son también accesibles desde GTEx y tienen formato textual (.txt).

Estos ficheros se leyeron y unificaron en una única matriz de conteos con la función *readDGE* del paquete edgeR (Robinson et al., 2010). Posteriormente, los datos de expresión génica de RNA-seq se normalizaron con la media recortada de valores M (TMM) de normalización utilizando el mismo paquete (Robinson et al., 2010).

8.6.1.2 Formato del archivo de diseño experimental

El diseño experimental es un archivo de valores separados por tabulaciones. Este archivo tiene dos columnas, la primera corresponde al nombre de la muestra y la segunda al fenotipo. En nuestro caso todas las muestras pertenecían a tejido hepático sano.

8.6.1.3 Formato del archivo de la lista de genes

La lista de genes es un archivo de valores separados por tabulaciones. Este archivo sólo tiene una columna, que es el Entrez ID de los genes.

En segundo lugar, se calcula un valor de intensidad de la señal transmitida a través de una proteína, teniendo en cuenta el nivel de expresión del ARNm correspondiente y la intensidad de la señal que le llega. El valor neto de la señal transmitida a través de la vía corresponde a los valores de la señal transmitida por las últimas proteínas de la vía que finalmente desencadenan las funciones celulares activadas por la vía (Hidalgo et al., 2017). Para ello, el algoritmo HiPathia requiere algunos pasos adicionales para el cálculo de las intensidades de la señal. Así, se realiza una transformación logarítmica (aplicar $\log(\text{matriz} + 1)$) seguida de un truncamiento por el cuantil 0,99 (todos los valores mayores que el cuantil 0,99 se truncan a este valor superior, todos los valores inferiores al cuantil 0,01 se truncan a este valor inferior valor) se aplicaron a los valores de expresión génica normalizados. Finalmente, por medio de la función *create_report* del paquete de Hpathia, se genera un reporte de los valores de la ruta, generando un HTML que permite la visualización de los resultados.

8.7 Análisis de la señal

Las vías se representan mediante gráficos dirigidos, que conectan nodos de entrada (receptores) con nodos de salida (efectores). La señal llega a un nodo de entrada inicial y se transmite a lo largo de la vía siguiendo la dirección de las interacciones hasta llegar a un nodo de salida que desencadena una acción dentro de la célula. Así, desde diferentes nodos de entrada la señal puede seguir diferentes rutas a lo largo de la vía para llegar a diferentes nodos de salida (Hidalgo et al., 2017).

Los nodos de salida son los responsables últimos de llevar a cabo la acción que la señal pretende desencadenar en la célula. Entonces, desde un punto de vista funcional, un circuito efector puede definirse como una entidad de señalización de nivel superior compuesta por la colección de todos los circuitos que terminan en un único nodo de salida (efector). Cuando se aplica a los circuitos efectores, el método devuelve la intensidad conjunta de la señal que llega al nodo efector correspondiente (Hidalgo et al., 2017).

Se simula un conjunto de datos afectados a partir del conjunto de datos de control reduciendo drásticamente la expresión del gen o genes con una mutación multiplicando sus valores de expresión por 0,01 en todas las muestras de control (*Knock-out in silico*). Esto simula un gen inactivo o un producto génico no funcional. A continuación, se recalculan las actividades del circuito en el conjunto de datos afectado y se compara con el conjunto de datos de referencia. Aunque no es completamente realista, dado que el modelo no tiene información sobre la forma en que el tejido reaccionará

transcripcionalmente a la perturbación inducida por los genes mutados, se espera que los resultados señalen las funciones celulares afectadas. Esto se representa en el visualizador de rutas del servidor.

Su comprobación se realiza por medio de un análisis de agrupación jerárquico y un análisis de expresión diferencial en muestras de tejido, para comprobar el *parseo* adecuado de las rutas. Su representación se expresará por medio de figuras tipo *Heatmap* y *Violin plot*.

8.8 Implementación

El paquete R/Bioconductor implementa el algoritmo HiPathia detrás del modelo mecanístico de señalización celular (Hidalgo et al., 2017). La aplicación HiPathia de R/Bioconductor implementa el modelo que transforma la matriz de expresión de genes (o proteínas) en una matriz de señalización del circuito y sólo ofrece la posibilidad de, dada la meta-información que distingue dos clases dentro de los datos, realizar un texto de señalización diferencial (Rian et al., 2021).

8.9 Anotación de la funcionalidad de los circuitos

Los circuitos por defecto de HiPathia se anotan utilizando tanto anotaciones de Uniprot (Bateman et al., 2015) como de Gene Ontology (GO) (Carbon et al., 2019) asociadas al nodo efector. Esto permite obtener perfiles funcionales además de perfiles de señalización. La intensidad de la señal recibida por el nodo efector puede propagarse a las funciones desencadenadas por ellos siguiendo el mismo razonamiento de propagación de señales a lo largo de los circuitos (Hidalgo et al., 2017)

9 Resultados

9.1 Proceso de automatización

La generación de código que permita el *parseo* automático de las rutas disponibles en los repositorios, permite mantener actualizada la información disponible para su uso e implementación en la herramienta HiPathia.

9.1.1 Primera iteración

El objetivo de la primera iteración fue la identificación de elementos únicos para la creación de un diccionario de moléculas codificadas. Para ello, creé 4 tablas iniciales a partir de la lectura de la ruta descargadas del repositorio:

Tabla original (ejemplo en anexos): se compone inicialmente de 11 columnas, que incluyen *from*, *to*, *reaction type*, *controller*, *control*, *cell type*, *PMID*, *genes*, *drugs*, *diseases*, *summary*. Esta tabla se transforma, seleccionando los elementos indispensables necesarios para crear los archivos para implementación de HiPathia. Las columnas finales son 4 *from*, *reaction type*, *to*, *genes*.

Tabla moléculas (ejemplo en anexos): se compone de 2 columnas, *molecules*, *coded_mol*. En la ruta de Fluoropirimidinas se obtuvieron 18 moléculas, en Tiopurinas 32 moléculas y en Metotrexato 42 moléculas.

Tabla genes (ejemplo en anexos): se compone de 3 columnas, *symbol*, *entrezid*, *coded_gene*. En la ruta de Fluoropirimidinas se obtuvieron 23 genes, en Tiopurinas 26 genes y en Metotrexato 17 genes.

Tabla codificada (ejemplo en anexos): incluye 4 columnas *mapped_mol*, *reaction type*, *mapped_mol2*, *coded_gen*. Las características de las moléculas “mapeadas” es el uso de las moléculas codificadas como base para reemplazar las moléculas cada una de las reacciones bioquímicas por su equivalente codificado.

Uno de los principales inconvenientes en este paso fue la heterogeneidad de los archivos que contenían las rutas, pues se utilizaba nomenclatura diversa para referirse a las reacciones bioquímicas. Por tanto, en esta primera iteración me enfoqué en unificar la nomenclatura. Para ello las expresiones “*Biochemical reaction*”, “*Transport*”, “*Leads to*” se homogenizaron para ser interpretadas como “*activation*”. La acción opuesta, inhibición, también se homogenizó en formato como “*inhibition*”.

Otro reto se limitaba a los genes. La mayoría de los genes venían clasificados en una columna específica como “controladores” de las distintas reacciones. En un primer intento de leer los elementos, se interpretaron como elementos inseparables, lo que generaba un diccionario erróneo. Al no separar los genes como elementos únicos, tampoco era capaz de anotar los genes correctamente con su Entrez ID correspondiente.

Para la creación de código correspondiente a cada ruta, era preciso identificar el nombre de la ruta a partir del nombre del archivo. Este punto, me permitió generar adicionalmente una tabla de rutas (tabla en anexos) que asocia cada fármaco con su

número de ruta correspondiente. De esta manera, se puede identificar todas las tablas provenientes de una única ruta, así como cada molécula y gen incluidos en dicha ruta.

9.1.2 Segunda iteración

En una segunda iteración resolví el problema para alguna de las rutas. Sin embargo, la separación entre genes y moléculas tampoco mantiene el mismo formato entre rutas. Por lo que tuve que incluir varias opciones de separación (“ “ , “ , ” , “ , “) para por separar los elementos.

Tabla expandida (ejemplo en anexos): Se crea una tabla de reacciones intermedias. Esto es necesario para transformar las reacciones de las rutas desde un formato “molécula-molécula” a un formato “molécula-gen-molécula”. Esto es necesario para que HiPathia pueda integrar los genes como elementos imprescindibles para que las reacciones bioquímicas tengan lugar. Consiste por tanto en la duplicación de las reacciones en las que participan 1 o más genes. En la ruta de Fluoropirimidinas se obtuvieron 40 interacciones, en Tiopurinas 72 interacciones y en Metotrexato 64 interacciones.

9.1.3 Tercera iteración

Finalmente, algunas rutas minoritarias incluían como elemento final de la reacción bioquímica un gen, en lugar de un compuesto. Esto supuso añadir al *parseo* una condición adicional, ya que no era suficiente con filtrar por columnas, sino por características de los elementos. Todos los genes están recogidos en las rutas en formato símbolo genético (*Gene Symbol*).

Un símbolo genético es una abreviatura de un gen concreto. Las reglas de nomenclatura de los genes fueron establecidas por el Comité de Nomenclatura de la Organización del Genoma Humano (HUGO) (*HUGO Gene Nomenclature Committee*, s. f.), a cada gen humano conocido su propio nombre "largo", también le asignan un símbolo "corto". En general, exigen que los símbolos de los genes: sean únicos, contengan sólo letras latinas y números arábigos (0, 1, 2, ...), no contengan signos de puntuación. Esto permitió su identificación para ser asignados a la tabla correspondiente.

9.2 Rutas farmacocinéticas/farmacodinámicas

Con el objetivo de probar si el *parseo* se realizaba de manera correcta, y el archivo de salida obtenido podría ser integrado posteriormente en la herramienta HiPathia, se seleccionaron una serie de rutas interesantes desde el punto de vista clínico para su

interpretación, así como por ofrecer una representación de rutas de farmacodinámica (PD) y farmacocinética (PK), con suficientes tipos de interacciones, complejos y moléculas, con un número adecuado de genes.

Demostramos las posibilidades que ofrecen los modelos mecanísticos para la interpretación de la variabilidad genómica en las rutas farmacocinéticas y farmacodinámicas. Las variantes que tienen un impacto en la vía, en ausencia de otras consideraciones, se asocian potencialmente a las diferencias en la eficacia y la toxicidad derivada del uso de fármacos.

9.2.1 Fluoropirimidinas

Las fluoropirimidinas, como el 5-Fluorouracilo (5-FU), la capecitabina y tegafur, son un ejemplo de los fármacos pertenecientes a la categoría de antimetabolitos. Un mapa detallado de la señalización de la farmacocinética (PK) de las fluoropirimidinas está disponible en PharmGKB (ID: PA150653776) y con este trabajo se busca su implementación en el modelo mecanístico correspondiente en HiPathia. La figura 6 muestra el mapa básico de la cadena de transducción de señales utilizado para relacionar la expresión génica con la actividad de los circuitos de señalización dentro de la vía de la farmacocinética de las fluoropirimidinas, que en última instancia desencadenan las actividades celulares metabolismo de sus compuestos en la célula hepática, y la inhibición del ciclo del ácido fólico.

Dado que el hígado es el principal tejido involucrado en el metabolismo de las fluoropirimidinas, los datos de hígado sano disponibles en GTEx, en el que se analizó el tejido hepático de 226 donantes sanos. De estos, 208 estaban genotipados que correspondían al 92% del total. El 70% correspondía a hombres.

Muestras	Total	Mujeres	Hombres
Total	226	65 (28,7%)	161 (71,3%)
Genotipadas	208 (92%)	62 (29,8%)	146 (70,2%)

Tabla 3. Información de la muestra de RNA-seq de tejido. Obtenido de GTEx (Genotype-Tissue Expression (GTEx) Project, 2022)

La dihidropirimidina deshidrogenasa (DPD) participa en el metabolismo del 5-FU convirtiendo hasta el 80% en metabolitos inactivos y, por tanto, es responsable de su eliminación. Está codificada por el gen DPYD. Debido a la gran variedad de DPD entre individuos, se han observado diferentes efectos del tratamiento con 5-FU: eficacia, resistencia y toxicidad (Miteva-Marcheva et al., 2020)(Miteva-Marcheva et al., 2020)

Sin embargo, no todos los casos con una actividad disminuida de DPD pueden explicarse con los polimorfismos de DPYD. Varios otros genes pueden tener un efecto sobre la actividad del 5-FU: ABCB1, metileno tetrahidrofolato reductasa y timidilato sintetasa (Miteva-Marcheva et al., 2020)

Al modelizar esta ruta, se obtuvieron un total de 46 nodos, y 4 circuitos alterados: fluorodesoxiuridina trifosfato, fluorouridina trifosfato, UPB1 y TYMS (esta última siendo la única ruta regulada al alza).

Los elementos más difíciles de modelizar en todos los casos fueron aquellas moléculas que aparecen de forma repetidas dentro de las rutas, por presentar una acción relevante en distintas reacciones bioquímicas. Generalmente, este caso se asociaba a proteínas transportadoras y genes relevantes. En el caso de las fluoropirimidinas, la relación con sus proteínas transportadoras: complejo ABCC4-ABCG2, complejo ABCC3-ABCC4-ABCC5, y SLC22A7 fue complicado de modelizar por sus interacciones de entrada y salida del hepatocito, como se observa en la figura anexos 1.

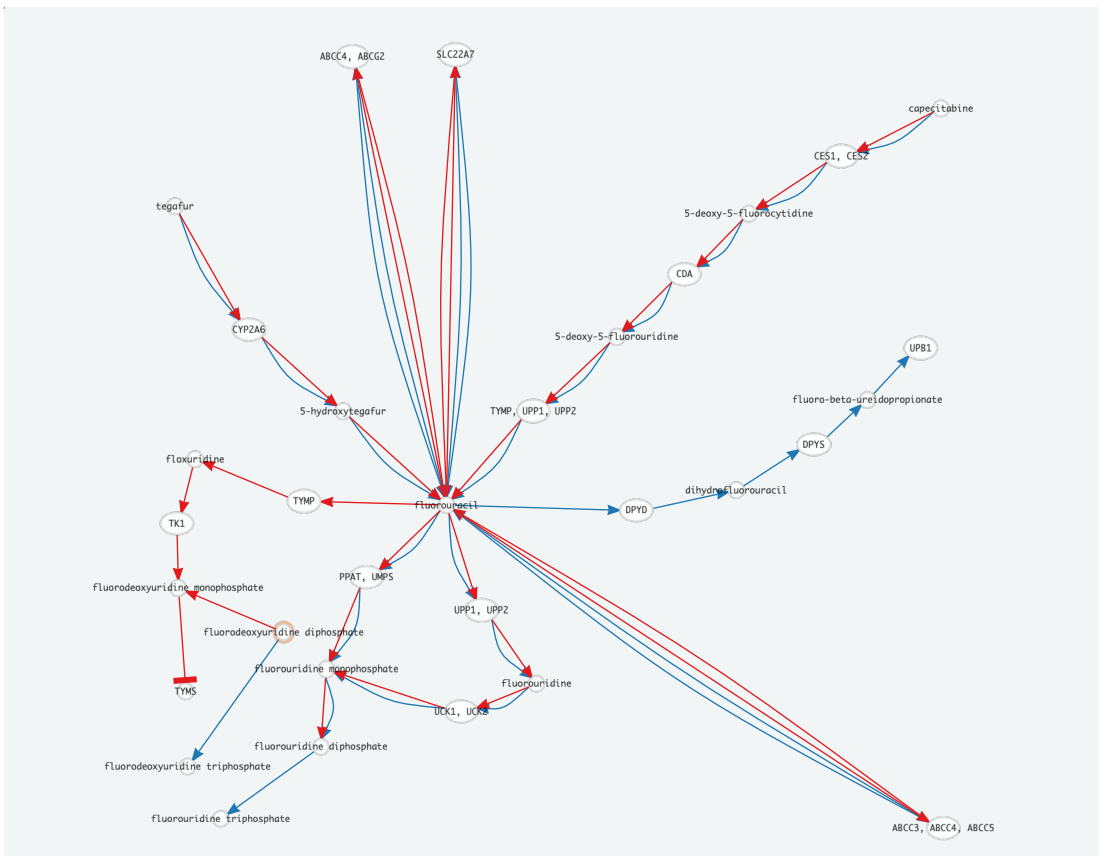


Figura 6. Ruta Fluoropirimidinas HiPathia

9.2.2 Metotrexate

El metotrexato (MTX) es un antagonista del ácido fólico. Inicialmente se usó como agente quimioterapéutico por sus propiedades antineoplásicas. Posteriormente demostró efectos inmunomoduladores y antiinflamatorios.

Un mapa detallado de la señalización de la farmacocinética y farmacodinámica del Metotrexate está disponible en PharmGKB (ID: PA2039) y su implementación en HiPathia se puede observar en la figura 7.

Al modelizar esta ruta, se obtuvieron un total de 80 nodos, y 10 circuitos: adenosina trifosfato, apoptosis, cistationina, deoxitimidina monofosfato, hipoxantina, l.serina, GART, PPAT, DNA metilado, proteínas metiladas.

Sin duda la más difícil de *parsear* por su gran tamaño, y por la gran densidad de compuestos metabólicos con múltiples interacciones.

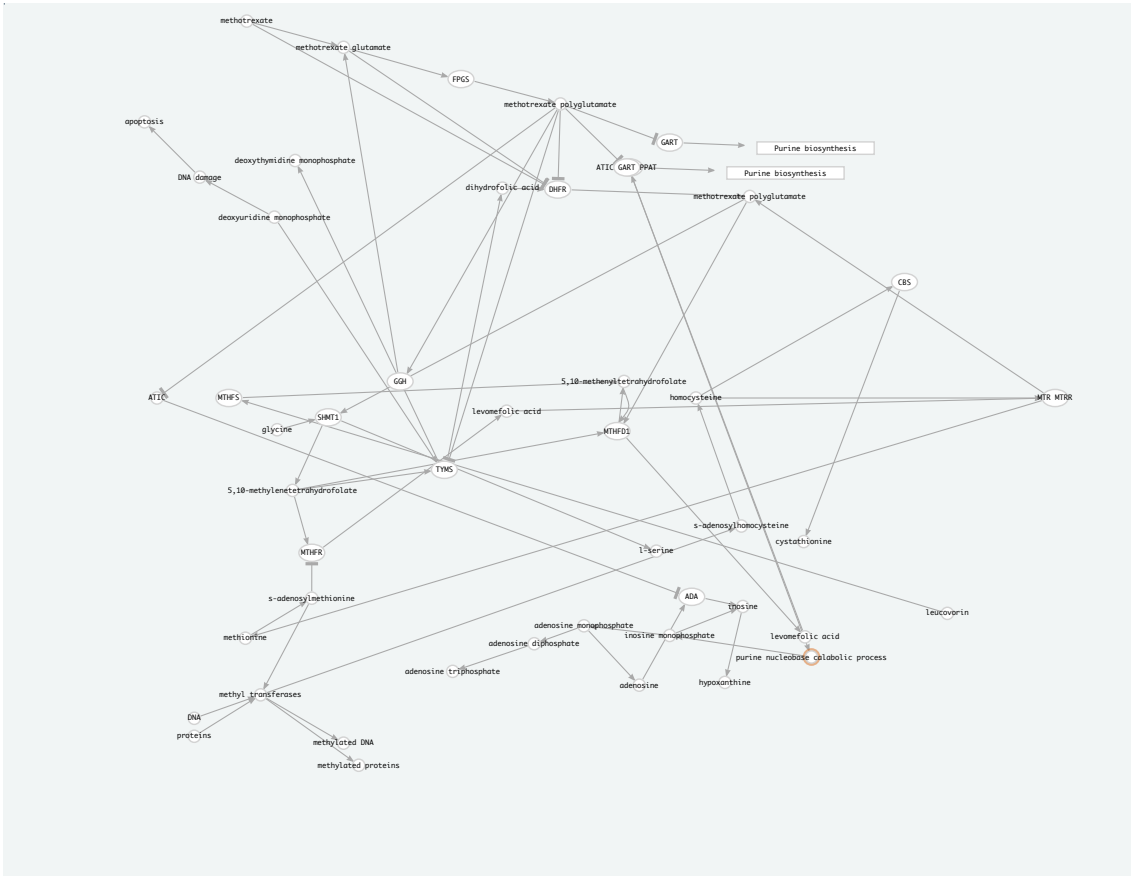


Figura 7. Ruta Metotrexate HiPathia

9.2.3 Tiopurinas

Las tiopurinas de uso clínico más frecuente son: azatioprina (AZA) (un profármaco de mercaptopurina), 6-Mercaptopurina (6MP) y 6-Tioguanina (6TG) (esta última no está actualmente comercializada en España). Se emplean para el tratamiento de diversas patologías autoinmunes, trasplantes o cáncer.

Un mapa detallado de la farmacocinética y farmacodinámica de las tiopurinas está disponible en PharmGKB (ID: PA2040) y su representación se puede observar en la figura 8.

La actividad de esta enzima está sujeta a una importante variabilidad interindividual asociada a variantes genéticas. En este contexto, TPMT es la enzima más estudiada en la eficacia de las tiopurinas y el gen TPMT el biomarcador farmacogenético más relevante para la predicción de respuesta al tratamiento con estos fármacos.

Más recientemente, se ha descrito que variantes genéticas en el gen NUDT15 están asociadas a mielosupresión (toxicidad) durante el tratamiento con tiopurinas.

Al modelizar esta ruta, se obtuvieron un total de 87 nodos, y 13 circuitos alterados: 7 rutas reguladas a la baja: 8-hidroxitioguanina, desoxy-tioguanosina monofosfato, ribosyltioguanina, tiinosina trifosfato, ácido tiúrico así como las relacionadas con efectos tóxicos como el daño hepático y la apoptosis. 4 rutas sin regulación diferencial: metil-tioguanosina monofosfato, metil-tiinosina monofosfato, metilmercaptopurina, fosforibosilamina. Finalmente 2 rutas reguladas al alza: PPAT y RAC1.

Los elementos más difíciles de modelizar en esta ruta fueron las proteínas transportadoras: complejo ABCC4-ABCC5, los genes TPMT (uno de los principales de la ruta), y las moléculas tiinosina monofosfato, tiioxantina monofosfato, tioguanosina monofosfato (elementos finales de alguno de los circuitos)

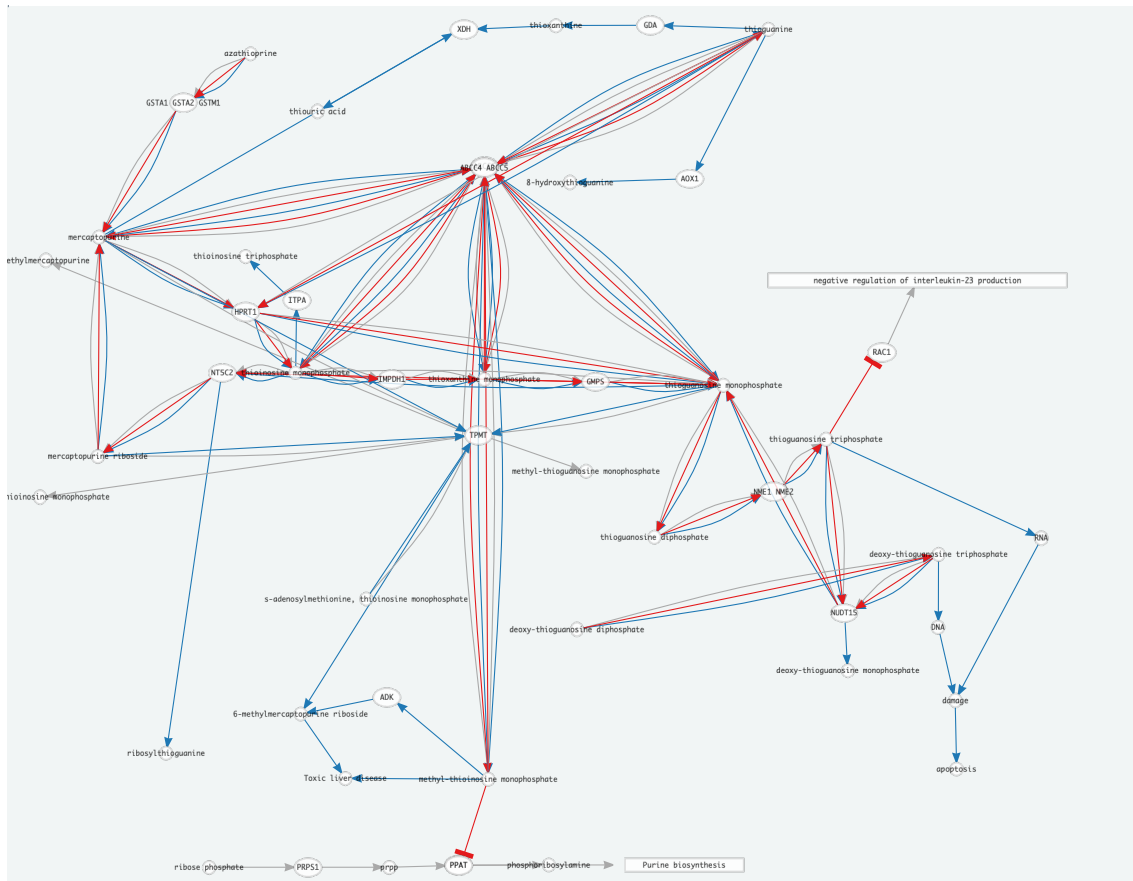


Figura 8. Ruta Tiopurinas HiPathia

4.3 Análisis de la variabilidad de rutas farmacogenéticas en tejido sano

Con el objetivo de evaluar las rutas *parseadas*, se emplearon datos de expresión procedentes de 208 muestras de tejido del proyecto GTEx, concretamente datos de hígado sano. Al realizar el análisis de expresión diferencial de las rutas, obtuvimos un total de 26 circuitos con expresión diferencial a lo largo de las 3 rutas. Para su estudio, primero se generó un mapa de agrupación jerárquica utilizando un *Heatmap*. Como se observa en la figura 9, 3 circuitos destacan como diferencialmente activados los 9 circuitos que presentan una mayor desviación estándar (*sd*) respecto a la media (*mean*) de los valores de activación: P-PA2040-22, PA150653776-15, PA150653776-16, P-PA2040-32, PA150653776-17, P-PA2040-29, P-PA2040-30, PA150653776-011, P-PA2040-27.

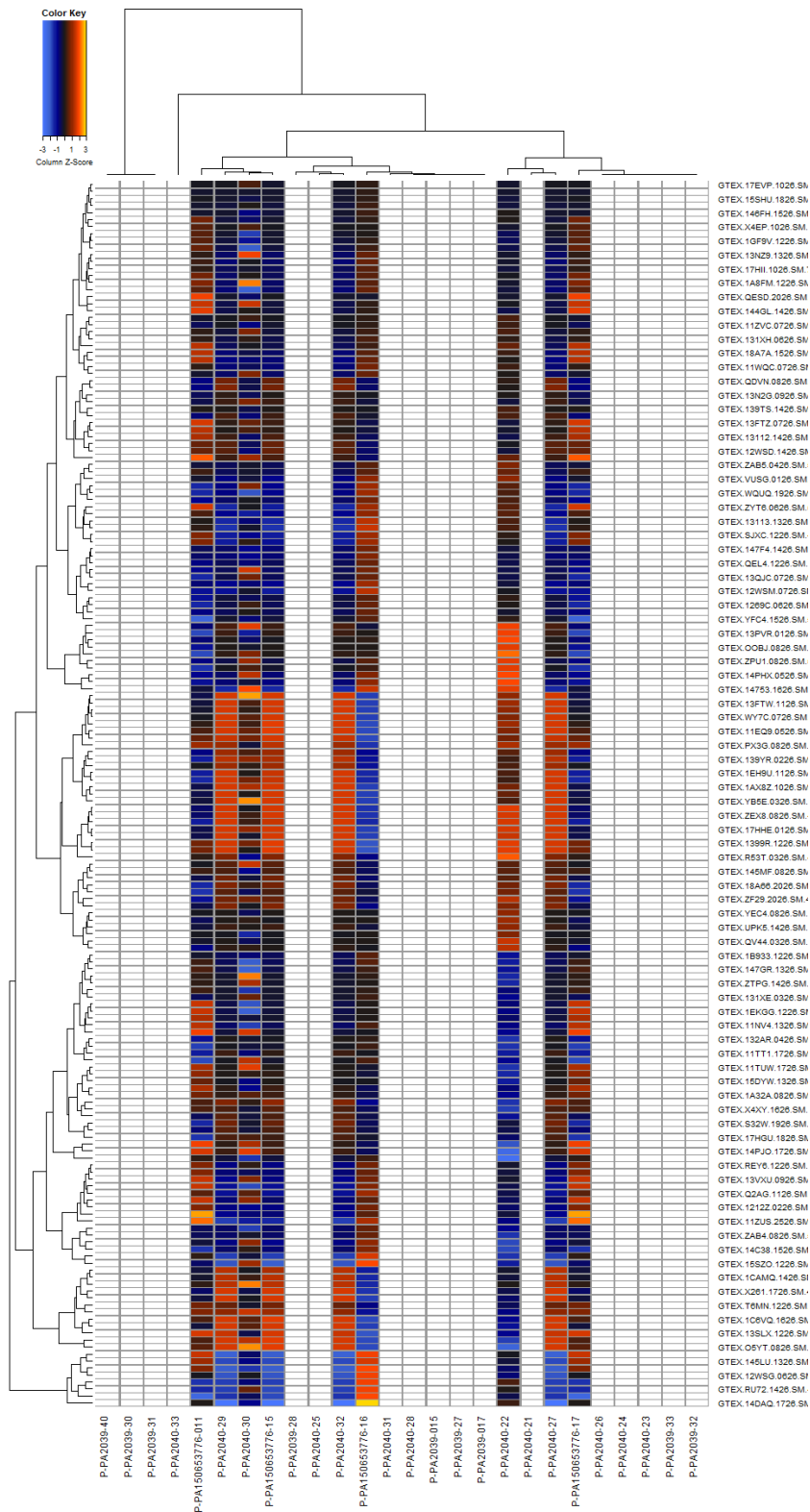


Figura 9. Heatmap. Agrupación jerárquica de activación diferencial de rutas

Para facilitar su interpretación, se incluye la tabla que incluye los valores para la generación del *Heatmap*, ordenados de forma descendente a partir de la desviación estándar (*sd*).

Ruta	Efactor (Entrez ID)	Anotación	mean	sd	min	max
P-PA2039-30	Metotrexate	methylated proteins	1	1	0	0
P-PA2039-31	Metotrexate	methylated DNA	1	1	0	0
P-PA2039-40	Metotrexate	apoptosis	1	1	0	0
P-PA2039-28	Metotrexate	hypoxanthine	0.118095800877748	6.96E-03	6.96E-03	0.118095800877748
P-PA2040-25	Tropurinas	thioinosine triphosphate	0.101682728885231	8.35E-03	8.35E-03	0.101682728885231
P-PA2039-27	Metotrexate	deoxythymidine monophosphate	0.0751639765065324	9.74E-03	9.74E-03	0.0751639765065324
P-PA2039-017	Metotrexate	PPAT (5471)	0.0751639765065324	9.74E-03	9.74E-03	0.0751639765065324
P-PA2039-015	Metotrexate	GART (2618)	0.0751639765065324	9.74E-03	9.74E-03	0.0751639765065324
P-PA2040-28	Tropurinas	phosphoribosylamine	0.0751639765065324	9.74E-03	9.74E-03	0.0751639765065324
P-PA2040-31	Tropurinas	PPAT (5471)	0.0696069456619318	2.51E-02	2.51E-02	0.0696069456619318
P-PA2039-33	Metotrexate	cystathionine	0.274160494066036	8.35E-02	8.35E-02	0.274160494066036
P-PA2039-32	Metotrexate	L-serine	0.274160494066036	8.35E-02	8.35E-02	0.274160494066036
P-PA2040-23	Tropurinas	methyl-thioguanosine monophosphate	0.274160494066036	8.35E-02	8.35E-02	0.274160494066036
P-PA2040-24	Tropurinas	methyl-thioinosine monophosphate	0.274160494066036	8.35E-02	8.35E-02	0.274160494066036
P-PA2040-26	Tropurinas	methylmercaptopyrimine	0.274160494066036	8.35E-02	8.35E-02	0.274160494066036
P-PA2040-21	Tropurinas	thiouric acid	0.17396513585214	8.35E-02	8.35E-02	0.17396513585214
P-PA2040-33	Tropurinas	Toxic liver disease	0.493142191894028	1E-01	1E-01	0.493142191894028
P-PA2040-27	Tropurinas	ribosylthioguanine	0.18197115246782	0.000146377838824262	0.000146377838824262	0.18197115246782
P-PA150653776-011	Fluoropirimidinas	UPB1 (51733)	0.0154176442743483	0.000287286251643787	0.000287286251643787	0.0154176442743483
P-PA2040-30	Tropurinas	RAC1 (5879)	0.0461747556697235	0.000849353653399571	0.000849353653399571	0.0461747556697235
P-PA2040-29	Tropurinas	deoxy-thioguanosine monophosphate	0.0358861817219986	0.00291453075854708	0.00291453075854708	0.0358861817219986
P-PA150653776-17	Fluoropirimidinas	fluorouridine triphosphate	0.248828880725206	0.0044308297984183	0.0044308297984183	0.248828880725206
P-PA2040-32	Tropurinas	apoptosis	0.104790614276434	0.0054607121192446	0.0054607121192446	0.104790614276434
P-PA150653776-16	Fluoropirimidinas	TYMS (7298)	0.0829283188132639	0.00776203338635881	0.00776203338635881	0.0829283188132639
P-PA150653776-15	Fluoropirimidinas	fluorodeoxyuridine triphosphate	0.0498517863036251	0.00978438423428264	0.00978438423428264	0.0498517863036251
P-PA2040-22	Tropurinas	8-hydroxythioguanine	0.205668844810337	0.0135670736487975	0.0135670736487975	0.205668844810337

Tabla 4. Resumen circuitos HiPathia (ordenados descendente por Desviación estándar). Mean: media, sd: desviación estándar, min: valor mínimo, max: valor máximo.

Especial mención merecen los circuitos: P-PA2039-40. P-PA2039-30. P-PA2039-31 que aparecen al inicio de la tabla con valores. Estos tres circuitos presentan la menor variabilidad (desviación estándar 0), lo que en este caso significa pasa total de la señal de activación sin resistencias (media 1). Probablemente estos resultados estén relacionados con dificultades de *parseo* debidas a la alta densidad de metabolitos, y de interacciones metabolito-metabolito presentes en esta ruta en particular.

9.2.4 Análisis de expresión génica en tejido

Para validar el *parseo*, se realizó un análisis de expresión diferencial en las muestras de tejido sano hepático. Este análisis permite validar los resultados de activación obtenidos por circuito. Para facilitar su visualización, se incluyen los datos de expresión separados por rutas: Fluoropirimidinas (figura 10), Metotrexato (figura anexo 31), Thiopurinas (figura anexo 32)

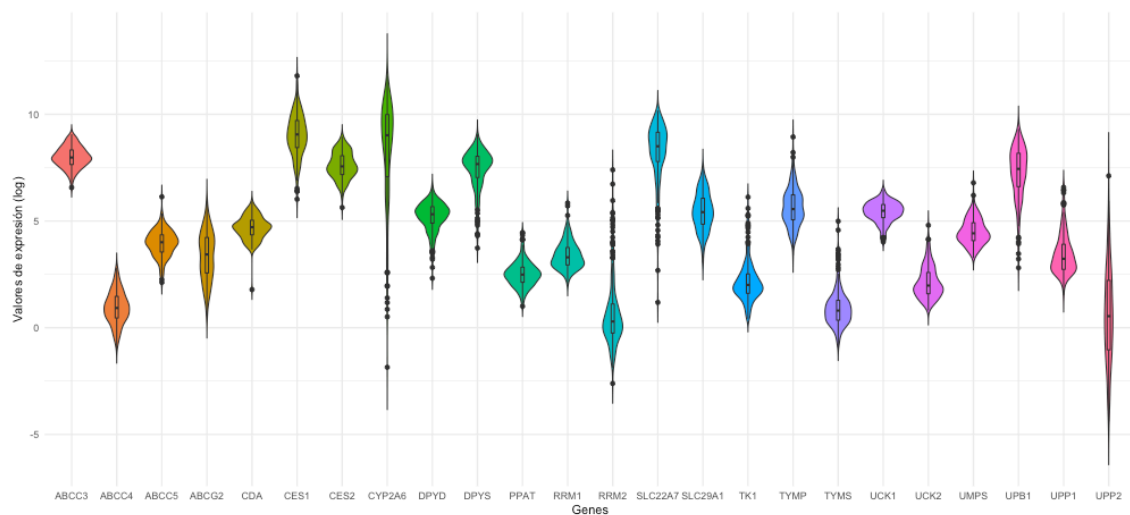


Figura 10. Violin plot. Análisis de expresión de genes relacionados con la ruta de las Fluoropirimidinas en tejido hepático sano

A continuación, se presenta la correlación de genes presentes en cada uno de los circuitos, separados por rutas.

Ruta	PA150653776
Circuito	Genes presentes
P-PA150653776-15	ABCC3, ABCC4, ABCC5, ABCG2, CDA, CES1, CES2, CYP2D6, SLC2A7, TYMP, UCK1, UCK2, UPP1, UPP2
P-PA150653776-17	ABCC3, ABCC4, ABCC5, ABCG2, CDA, CES1, CES2, CYP2D6, SLC2A7, TYMP, UCK1, UCK2, UPP1, UPP2
P-PA150653776-16	ABCC3, ABCC4, ABCC5, ABCG2, CDA, CES1, CES2, CYP2D6, SLC2A7, TK1, TYMP, TYMS, UCK1, UCK2, UPP1, UPP2
P-PA150653776-011	ABCC3, ABCC4, ABCC5, ABCG2, CDA, CES1, CES2, CYP2D6, DPYD, DPYS, SLC2A7, TYMP, UPB1, UPP1, UPP2
Ruta	PA2039
Circuito	Genes presentes
P-PA2039-40	
P-PA2039-33	CBS, DHFR, FPGS, GGH, MTHFR, MTR, MTRR, SHMT1, TYMS
P-PA2039-27	DHFR, FPGS, GGH, MTHFR, MTR, MTRR, SHMT1, TYMS
P-PA2039-015	FPGS, GART
P-PA2039-28	ADA, ATIC, DHFR, FPGS, GART, GGH, MTHFD1, MTHFR, MTHFS, MTR, MTRR, PPAT, SHMT1, TYMS
P-PA2039-32	ADA, ATIC, DHFR, FPGS, GART, GGH, MTHFD1, MTHFR, MTHFS, MTR, MTRR, PPAT, SHMT1, TYMS
P-PA2039-31	DHFR, FPGS, GGH, MTHFR, MTR, MTRR, SHMT1, TYMS
P-PA2039-30	DHFR, FPGS, GGH, MTHFR, MTR, MTRR, SHMT1, TYMS
P-PA2039-017	FPGS, PPAT
Ruta	PA2040
Circuito	Genes presentes
P-PA2040-22	ABCC4, ABCC5, AOX1, GMPS, GSTA1, GSTA2, GSTM1, HPRT1, IMPDH1, NME1, NME2, NTSC2, NUDT15
P-PA2040-32	ABCC4, ABCC5, GMPS, GSTA1, GSTA2, GSTM1, HPRT1, IMPDH1, NME1, NME2, NTSC2, NUDT15
P-PA2040-29	ABCC4, ABCC5, GMPS, GSTA1, GSTA2, GSTM1, HPRT1, IMPDH1, NME1, NME2, NTSC2, NUDT15
P-PA2040-23	ABCC4, ABCC5, GMPS, GSTA1, GSTA2, GSTM1, HPRT1, IMPDH1, NME1, NME2, NTSC2, NUDT15, TPMT
P-PA2040-24	ABCC4, ABCC5, GMPS, GSTA1, GSTA2, GSTM1, HPRT1, IMPDH1, NME1, NME2, NTSC2, NUDT15, TPMT
P-PA2040-26	ABCC4, ABCC5, GMPS, GSTA1, GSTA2, GSTM1, HPRT1, IMPDH1, NME1, NME2, NTSC2, NUDT15, TPMT
P-PA2040-28	PPAT, PRPS1
P-PA2040-31	ABCC4, ABCC5, GMPS, GSTA1, GSTA2, GSTM1, HPRT1, IMPDH1, NME1, NME2, NTSC2, NUDT15, PPAT
P-PA2040-30	ABCC4, ABCC5, GMPS, GSTA1, GSTA2, GSTM1, HPRT1, IMPDH1, NME1, NME2, NTSC2, NUDT15, RAC1
P-PA2040-27	ABCC4, ABCC5, GMPS, GSTA1, GSTA2, GSTM1, HPRT1, IMPDH1, NME1, NME2, NTSC2, NUDT15
P-PA2040-25	ABCC4, ABCC5, GMPS, GSTA1, GSTA2, GSTM1, HPRT1, IMPDH1, ITPA, NME1, NME2, NTSC2, NUDT15
P-PA2040-21	ABCC4, ABCC5, GDA, GMPS, GSTA1, GSTA2, GSTM1, HPRT1, IMPDH1, NME1, NME2, NTSC2, NUDT15, XDH
P-PA2040-33	ADK, ABCC4, ABCC5, GMPS, GSTA1, GSTA2, GSTM1, HPRT1, IMPDH1, NME1, NME2, NTSC2, NUDT15, TPMT

Tabla 5. Relación Circuito-Genes presentes, pertenecientes a cada una de las rutas

10 Discusión

La farmacogenómica es el aspecto de la medicina personalizada, y de la genómica clínica, que verá su implementación en la práctica clínica en el futuro cercano, formando parte del tratamiento racional guiado por evidencias de todos los pacientes.

El rápido aumento de conocimiento farmacogenético clínicamente relevante en los últimos años pone de manifiesto el reto que supone su implementación efectiva en la práctica clínica habitual. La investigación en este campo ha conducido a la identificación, caracterización, validación e implementación de docenas de biomarcadores genéticos para mejorar la selección y eficacia de los medicamentos y prevenir las reacciones adversas a los mismos (Gonzalez-Covarrubias et al., 2020).

La llegada de las tecnologías genómicas ha facilitado la investigación de los genes que influyen en la seguridad y eficacia de los medicamentos y su enlace con los ensayos de fenotipado. En la actualidad, es posible utilizar la información farmacogenética para personalizar la prescripción de más de 40 fármacos mediante un variado conjunto de plataformas genómicas, como el genotipado con sondas, los *microarrays*, la secuenciación tipo Sanger y la secuenciación de nueva generación. Estos marcadores incluyen polimorfismos de un solo nucleótido, variaciones en el número de copias, inserciones o deleciones cortas y repeticiones en tándem de número variable, que afectan a la farmacodinámica o la farmacocinética de un fármaco. En comparación con la farmacodinamia, la farmacocinética se ha caracterizado más intensamente, por lo que es posible estratificar a los pacientes en cuatro fenotipos: fenotipos lentos, intermedios, normales o ultrarrápidos (Gonzalez-Covarrubias et al., 2020).

Las grandes agencias reguladoras del medicamento, *Food and Drug Administration* (FDA) y *European Medicines Agency* (EMA), han desarrollado diferentes documentos dirigidos a impulsar el desarrollo de implementación de estrategias farmacogenómicas en investigación clínica (Medicines Agency, 2018).

De los 261 pares fármaco-gen, el 35% son para la categoría de fármacos quimioterápicos, seguidos en proporción por antipsicóticos, cardiovasculares, metabólicos, gastroenterología, hematología, urología y enfermedades autoinmunes. Los marcadores farmacogenéticos más comunes, se localizan en los genes CYP2D6, CYP2C19, CYP2C9 y UGT1A1, e influyen en al menos 20 fármacos diferentes (Gonzalez-Covarrubias et al., 2020).

A partir de estas asociaciones, organismos como el Consorcio para la implementación clínica de la farmacogenética (CPIC, por sus siglas en inglés) publica guías para la interpretación los resultados de las pruebas genéticas disponibles, incluyendo un sistema de gradación de los niveles de evidencia de cada recomendación de prescripción derivada de la relación genotipo-fenotipo (Caudle et al., 2014).

10.1 HiPathia y sus usos

Los métodos MPA constituyen una evolución de los métodos de análisis de vías en los que éstas se descomponen en subvías o circuitos elementales que potencialmente dan cuenta de los resultados celulares que pueden ayudar a explicar las características mecanísticas de los fenotipos (mecanismo de la enfermedad, mecanismo de acción de los fármacos, etc.). Aquí ofrecemos una revisión de tres casos de uso para el análisis de rutas farmacocinéticas y farmacodinámicas. Puede concluirse que una definición

biológicamente realista de los circuitos dentro de las vías analizadas es un factor determinante del éxito del método. Sin embargo, la metodología de puntuación de cada una de las interacciones presentes en un circuito debe ser representativa de la actividad biológica de la célula. Así, el método de propagación utilizado por HiPathia parece ser la solución eficaz para la interpretación de las mediciones genómicas.

Por lo tanto, los modelos mecanísticos proporcionan un marco preciso para la interpretación de las consecuencias de la variación genética encontrada en el exoma completo o secuenciación del genoma, especialmente en el caso de rutas o circuitos de activación celular complejos, en las que el efecto puede depender en gran medida de la condición específica de cada paciente (Peña-Chilet et al., 2019).

A continuación, discutiremos una de las rutas utilizadas para el modelado.

10.2 Fluoropirimidinas

Se utilizan ampliamente en el tratamiento de muchos tipos de cáncer, como el colorrectal, cáncer gástrico y tumores gastrointestinales avanzados o recurrentes (incluyendo esófago y páncreas). También se utilizan en carcinoma metastásico de mama y en el tratamiento de tumores de cabeza y cuello (estadios III y IV) (Longley et al., 2003).

El mecanismo de acción de estos agentes radica en su unión irreversible a la enzima timidilato sintasa (TYMS), esencial para la síntesis del nucleótido timina. La carencia de este nucleótido implica que el ADN no se pueda replicar y, por lo tanto, interfiere en la división celular y el crecimiento tumoral (Miteva-Marcheva et al., 2020).

Los fármacos inhibidores de pirimidinas son i) 5- Fluorouracilo (5- FU) que se administra por vía intravenosa y el 80% se metaboliza en el hígado. ii) Capecitabina un profármaco de 5- FU, que se administra por vía oral y atraviesa sin modificaciones la pared intestinal. Se transforma en el hígado a 5'-desoxi- 5-fluorocitidina (5'dFCR) y, posteriormente, en 5´desoxi-5-fluorouridina (5'dFUR), gracias a las enzimas Carboxilesterasas (CES) y Citidina deaminasa (CDA), respectivamente. Finalmente, el 5'dFUR se trasforma en 5-FU gracias a la vía de la timidina fosforilasa (TYMP) o la uridina fosforilasa (UPP). iii) Tegafur: es un profármaco de 5- FU administrado por vía oral. El enzima CYP2D6 lo convierte en un compuesto intermedio inestable (5-hidroxitegafur) que espontáneamente se descompone a 5- FU.

Existe un enzima limitante en la degradación de 5-FU, la dihidropiridina deshidrogenasa (DPD) codificada por el gen DPYD. Los efectos secundarios más comunes en los

pacientes con deficiencia de la enzima DPD son mielosupresión, mucositis, neurotoxicidad y diarrea. (Miteva-Marcheva et al., 2020)

Sin embargo, no todos los casos toxicidad pueden explicarse por una actividad disminuida de DPD secundaria a polimorfismos de DPYD. Varios genes podrían tener un efecto sobre la actividad del 5-FU: ABCB1, metileno tetrahidrofolato reductasa y timidilato sintetasa (Miteva-Marcheva et al., 2020). Nuestro estudio sugiere que otros elementos deben ser tenidos en cuenta: TYMS, UPB1, que hasta ahora no hacen parte de los estudios farmacogenéticos rutinarios empleados previo a la utilización de 5-FU.

10.2.1 TYMS

Cataliza la metilación reductora de la 2'-desoxiuridina 5'-monofosfato (dUMP) a timidina 5'-monofosfato (dTMP), utilizando el cosustrato, el 5,10-metilentetrahidrofolato (CH₂H₄folato) como donante de 1 carbono y reductor y contribuye a la vía de biosíntesis mitocondrial de novo del timidilato (Phan et al., 2001).

El TYMS cataliza la metilación del dUMP a dTMP. Como única fuente *de novo* de timidilato en la célula, es una importante diana para el 5-fluorouracilo (5-FU) y la capecitabina (CAP). Existe evidencia en la literatura que la sobreexpresión de TYMS se ha relacionado con la resistencia a estos agentes tanto *in vitro* como *in vivo*. La causa de la variabilidad en la expresión de TYMS no está clara, sin embargo, se han descrito previamente polimorfismos en las regiones no traducidas 5' y 3' (5'UTR y 3'UTR) del gen TYMS y se ha sugerido que son tanto predictivos de la toxicidad como pronósticos de la eficacia con fluoropirimidinas (Saif et al., 2020).

10.2.2 UPB1

Este gen codifica una proteína que pertenece a la familia de las hidrolasas. La beta-ureidopropionasa cataliza el último paso en la vía de degradación de las pirimidinas. Las bases pirimidínicas uracilo y timina se degradan mediante la acción consecutiva de la dihidropirimidina deshidrogenasa (DPD), la dihidropirimidasa (DHP) y la beta-ureidopropionasa (UP) a beta-alanina y ácido beta-aminoisobutírico, respectivamente (Saif et al., 2020).

En un estudio realizado por Vreken (Vreken et al., 1999), se detectaron nueve variantes de UPB1 en la subpoblación de pacientes con cáncer que desarrollaron toxicidad grave tras el tratamiento. Un análisis de las variantes de UPB1 sobre la toxicidad relacionada con el 5-FU en la población de todos los pacientes analizados reveló una asociación entre la variante c.-80C > G (rs2070474) y la toxicidad gastrointestinal. Se encontró una fuerte correlación positiva entre los portadores del genotipo c.-80 GG y el desarrollo de mucositis grave (grado 3-4). Sin embargo, dado que el papel de las alteraciones de

UPB1 es probablemente menos significativo que el de las alteraciones de DPYD, hasta ahora tampoco se ha empleado en los estudios previos a la administración de 5- FU.

10.3 Tiopurinas

Las tiopurinas son fármacos inmunosupresores empleados en aquellos procesos en los que es necesario modificar la respuesta inmunitaria de los pacientes. Además, azatioprina está también indicada para el tratamiento de enfermedades como la artritis reumatoide, enfermedad inflamatoria intestinal, lupus eritematoso sistémico, dermatomiositis y polimiositis entre otras. Al tratarse de agentes citotóxicos estos fármacos cuentan con un estrecho margen terapéutico pudiendo causar reacciones adversas graves, principalmente mielosupresión (Wang & Weinshilboum, 2006).

La principal ruta de inactivación de las Tiopurinas a nivel intracelular es la metilación catalizada por la enzima Tiopurina S-metiltransferasa (TPMT). Esta enzima produce la inactivación de 6MP y 6TG disminuyendo así la cantidad de nucleótidos de tioguanina (TGN) TGTP y TdGTP disponibles para ejercer la acción citotóxica del fármaco. De este modo, existe una relación inversamente proporcional entre la actividad de la enzima TPMT y la cantidad de metabolitos activos y por tanto la toxicidad del fármaco (Zaza et al., 2010).

Esta enzima también participa en la metilación de 6MP para dar lugar a sus metabolitos secundarios MetTIMP y MetMPN parcialmente activos. La actividad de esta enzima está sujeta a una importante variabilidad interindividual asociada a variantes genéticas. En este contexto, TPMT es la enzima más estudiada en la eficacia de las tiopurinas y el gen TPMT el biomarcador farmacogenético más relevante para la predicción de respuesta al tratamiento con estos fármacos. En torno al 90 % de los individuos europeos tienen actividad completa, en torno al 9 % parcial y un bajo porcentaje (0.2%-0.3%) tienen actividad nula (Gisbert & Gomollón, 2008).

Más recientemente, se ha descrito que variantes genéticas en el gen NUDT15 están asociadas a mielosupresión durante el tratamiento con tiopurinas. La encima nudix hidrolasa 15 (NUDT15) codifica un nucleótido difosfatasa específica de purina que desfosforila los metabolitos activos de tiopurina TGTP y TdGTP, evitando así su incorporación al ADN y afectando negativamente los efectos citotóxicos deseados de las tiopurinas. Defectos en la degradación de TGTP por NUDT15 resultan en mayor TGTP disponible para su incorporación al ADN. En el caso de NUDT15 se establecen también tres fenotipos farmacogenéticos: metabolizador normal (98%), metabolizador

intermedio (0.8%) y metabolizador lento (frecuencia muy baja en europeos) (Moriyama et al., 2016).

Los pacientes con deficiencia de TPMT y NUDT15, especialmente aquellos con una deficiencia completa, tienen un alto riesgo de desarrollar mielosupresión severa (que puede llegar en ocasiones a ser fatal) (Yang et al., 2014). Nuestro trabajo propone el estudio de otros dos elementos relevantes en esta ruta: PPAT y RAC1.

10.3.1 PPAT

La proteína codificada por este gen es un miembro de la familia de la purina/pirimidina fosforribosiltransferasa. Esta proteína es una enzima alostérica reguladora que cataliza el primer paso de la biosíntesis *de novo* de los nucleótidos de purina. Este gen y el PAICS/AIRC, una enzima bifuncional que cataliza los pasos seis y siete de la vía de biosíntesis de los nucleótidos de purina, se encuentran localizados en el cromosoma 4 (Zalkin & Dixon, 1992).

10.3.2 RAC1

Este gen localizado en el cromosoma 7 codifica una GTPasa asociada a la membrana plasmática que alterna entre los estados activo unido a GTP e inactivo unido a GDP. En su estado activo, se une a una variedad de proteínas efectoras para regular respuestas celulares como los procesos de secreción, la fagocitosis de células apoptóticas, la polarización de las células epiteliales, la adhesión de las neuronas, la migración y la diferenciación inducida por factores de crecimiento (Marei et al., 2016).

La diversidad del funcionamiento de Rac1 en las células se atribuye principalmente a su capacidad de unirse a una multitud de efectores descendentes tras su activación por los factores de intercambio de nucleótidos de guanina (GEF) (Marei et al., 2016).

El efecto inmunosupresor de las tiopurinas se basa principalmente en el bloqueo del sustrato 1 de la toxina botulínica C3 relacionado con Ras (Rac1) que provoca la apoptosis de los linfocitos T mediante la inhibición del factor de transcripción descendente fosforilado Transductor de señales y activador de la transcripción 3 (pSTAT3). Por lo que recientemente se ha propuesto como posible marcador farmacodinámico funcional en los linfocitos T para predecir el resultado terapéutico de la terapia con tiopurinas (Deben et al., 2022).

10.4 Metotrexato

El Metotrexato (MTX) es un antagonista del ácido fólico. Inicialmente se usó como agente quimioterapéutico por sus propiedades antineoplásicas. Posteriormente demostró efectos inmunomoduladores y antiinflamatorios, por lo que se indica para el tratamiento de la psoriasis, artritis psoriática, artritis reumatoide, artritis idiopática juvenil, enfermedad de Crohn, colitis ulcerosa, síndrome de Reiter o en la profilaxis de la enfermedad injerto contra huésped en el trasplante alogénico de medula ósea (Cronstein, 1996).

Presenta un mecanismo de acción complejo, ya que interacciona con numerosas dianas y, además, es sustrato de numerosos transportadores que determinan su farmacocinética: Tras entrar en la célula, es transformado en poliglutamatos (MTX-PG) mediante la folatopoliglutamato sintasa (FPGS) que permanecen estables en el citoplasma. Estos compuestos (MTX-PG) inhiben la dihidrofolato reductasa (DHFR), la enzima encargada de reducir el ácido fólico a ácido tetrahidrofólico (THF). El THF es necesario para sintetizar unos derivados de un carbono que son necesarios en la síntesis de timidilato y de nucleótidos purínicos. La inhibición de DHFR por parte de MTX reduce los niveles de THF en el citoplasma, lo que da lugar a una reducción de estos compuestos. Como consecuencia, se inhibe la síntesis y reparación de ácidos nucleicos y se altera el ciclo celular (Mikkelsen et al., 2011).

El MTX también inhibe la proteína bifuncional de biosíntesis de proteínas, PURH (codificada por el gen ATIC), enzima necesaria a otro nivel para la síntesis *de novo* de purinas. Las células que se dividen a gran velocidad, como las células tumorales, son más sensibles a la inhibición de DHFR del MTX, lo que explica sus propiedades antineoplásicas. Las propiedades inmunosupresoras de MTX podrían deberse a la inhibición de la proliferación de los linfocitos, lo que explicaría su indicación para pacientes con artritis reumatoide (Schmiegelow, 2009).

Existen numerosos estudios farmacogenéticos con Metotrexate con nivel de evidencia moderado-alto (PharmGKB niveles 2A y 2B). Sin embargo, la implementación de la farmacogenética de Metotrexate en la práctica clínica es limitada. El CPIC considera genes de alta importancia relacionados con esta: ABCB1, SLCO1B1, SLC19A1 y MTHFR (Schmiegelow, 2009). Nosotros presentamos adicionalmente PPAT (explicado en el apartado 10.3.1) y GART.

10.4.1 GART

La proteína codificada por este gen es un polipéptido trifuncional. Tiene actividad fosforibosilglicinamida formiltransferasa, fosforibosilglicinamida sintetasa,

fosforibosilaminoimidazol sintetasa que se requiere para la biosíntesis *de novo* de purinas. El empalme alternativo de este gen da lugar a dos variantes de transcripción que codifican diferentes isoformas. Este gen GART se localiza en el cromosoma 21 (Gnirke et al., 1991).

En un estudio elaborado por Zarza, (Zaza et al., 2004) se propone que en tejidos con alto recambio, una menor expresión de este gen podría estar relacionado con una síntesis de purinas *de novo* significativamente menor y una expresión diferencial de los genes implicados en el metabolismo de las purinas, por lo que podría estar relacionado con la eficacia del tratamiento.

11 Limitaciones

El modelo presentado en este trabajo pretende explicar el funcionamiento de rutas relacionadas con el mecanismo de acción de fármacos, teniendo en cuenta la actividad inferida de los genes participantes de los circuitos que integran las distintas rutas de metabolismo de los medicamentos. Es importante destacar que las rutas relacionadas con el metabolismo de fármacos están compuestas principalmente por metabolitos, cuya actividad fue inferida por medio de la actividad de los genes presentes en cada una de las rutas. Sin embargo, hemos logrado entender con este trabajo que, en el caso particular de las rutas utilizadas para representar la farmacocinética, la mayoría de las interacciones son del tipo metabolito-metabolito, que no sin la presencia de un gen intermediario, no estarían incluidas en este modelo.

Adicionalmente, al realizar el *knock-out in silico* de los genes de las rutas, hemos podido comprobar la transmisión de la señal en toda la ruta, e identificar los elementos efectores de las rutas, que se presuponen los más importantes. Sería interesante añadir al estudio de cada ruta un *knock-out* gen a gen, que permitiría identificar con mayor precisión los elementos fundamentales de cada circuito.

Finalmente, algunas rutas son muy pequeñas, integradas por un número reducido de genes e interacciones gen-molécula, por lo que los resultados de estas rutas podrían verse alterados por este hecho.

12 Futuros pasos

Proponemos la integración de datos de metabolómica, especialmente de datos relacionados con el metabolismo de fármacos, para probar la totalidad de interacciones pertenecientes a las rutas, y así mejorar su capacidad de representación de procesos celulares reales.

Emplear datos de transcriptómica de tejido patológico para las pruebas de las rutas relacionadas con farmacodinámica, con la intención de estudiar elementos adicionales que pudieran identificarse como determinantes de la eficacia.

Validar los resultados obtenidos con datos de variación poblacional española. Conocer la variabilidad genética de la población local es uno de los aspectos más importantes a la hora de implementar estrategias de medicina personalizada. El conocimiento de la variabilidad genética de la población local podría permitir el estudio de genes y variantes presentes en nuestra población.

13 Conclusiones

Actualmente la farmacogenómica se aplica a partir del conocimiento de un número limitado de biomarcadores asociados a la respuesta a fármacos. En un futuro, será necesario la búsqueda e identificación de biomarcadores que permitan la implementación de la farmacogenómica en base a millones de datos derivados del estudio del genoma completo del paciente contribuyendo así a alcanzar el paradigma de los tratamientos personalizados.

Concretamente, en este trabajo hemos conseguido desarrollar un código que permite *parsear* rutas relacionadas con farmacocinética y farmacodinámica disponible en uno de los repositorios más importantes relacionados con la farmacogenética. Como prueba de concepto hemos *parseado* las rutas de Fluoropirimidinas, Metotrexate y Tiopurinas, que posteriormente han sido modeladas empleando HiPathia.

Los métodos de análisis mecanístico de rutas se centran en la actividad de las subestructuras de las vías definidas en nuestro caso como circuitos elementales. Las subestructuras de vías inspiradas en la biología, como los circuitos receptor-efector, y los métodos para puntuar su actividad son relevantes para el rendimiento de este

método. Estos modelos abren nuevas vías para comprender las complejas relaciones entre los genes que, en última instancia, permiten la inferencia del fenotipo.

Por medio de su implementación con 3 casos de uso, se demuestra que los métodos de análisis mecanístico de rutas, por medio de la herramienta HiPathia, pueden ayudar a descubrir mecanicistas biomarcadores predictivos de la eficacia y la toxicidad de fármacos.

14 Referencias

- Amadoz, A., Hidalgo, M. R., Çubuk, C., Carbonell-Caballero, J., & Dopazo, J. (2019). A comparison of mechanistic signaling pathway activity analysis methods. *Briefings in Bioinformatics*, 20(5), 1655-1668. <https://doi.org/10.1093/BIB/BBY040>
- Auffray, C., & Hood, L. (2012). Editorial: Systems biology and personalized medicine - the future is now. *Biotechnology Journal*, 7(8), 938-939. <https://doi.org/10.1002/BIOT.201200242>
- Bateman, A., Martin, M. J., O'Donovan, C., Magrane, M., Apweiler, R., Alpi, E., Antunes, R., Arganiska, J., Bely, B., Bingley, M., Bonilla, C., Britto, R., Bursteinas, B., Chavali, G., Cibrian-Uhalte, E., da Silva, A., de Giorgi, M., Dogan, T., Fazzini, F., ... Zhang, J. (2015). UniProt: a hub for protein information. *Nucleic Acids Research*, 43(Database issue), D204. <https://doi.org/10.1093/NAR/GKU989>
- Bechtel, W., & Richardson, R. C. (2010). Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research. *Discovering Complexity*. <https://doi.org/10.7551/MITPRESS/8328.001.0001>
- Boogerd, F. C., Bruggeman, F. J., Hofmeyr, J. H. S., & Westerhoff, H. v. (2007). Afterthoughts as foundations for systems biology. *Systems Biology*, 321-336. <https://doi.org/10.1016/B978-044452085-2/50016-4>
- Brigandt, I. (2013). Systems biology and the integration of mechanistic explanation and mathematical explanation. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 44(4), 477-492. <https://doi.org/10.1016/J.SHPSC.2013.06.002>
- Carbon, S., Douglass, E., Dunn, N., Good, B., Harris, N. L., Lewis, S. E., Mungall, C. J., Basu, S., Chisholm, R. L., Dodson, R. J., Hartline, E., Fey, P., Thomas, P. D., Albou, L. P., Ebert, D., Kesling, M. J., Mi, H., Muruganujan, A., Huang, X., ... Westerfield, M. (2019). The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Research*, 47(Database issue), D330. <https://doi.org/10.1093/NAR/GKY1055>
- Caudle, K. E., Klein, T. E., Hoffman, J. M., Müller, D. J., Whirl-Carrillo, M., Gong, L., McDonagh, E. M., Sangkuhl, K., Thorn, C. F., Schwab, M., Agúndez, J. A. G., Freimuth, R. R., Huser, V., Lee, M. T. M., Iwuchukwu, O. F., Crews, K. R., Scott, S. A., Wadelius, M., Swen, J. J., ... Johnson, S. G. (2014). Incorporation of Pharmacogenomics into Routine Clinical Practice: the Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline Development Process. *Current Drug Metabolism*, 15(2), 209. <https://doi.org/10.2174/1389200215666140130124910>
- Cronstein, B. N. (1996). Molecular therapeutics: Methotrexate and its mechanism of action. *Arthritis and Rheumatism*, 39(12), 1951-1960. <https://doi.org/10.1002/ART.1780391203>
- Deben, D. S., Creemers, R. H., van Adrichem, A. J., Drent, R., Merry, A. H. H., Leers, M. P. G., van Bodegraven, A. A., & Wong, D. R. (2022). A report on the potential of Rac1/pSTAT3 protein levels in T lymphocytes to assess the pharmacodynamic effect of thiopurine therapy in Inflammatory Bowel Disease patients. *Scientific Reports 2022 12:1*, 12(1), 1-9. <https://doi.org/10.1038/s41598-022-20197-5>

- Frolkis, A., Knox, C., Lim, E., Jewison, T., Law, V., Hau, D. D., Liu, P., Gautam, B., Ly, S., Guo, A. C., Xia, J., Liang, Y., Shrivastava, S., & Wishart, D. S. (2010). SMPDB: The Small Molecule Pathway Database. *Nucleic Acids Research*, *38*(Database issue). <https://doi.org/10.1093/NAR/GKP1002>
- Genotype-Tissue Expression (GTEx) Project*. (2022, septiembre 15). Accession Number Phs000424.vN.PN. <https://www.gtexportal.org/>
- Gisbert, J. P., & Gomollón, F. (2008). Thiopurine-induced myelotoxicity in patients with inflammatory bowel disease: a review. *The American Journal of Gastroenterology*, *103*(7), 1783-1800. <https://doi.org/10.1111/J.1572-0241.2008.01848.X>
- Gnirke, A., Barnes, T. S., Patterson, D., Schild, D., Featherstone, T., & Olson, M. v. (1991). Cloning and in vivo expression of the human GART gene using yeast artificial chromosomes. *The EMBO Journal*, *10*(7), 1629-1634. <https://doi.org/10.1002/J.1460-2075.1991.TB07685.X>
- Gong, L., Whirl-Carrillo, M., & Klein, T. E. (2021). PharmGKB, an Integrated Resource of Pharmacogenomic Knowledge. *Current Protocols*, *1*(8). <https://doi.org/10.1002/CPZ1.226>
- Gonzalez-Covarrubias, V., Lozano, K., Taxis, T., Guzman-Cruz, C. K., Rodriguez-Dorantes, M., Rubio-Carrasco, K., Mendez-Lorenzo, L. H., & Soberon, X. (2020). Pharmacogenomics: Current actionable variants. *Revista de Investigacion Clinica*, *72*(5), 271-279. <https://doi.org/10.24875/RIC.20003239>
- Helikar, T., Konvalina, J., Heidel, J., & Rogers, J. A. (2008). Emergent decision-making in biological signal transduction networks. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(6), 1913. <https://doi.org/10.1073/PNAS.0705088105>
- Hidalgo, M. R., Cubuk, C., Amadoz, A., Salavert, F., Carbonell-Caballero, J., & Dopazo, J. (2017). High throughput estimation of functional cell activities reveals disease mechanisms and predicts relevant clinical outcomes. *Oncotarget*, *8*(3), 5160. <https://doi.org/10.18632/ONCOTARGET.14107>
- HUGO Gene Nomenclature Committee*. (s. f.). Recuperado 15 de septiembre de 2022, a partir de <https://www.genenames.org>
- Ideker, T., Galitski, T., & Hood, L. (2001). A new approach to decoding life: systems biology. *Annual Review of Genomics and Human Genetics*, *2*, 343-372. <https://doi.org/10.1146/ANNUREV.GENOM.2.1.343>
- Jacob, L., Neuvial, P., & Dudoit, S. (2012). More power via graph-structured tests for differential expression of gene networks. *Annals of Applied Statistics*, *6*(2), 561-600. <https://doi.org/10.1214/11-AOAS528>
- Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., & Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research*, *45*(D1), D353-D361. <https://doi.org/10.1093/NAR/GKW1092>
- Kitano, H. (2002). Systems Biology: A Brief Overview. *Science*, *295*(5560), 1662-1664. <https://doi.org/10.1126/SCIENCE.1069492>

- Longley, D. B., Harkin, D. P., & Johnston, P. G. (2003). 5-fluorouracil: mechanisms of action and clinical strategies. *Nature Reviews. Cancer*, 3(5), 330-338.
<https://doi.org/10.1038/NRC1074>
- Marei, H., Carpy, A., Macek, B., & Malliri, A. (2016). Proteomic analysis of Rac1 signaling regulation by guanine nucleotide exchange factors. *Cell Cycle (Georgetown, Tex.)*, 15(15), 1961-1974. <https://doi.org/10.1080/15384101.2016.1183852>
- Medicines Agency, E. (2018). *Work plan for the Pharmacogenomics Working Party (PGWP) for 2018 1. Meetings scheduled for 2018*. www.ema.europa.eu/contact
- Mesarović, M. D. (1968). Systems Theory and Biology—View of a Theoretician. *Systems Theory and Biology*, 59-87. https://doi.org/10.1007/978-3-642-88343-9_3
- Mikkelsen, T. S., Thorn, C. F., Yang, J. J., Ulrich, C. M., French, D., Zaza, G., Dunnenberger, H. M., Marsh, S., McLeod, H. L., Giacomini, K., Becker, M. L., Gaedigk, R., Leeder, J. S., Kager, L., Relling, M. v., Evans, W., Klein, T. E., & Altman, R. B. (2011). PharmGKB summary: Methotrexate pathway. *Pharmacogenetics and Genomics*, 21(10), 679-686.
<https://doi.org/10.1097/FPC.0B013E328343DD93>
- Miteva-Marcheva, N. N., Ivanov, H. Y., Dimitrov, D. K., & Stoyanova, V. K. (2020). Application of pharmacogenetics in oncology. *Biomarker Research*, 8(1), 1-10.
<https://doi.org/10.1186/S40364-020-00213-4/TABLES/1>
- Moriyama, T., Nishii, R., Perez-Andreu, V., Yang, W., Klussmann, F. A., Zhao, X., Lin, T. N., Hoshitsuki, K., Nersting, J., Kihira, K., Hofmann, U., Komada, Y., Kato, M., McCorkle, R., Li, L., Koh, K., Najera, C. R., Kham, S. K. Y., Isobe, T., ... Yang, J. J. (2016). NUDT15 polymorphisms alter thiopurine metabolism and hematopoietic toxicity. *Nature Genetics*, 48(4), 367-373. <https://doi.org/10.1038/NG.3508>
- Niarakis, A., & Helikar, T. (2021). A practical guide to mechanistic systems modeling in biology using a logic-based approach. *Briefings in Bioinformatics*, 22(4).
<https://doi.org/10.1093/BIB/BBAA236>
- Oti, M., & Brunner, H. G. (2007). The modular nature of genetic diseases. *Clinical Genetics*, 71(1), 1-11. <https://doi.org/10.1111/J.1399-0004.2006.00708.X>
- Peña-Chilet, M., Esteban-Medina, M., Falco, M. M., Rian, K., Hidalgo, M. R., Loucera, C., & Dopazo, J. (2019). Using mechanistic models for the clinical interpretation of complex genomic variation. *Scientific Reports 2019 9:1*, 9(1), 1-12.
<https://doi.org/10.1038/s41598-019-55454-7>
- Peña-Chilet, M., Roldán, G., Perez-Florida, J., Ortuño, F. M., Carmona, R., Aquino, V., Lopez-Lopez, D., Loucera, C., Fernandez-Rueda, J. L., Gallego, A., García-García, F., González-Neira, A., Pita, G., Núñez-Torres, R., Santoyo-López, J., Ayuso, C., Minguez, P., Avila-Fernandez, A., Corton, M., ... Dopazo, J. (2021). CSVS, a crowdsourcing database of the Spanish population genetic variability. *Nucleic Acids Research*, 49(D1), D1130-D1137.
<https://doi.org/10.1093/NAR/GKAA794>
- Phan, J., Steadman, D. J., Koli, S., Ding, W. C., Minor, W., Dunlap, R. B., Berger, S. H., & Lebioda, L. (2001). Structure of human thymidylate synthase suggests advantages of chemotherapy with noncompetitive inhibitors. *The Journal of Biological Chemistry*, 276(17), 14170-14177. <https://doi.org/10.1074/JBC.M009493200>

- Prioritization of CPIC Guidelines – CPIC.* (s. f.). Recuperado 9 de octubre de 2022, a partir de <https://cpicpgx.org/prioritization-of-cpic-guidelines/>
- Puniya, B. L., Todd, R. G., Mohammed, A., Brown, D. M., Barberis, M., & Helikar, T. (2018). A mechanistic computational model reveals that plasticity of CD4+ T cell differentiation is a function of cytokine composition and dosage. *Frontiers in Physiology*, 9(AUG). <https://doi.org/10.3389/FPHYS.2018.00878/FULL>
- Razzoli, M., Frontini, A., Gurney, A., Mondini, E., Cubuk, C., Katz, L. S., Cero, C., Bolan, P. J., Dopazo, J., Vidal-Puig, A., Cinti, S., & Bartolomucci, A. (2016). Stress-induced activation of brown adipose tissue prevents obesity in conditions of low adaptive thermogenesis. *Molecular Metabolism*, 5(1), 19. <https://doi.org/10.1016/J.MOLMET.2015.10.005>
- Rian, K., Hidalgo, M. R., Çubuk, C., Falco, M. M., Loucera, C., Esteban-Medina, M., Alamo-Alvarez, I., Peña-Chilet, M., & Dopazo, J. (2021). Genome-scale mechanistic modeling of signaling pathways made easy: A bioconductor/cytoscape/web server framework for the analysis of omic data. *Computational and Structural Biotechnology Journal*, 19, 2968. <https://doi.org/10.1016/J.CSBJ.2021.05.022>
- Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), 139-140. <https://doi.org/10.1093/BIOINFORMATICS/BTP616>
- Saif, M. W., Hachem, H., Purvey, S., Hamal, R., Zhang, L., Siddiqui, N. S., Godara, A., & Diasio, R. B. (2020). Pharmacogenetic Variants in the DPYD and TYMS Genes are Clinically Significant Predictors of Fluoropyrimidine Toxicity: Are We Ready for Use in our Clinical Practice. *Archives of Pharmacology and Therapeutics*, 2(1), 6. <https://doi.org/10.33696/pharmacol.2.012>
- Salavert, F., Hidalgo, M. R., Amadoz, A., Çubuk, C., Medina, I., Crespo, D., Carbonell-Caballero, J., & Dopazo, J. (2016a). Actionable pathways: interactive discovery of therapeutic targets using signaling pathway models. *Nucleic Acids Research*, 44(Web Server issue), W212. <https://doi.org/10.1093/NAR/GKW369>
- Salavert, F., Hidalgo, M. R., Amadoz, A., Çubuk, C., Medina, I., Crespo, D., Carbonell-Caballero, J., & Dopazo, J. (2016b). Actionable pathways: interactive discovery of therapeutic targets using signaling pathway models. *Nucleic Acids Research*, 44(Web Server issue), W212. <https://doi.org/10.1093/NAR/GKW369>
- Schmiegelow, K. (2009). Advances in individual prediction of methotrexate toxicity: a review. *British Journal of Haematology*, 146(5), 489-503. <https://doi.org/10.1111/J.1365-2141.2009.07765.X>
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*, 13(11), 2498. <https://doi.org/10.1101/GR.1239303>
- Vreken, P., van Kuilenburg, A. B. P., Hamajima, N., Meinsma, R., van Lenthe, H., Göhlich-Ratmann, G., Assmann, B. E., Wevers, R. A., & van Gennip, A. H. (1999). cDNA cloning, genomic structure and chromosomal localization of the human BUP-1 gene encoding β -

- ureidopropionase. *Biochimica et Biophysica Acta - Gene Structure and Expression*, 1447(2-3), 251-257. [https://doi.org/10.1016/S0167-4781\(99\)00182-7](https://doi.org/10.1016/S0167-4781(99)00182-7)
- Wang, L., & Weinshilboum, R. (2006). Thiopurine S-methyltransferase pharmacogenetics: insights, challenges and future directions. *Oncogene* 2006 25:11, 25(11), 1629-1638. <https://doi.org/10.1038/sj.onc.1209372>
- Weinshilboum, R. M., & Wang, L. (2017). Pharmacogenomics: Precision Medicine and Drug Response. *Mayo Clinic Proceedings*, 92(11), 1711-1722. <https://doi.org/10.1016/J.MAYOCP.2017.09.001>
- Westerhoff, H. v., & Alberghina, L. (2005). Systems Biology: Did we know it all along? *Systems Biology*, 3-9. <https://doi.org/10.1007/B137744>
- Wishart, D. S., Knox, C., Guo, A. C., Shrivastava, S., Hassanali, M., Stothard, P., Chang, Z., & Woolsey, J. (2006). DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Research*, 34(Database issue). <https://doi.org/10.1093/NAR/GKJ067>
- Yang, S. K., Hong, M., Baek, J., Choi, H., Zhao, W., Jung, Y., Haritunians, T., Ye, B. D., Kim, K. J., Park, S. H., Park, S. K., Yang, D. H., Dubinsky, M., Lee, I., McGovern, D. P. B., Liu, J., & Song, K. (2014). A common missense variant in NUDT15 confers susceptibility to thiopurine-induced leukopenia. *Nature Genetics*, 46(9), 1017-1020. <https://doi.org/10.1038/NG.3060>
- Yu, M. K., Kramer, M., Dutkowski, J., Srivas, R., Licon, K., Kreisberg, J. F., Ng, C. T., Krogan, N., Sharan, R., & Ideker, T. (2016a). Translation of Genotype to Phenotype by a Hierarchy of Cell Subsystems. *Cell Systems*, 2(2), 77. <https://doi.org/10.1016/J.CELS.2016.02.003>
- Yu, M. K., Kramer, M., Dutkowski, J., Srivas, R., Licon, K., Kreisberg, J. F., Ng, C. T., Krogan, N., Sharan, R., & Ideker, T. (2016b). Translation of Genotype to Phenotype by a Hierarchy of Cell Subsystems. *Cell Systems*, 2(2), 77-88. <https://doi.org/10.1016/J.CELS.2016.02.003>
- Zalkin, H., & Dixon, J. E. (1992). De novo purine nucleotide biosynthesis. *Progress in Nucleic Acid Research and Molecular Biology*, 42(C), 259-287. [https://doi.org/10.1016/S0079-6603\(08\)60578-4](https://doi.org/10.1016/S0079-6603(08)60578-4)
- Zaza, G., Cheok, M., Krynetskaia, N., Thorn, C., Stocco, G., Hebert, J. M., McLeod, H., Weinshilboum, R. M., Relling, M. v., Evans, W. E., Klein, T. E., & Altman, R. B. (2010). Thiopurine pathway. *Pharmacogenetics and Genomics*, 20(9), 573-574. <https://doi.org/10.1097/FPC.0B013E328334338F>
- Zaza, G., Yang, W., Kager, L., Cheok, M., Downing, J., Pui, C. H., Cheng, C., Relling, M. v., & Evans, W. E. (2004). Acute lymphoblastic leukemia with TEL-AML1 fusion has lower expression of genes involved in purine metabolism and lower de novo purine synthesis. *Blood*, 104(5), 1435-1441. <https://doi.org/10.1182/BLOOD-2003-12-4306>

15 Anexos

15.1 Lista de paquetes de R usados

R version 4.2.1 (2022-06-23)
Platform: aarch64-apple-darwin20 (64-bit)
Running under: macOS Monterey 12.4

Matrix products: default
LAPACK: /Library/Frameworks/R.framework/Versions/4.2-arm64/Resources/lib/libRlapack.dylib

locale:
[1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8

attached base packages:
[1] stats graphics grDevices utils datasets methods base

other attached packages:
[1] stringr_1.4.1

loaded via a namespace (and not attached):

[1] Rcpp_1.0.9	circlize_0.4.15	png_0.1-7	Biostrings_2.64.1
[5] assertthat_0.2.1	digest_0.6.30	foreach_1.5.2	utf8_1.2.2
[9] R6_2.5.1	GenomeInfoDb_1.32.4	plyr_1.8.7	stats4_4.2.1
[13] RSQLite_2.2.18	httr_1.4.4	ggplot2_3.3.6	pillar_1.8.1
[17] zlibbioc_1.42.0	GlobalOptions_0.1.2	rlang_1.0.6	rstudioapi_0.14
[21] blob_1.2.3	S4Vectors_0.34.0	GetoptLong_1.0.5	RCurl_1.98-1.9
[25] bit_4.0.4	munsell_0.5.0	compiler_4.2.1	xfun_0.34
[29] pkgconfig_2.0.3	BiocGenerics_0.42.0	shape_1.4.6	htmltools_0.5.3
[33] tidyselect_1.2.0	KEGGREST_1.36.3	tibble_3.1.8	gridExtra_2.3
[37] GenomeInfoDb	Data_1.2.8	IRanges_2.30.1	matrixStats_0.62
[41] fansi_1.0.3	viridisLite_0.4.1	crayon_1.5.2	dplyr_1.0.10
[45] bitops_1.0-7	grid_4.2.1	gtable_0.3.1	lifecycle_1.0.3
[49] DBI_1.1.3	magrittr_2.0.3	scales_1.2.1	stringi_1.7.8
[53] cli_3.4.1	cachem_1.0.6	XVector_0.36.0	remotes_2.4.2
[57] viridis_0.6.2	doParallel_1.0.17	generics_0.1.3	vctrs_0.5.0
[61] rjson_0.2.21	RColorBrewer_1.1-3	iterators_1.0.14	tools_4.2.1
[65] bit64_4.0.5	Biobase_2.56.0	glue_1.6.2	parallel_4.2.1
[69] fastmap_1.1.0	clue_0.3-62	AnnotationDbi_1.58.0	colorspace_2.0-3
[73] cluster_2.1.4	ComplexHeatmap_2.12	memoise_2.0.1	knitr_1.40

15.2 Código

El código se encuentra disponible de forma pública en el repositorio Github, y se puede acceder por medio del siguiente enlace: <https://github.com/RaulParraG/Pharmapathia>

El proyecto consta de los siguientes archivos:

- Archivo README: Con una pequeña descripción del proyecto.
- Pharmapathia_one_path.R: Código para el *parseo* de cada una de las rutas de forma independiente. Fue creado inicialmente para comprobar que se realizara correctamente cada paso en una única ruta.

- Pharmapathia_all_paths.R: Código para el *parseo* automático de todas las rutas en una carpeta.

15.3 Visualizador de reportes HiPathia

Para acceder al HTML que contiene los resultados de las rutas modelizadas con HiPathia, se puede acceder al servidor por medio del enlace: <https://clinbioinfospa.es/pharmapathia/pathway-viewer/>

15.4 Enlace archivos

El resto de los archivos se encuentran almacenados en la nube para facilitar su consulta, y se puede acceder siguiendo este enlace: https://drive.google.com/drive/folders/1OC-yunTflyHwfp95IWgazrQQaTkAuUy4?usp=share_link. Estos archivos son necesarios para la implementación de las rutas en HiPathia.

Consta de los siguientes archivos:

- Carpetas Fluropirimidinas_pk, MTX_pkpd, Tiopurinas_pkpd. Su nombre hace referencia al tipo de ruta modelada: farmacocinética (PK) y farmacocinética-farmacodinámica (PKPD). Contiene los archivos necesarios para su implementación: archivos .sif y .att, el archivo original en formato .tsv, diccionario de genes y moléculas en formato .tsv, listado de genes y anotación funcional de circuitos en formato.txt, así como la imagen original de la ruta en formato .pdf.
- Datos LiverGTEx: Contiene todos los datos relacionados con el conteo crudo de los datos de tejido en formato.rds y .tsv, la anotación de las muestras en formato .tsv, listado de genes en formato ENSEMBL, Entrez ID y SYMBOL en .txt
- HiPathia pathway viewer (local): contiene todos los archivos necesarios para abrir el servidor de HiPathia de forma local.
- R – código adicional: código adicional no necesario para la implementación del proyecto Pharmapathia, pero que fue utilizado para obtener parte el manejo e interpretación de los datos de expresión, así como para la generación de las gráficas incluidas en el trabajo.

PA15063576-Fluoropyrimidine_Pathway_Pharmacokinetics

From	To	Reaction Type	Controller	Control Type	Cell Type	PMIDs	Genes	Drugs	Diseases	Summary
fluorodeoxyuridine diphosphate	fluorodeoxyuridine triphosphate	Biochemical Reaction			hepatocyte		12724731			Biochemical reaction from fluorodeoxyuridine diphosphate to fluorodeoxyuridine triphosphate.
fluoridine	fluorodeoxyuridine monophosphate	Biochemical Reaction	TK1	Gene	hepatocyte	12724731				Biochemical reaction from fluoridine to fluorodeoxyuridine monophosphate. This interaction is catalyzed by TK1.
fluorouridine diphosphate	fluorodeoxyuridine diphosphate	Biochemical Reaction	RRM1, RRM2	Gene	hepatocyte	12724731				Biochemical reaction from fluorouridine diphosphate to fluorodeoxyuridine diphosphate. This interaction can be catalyzed by RRM1 or RRM2.
5'-deoxy-5-fluorocytidine	5'-deoxy-5-fluorouridine	Biochemical Reaction	CDA	Gene	hepatocyte	28827188, 18172246, 9849481				Biochemical reaction from 5'-deoxy-5-fluorocytidine to 5'-deoxy-5-fluorouridine. This interaction is catalyzed by CDA.
tegatar	5'-hydroxytegatar	Biochemical Reaction	CYP2A6	Gene	hepatocyte	20463005, 18172246		tegatar		Biochemical reaction from tegatar to 5'-hydroxytegatar. This interaction is catalyzed by CYP2A6.
5'-deoxy-5-fluorouridine	fluorouracil	Biochemical Reaction	TYMP, UPP1, UPP2	Gene	hepatocyte	18172246, 11566088, 9848491		fluorouracil		Biochemical reaction from 5'-deoxy-5-fluorouridine to fluorouracil. This interaction can be catalyzed by UPP2, UPP1 or TYMP.
5'-hydroxytegatar	fluorouracil	Biochemical Reaction		Gene	hepatocyte	20463005, 18172246		fluorouracil		Biochemical reaction from 5'-hydroxytegatar to fluorouracil.
capcitabine	5'-deoxy-5-fluorocytidine	Biochemical Reaction	CES1, CES2	Gene	hepatocyte	28827188, 28138840, 18172246		capcitabine		Biochemical reaction from capcitabine to 5'-deoxy-5-fluorocytidine. This interaction can be catalyzed by CES1 or CES2.
fluorodeoxyuridine diphosphate	fluorodeoxyuridine monophosphate	Biochemical Reaction		hepatocyte	hepatocyte	12724731				Biochemical reaction from fluorodeoxyuridine diphosphate to fluorodeoxyuridine monophosphate.
fluorouracil	dihydrofluorouracil	Biochemical Reaction	DPYD	Gene	hepatocyte	14555807		fluorouracil		Biochemical reaction from fluorouracil to dihydrofluorouracil. This interaction is catalyzed by DPYD.
fluorouridine	fluoridine	Biochemical Reaction	TYMP	Gene	hepatocyte	12724731, 10741735		fluorouracil		Biochemical reaction from fluorouridine to fluoridine. This interaction is catalyzed by TYMP.
fluorouridine	fluorouridine monophosphate	Biochemical Reaction	UCK1, LUCK2	Gene	hepatocyte	12724731		fluorouracil		Biochemical reaction from fluorouridine to fluorouridine monophosphate. This interaction can be catalyzed by UCK1 or LUCK2.
fluorouracil	fluorouridine monophosphate	Biochemical Reaction	PPAT, UMPS	Gene	hepatocyte	12724731		fluorouracil		Biochemical reaction from fluorouracil to fluorouridine monophosphate. This interaction can be catalyzed by PPAT or UMPS.
fluorouridine diphosphate	fluorouridine triphosphate	Biochemical Reaction		hepatocyte	hepatocyte	12724731				Biochemical reaction from fluorouridine diphosphate to fluorouridine triphosphate.
dihydrofluorouracil	fluoro-beta-ureidopropionate	Biochemical Reaction	DPYS	Gene	hepatocyte	18075467				Biochemical reaction from dihydrofluorouracil to fluoro-beta-ureidopropionate. This interaction is catalyzed by DPYS.
fluorouridine monophosphate	fluorouridine diphosphate	Biochemical Reaction		hepatocyte	hepatocyte	12724731				Biochemical reaction from fluorouridine monophosphate to fluorouridine diphosphate.
fluoro-beta-ureidopropionate	fluoro-beta-alanine	Biochemical Reaction	UPB1	Gene	hepatocyte	14555807				Biochemical reaction from fluoro-beta-ureidopropionate to fluoro-beta-alanine. This interaction is catalyzed by UPB1.
fluorouracil	fluorouridine	Biochemical Reaction	UPP1, UPP2	Gene	hepatocyte	11956089		fluorouracil		Biochemical reaction from fluorouracil to fluorouridine. This interaction can be catalyzed by UPP2 or UPP1.
fluorodeoxyuridine monophosphate	TYMS	Inhibition		hepatocyte	hepatocyte	18339911, 12724731		TYMS		fluorodeoxyuridine monophosphate inhibits TYMS.
fluorodeoxyuridine triphosphate	Fluoropyrimidine Pathway, Pharmacodynamics	Leads To		hepatocyte, cell	hepatocyte, cell	12724731				fluorodeoxyuridine triphosphate leads to Fluoropyrimidine Pathway, Pharmacodynamics.
TYMS	Fluoropyrimidine Pathway, Pharmacodynamics	Leads To		hepatocyte, cell	hepatocyte, cell	12724731		TYMS		TYMS leads to Fluoropyrimidine Pathway, Pharmacodynamics.
fluorouridine triphosphate	Fluoropyrimidine Pathway, Pharmacodynamics	Leads To		hepatocyte, cell	hepatocyte, cell	12724731				fluorouridine triphosphate leads to Fluoropyrimidine Pathway, Pharmacodynamics.
fluorouracil	fluorouracil	Transport	SLC22A7	Gene	null, hepatocyte	15901346		fluorouracil		Transport from fluorouracil to fluorouracil. This interaction is catalyzed by SLC22A7.
fluorouracil	fluorouracil	Transport	ABCC4, ABCG2	Gene	hepatocyte, null	18820913, 19077464, 18837281		fluorouracil		Transport from fluorouracil to fluorouracil. This interaction can be catalyzed by ABCC4 or ABCG2.
fluorouracil	fluorouracil	Transport	ABCC3, ABCC4, ABCG5	Gene	hepatocyte, null	19077464		fluorouracil		Transport from fluorouracil to fluorouracil. This interaction can be catalyzed by ABCC3, ABCG5 or ABCG4.
fluorouracil	fluorouracil	Transport	SLC29A1	Gene	null, hepatocyte	18992248, 17955509		fluorouracil		Transport from fluorouracil to fluorouracil. This interaction is catalyzed by SLC29A1.

Tabla Anexos 1. Ruta Fluoropirimidinas PharmGKB

From	map_mol	Reaction Type	To	map_mol2	Gene1	Entrez1	coded_gene1
fluorodeoxyuridine diphosphate	N-PA150653776-01	activation	fluorodeoxyuridine triphosphate	N-PA150653776-15	NA	NA	NA
floxuridine	N-PA150653776-02	activation	fluorodeoxyuridine monophosphate	N-PA150653776-14	TK1	7083	N-PA150653776-001
fluorouridine diphosphate	N-PA150653776-03	activation	fluorodeoxyuridine diphosphate	N-PA150653776-01	RRM1	6240	N-PA150653776-002
5'-deoxy-5-fluorocytidine	N-PA150653776-04	activation	5'-deoxy-5-fluorouridine	N-PA150653776-06	CDA	978	N-PA150653776-003
tegafur	N-PA150653776-05	activation	5'-hydroxytegafur	N-PA150653776-07	CYP2A6	1548	N-PA150653776-004
5'-deoxy-5-fluorouridine	N-PA150653776-06	activation	fluorouracil	N-PA150653776-09	TYMP	1890	N-PA150653776-005
5'-hydroxytegafur	N-PA150653776-07	activation	fluorouracil	N-PA150653776-09	NA	NA	NA
capecitabine	N-PA150653776-08	activation	5'-deoxy-5-fluorocytidine	N-PA150653776-04	CES1	1066	N-PA150653776-006
fluorodeoxyuridine diphosphate	N-PA150653776-01	activation	fluorodeoxyuridine monophosphate	N-PA150653776-14	NA	NA	NA
fluorouracil	N-PA150653776-09	activation	dihydrofluorouracil	N-PA150653776-11	DPYD	1806	N-PA150653776-007
fluorouracil	N-PA150653776-09	activation	floxuridine	N-PA150653776-02	TYMP	1890	N-PA150653776-005
fluorouridine	N-PA150653776-10	activation	fluorouridine monophosphate	N-PA150653776-12	UCK1	83549	N-PA150653776-008
fluorouracil	N-PA150653776-09	activation	fluorouridine monophosphate	N-PA150653776-12	PPAT	5471	N-PA150653776-009
fluorouridine diphosphate	N-PA150653776-03	activation	fluorouridine triphosphate	N-PA150653776-17	NA	NA	NA
dihydrofluorouracil	N-PA150653776-11	activation	fluoro-beta-ureidopropionate	N-PA150653776-13	DPYS	1807	N-PA150653776-010
fluorouridine monophosphate	N-PA150653776-12	activation	fluorouridine diphosphate	N-PA150653776-03	NA	NA	NA

Tabla Anexos 2. Tabla referencia Fluoropirimidinas

moleculas	coded_mol
fluorodeoxyuridine diphosphate	N-PA150653776-01
floxuridine	N-PA150653776-02
fluorouridine diphosphate	N-PA150653776-03
5'-deoxy-5-fluorocytidine	N-PA150653776-04
tegafur	N-PA150653776-05
5'-deoxy-5-fluorouridine	N-PA150653776-06
5'-hydroxytegafur	N-PA150653776-07
capecitabine	N-PA150653776-08
fluorouracil	N-PA150653776-09
fluorouridine	N-PA150653776-10
dihydrofluorouracil	N-PA150653776-11
fluorouridine monophosphate	N-PA150653776-12
fluoro-beta-ureidopropionate	N-PA150653776-13
fluorodeoxyuridine monophosphate	N-PA150653776-14
fluorodeoxyuridine triphosphate	N-PA150653776-15

Tabla Anexos 3. Tabla moléculas Fluoropirimidinas

SYMBOL	EntrezID	coded_gen
TK1	7083	N-PA150653776-001
RRM1	6240	N-PA150653776-002
CDA	978	N-PA150653776-003
CYP2A6	1548	N-PA150653776-004
TYMP	1890	N-PA150653776-005
CES1	1066	N-PA150653776-006
DPYD	1806	N-PA150653776-007
UCK1	83549	N-PA150653776-008
PPAT	5471	N-PA150653776-009
DPYS	1807	N-PA150653776-010
UPB1	51733	N-PA150653776-011
UPP1	7378	N-PA150653776-012
SLC22A7	10864	N-PA150653776-013
ABCC4	10257	N-PA150653776-014
ABCC3	8714	N-PA150653776-015
SLC29A1	2030	N-PA150653776-016

Tabla Anexos 4. Tabla genes Fluoropirimidinas

map_mol	Reaction Type	map_mol2	coded_gene
N-PA150653776-01	activation	N-PA150653776-15	
N-PA150653776-02	activation	N-PA150653776-14	N-PA150653776-001
N-PA150653776-03	activation	N-PA150653776-01	N-PA150653776-002 N-PA150653776-017
N-PA150653776-04	activation	N-PA150653776-06	N-PA150653776-003
N-PA150653776-05	activation	N-PA150653776-07	N-PA150653776-004
N-PA150653776-06	activation	N-PA150653776-09	N-PA150653776-005 N-PA150653776-012 N-PA15...
N-PA150653776-07	activation	N-PA150653776-09	
N-PA150653776-08	activation	N-PA150653776-04	N-PA150653776-006 N-PA150653776-018
N-PA150653776-01	activation	N-PA150653776-14	
N-PA150653776-09	activation	N-PA150653776-11	N-PA150653776-007
N-PA150653776-09	activation	N-PA150653776-02	N-PA150653776-005
N-PA150653776-10	activation	N-PA150653776-12	N-PA150653776-008 N-PA150653776-019
N-PA150653776-09	activation	N-PA150653776-12	N-PA150653776-009 N-PA150653776-020
N-PA150653776-03	activation	N-PA150653776-17	
N-PA150653776-11	activation	N-PA150653776-13	N-PA150653776-010
N-PA150653776-12	activation	N-PA150653776-03	

Tabla Anexos 5. Tabla codificada Fluoropirimidinas

map_mol	Reaction Type	map_mol2
N-PA150653776-01	activation	N-PA150653776-15
N-PA150653776-07	activation	N-PA150653776-09
N-PA150653776-01	activation	N-PA150653776-14
N-PA150653776-03	activation	N-PA150653776-17
N-PA150653776-12	activation	N-PA150653776-03
N-PA150653776-14	inhibition	N-PA150653776-16
N-PA150653776-15	activation	N-PA150653776-19
N-PA150653776-16	activation	N-PA150653776-19
N-PA150653776-17	activation	N-PA150653776-19
N-PA150653776-02	activation	N-PA150653776-001
N-PA150653776-001	activation	N-PA150653776-14
N-PA150653776-03	activation	N-PA150653776-002 N-PA150653776-017
N-PA150653776-002 N-PA150653776-017	activation	N-PA150653776-01
N-PA150653776-04	activation	N-PA150653776-003
N-PA150653776-003	activation	N-PA150653776-06
N-PA150653776-05	activation	N-PA150653776-004
N-PA150653776-004	activation	N-PA150653776-07

Tabla Anexos 6. Tabla pasos intermedios (expanded) Fluoropirimidinas

"N-PA150653776-01"	"activation"	"N-PA150653776-15"
"N-PA150653776-07"	"activation"	"N-PA150653776-09"
"N-PA150653776-01"	"activation"	"N-PA150653776-14"
"N-PA150653776-03"	"activation"	"N-PA150653776-17"
"N-PA150653776-12"	"activation"	"N-PA150653776-03"
"N-PA150653776-14"	"inhibition"	"N-PA150653776-16"
"N-PA150653776-02"	"activation"	"N-PA150653776-001"
"N-PA150653776-001"	"activation"	"N-PA150653776-14"
"N-PA150653776-03"	"activation"	"N-PA150653776-002 017"
"N-PA150653776-002 017"	"activation"	"N-PA150653776-01"
"N-PA150653776-04"	"activation"	"N-PA150653776-003"
"N-PA150653776-003"	"activation"	"N-PA150653776-06"
"N-PA150653776-05"	"activation"	"N-PA150653776-004"
"N-PA150653776-004"	"activation"	"N-PA150653776-07"
"N-PA150653776-06"	"activation"	"N-PA150653776-005 012 021"
"N-PA150653776-005 012 021"	"activation"	"N-PA150653776-09"
"N-PA150653776-08"	"activation"	"N-PA150653776-006 018"
"N-PA150653776-006 018"	"activation"	"N-PA150653776-04"
"N-PA150653776-09"	"activation"	"N-PA150653776-007"
"N-PA150653776-007"	"activation"	"N-PA150653776-11"
"N-PA150653776-09"	"activation"	"N-PA150653776-005"
"N-PA150653776-005"	"activation"	"N-PA150653776-02"
"N-PA150653776-10"	"activation"	"N-PA150653776-008 019"
"N-PA150653776-008 019"	"activation"	"N-PA150653776-12"
"N-PA150653776-09"	"activation"	"N-PA150653776-009 020"
"N-PA150653776-009 020"	"activation"	"N-PA150653776-12"
"N-PA150653776-11"	"activation"	"N-PA150653776-010"
"N-PA150653776-010"	"activation"	"N-PA150653776-13"
"N-PA150653776-13"	"activation"	"N-PA150653776-011"
"N-PA150653776-011"	"activation"	"N-PA150653776-18"
"N-PA150653776-09"	"activation"	"N-PA150653776-012 021"
"N-PA150653776-012 021"	"activation"	"N-PA150653776-10"
"N-PA150653776-09"	"activation"	"N-PA150653776-013"
"N-PA150653776-013"	"activation"	"N-PA150653776-09"
"N-PA150653776-09"	"activation"	"N-PA150653776-014 022"
"N-PA150653776-014 022"	"activation"	"N-PA150653776-09"
"N-PA150653776-09"	"activation"	"N-PA150653776-015 014 023"
"N-PA150653776-015 014 023"	"activation"	"N-PA150653776-09"
"N-PA150653776-09"	"activation"	"N-PA150653776-016"
"N-PA150653776-016"	"activation"	"N-PA150653776-09"

Tabla Anexos 7. Archivo SIF Fluoropirimidinas

ID	label	X	Y	color	shape	type	label.cex	label.color	width	height	genesList						
N-PA150653776-01			fluorodeoxyuridine diphosphate					white	circle	compound	0.5	black	46	17			
N-PA150653776-07			5-hydroxytegafur				white	circle	compound	0.5	black	46	17				
N-PA150653776-01			fluorodeoxyuridine diphosphate					white	circle	compound	0.5	black	46	17			
N-PA150653776-03			fluorouridine diphosphate					white	circle	compound	0.5	black	46	17			
N-PA150653776-12			fluorouridine monophosphate					white	circle	compound	0.5	black	46	17			
N-PA150653776-14			fluorodeoxyuridine monophosphate					white	circle	compound	0.5	black	46	17			
N-PA150653776-15			fluorodeoxyuridine triphosphate					white	circle	compound	0.5	black	46	17			
N-PA150653776-16			TYMS		white	circle	gene	0.5	black	46	17	7298					
N-PA150653776-17			fluorouridine triphosphate					white	circle	compound	0.5	black	46	17			
N-PA150653776-02			floxuridine		white	circle	gene	0.5	black	46	17						
N-PA150653776-001			TK1		white	rectangle		white	rectangle	compound	0.5	black	46	17	7083		
N-PA150653776-03			fluorouridine diphosphate					white	circle	compound	0.5	black	46	17			
N-PA150653776-002	017		RRM1, RRM2				white	rectangle	gene	0.5	black	46	17	6240,,6241			
N-PA150653776-04			5-deoxy-5-fluorocytidine					white	circle	compound	0.5	black	46	17			
N-PA150653776-003			CDA		white	rectangle		white	rectangle	compound	0.5	black	46	17	978		
N-PA150653776-05			tegafur		white	circle	compound	white	circle	compound	0.5	black	46	17			
N-PA150653776-004			CYP2A6		white	rectangle		white	rectangle	compound	0.5	black	46	17	1548		
N-PA150653776-06			5-deoxy-5-fluorouridine				white	circle	compound	0.5	black	46	17				
N-PA150653776-005	012	021	TYMP, UPP1, UPP2					white	rectangle	gene	0.5	black	46	17	1890,,7378,,151531		
N-PA150653776-08			capecitabine				white	circle	compound	0.5	black	46	17				
N-PA150653776-006	018		CES1, CES2				white	rectangle	gene	0.5	black	46	17	1066,,8824			
N-PA150653776-09			fluorouracil				white	circle	compound	0.5	black	46	17				
N-PA150653776-007			DPYD		white	rectangle		white	rectangle	compound	0.5	black	46	17	1806		
N-PA150653776-09			fluorouracil				white	circle	compound	0.5	black	46	17				
N-PA150653776-005			TYMP		white	rectangle		white	rectangle	compound	0.5	black	46	17	1890		
N-PA150653776-10			fluorouridine				white	circle	compound	0.5	black	46	17				
N-PA150653776-008	019		UCK1, UCK2				white	rectangle	gene	0.5	black	46	17	83549,,7371			
N-PA150653776-09			fluorouracil				white	circle	compound	0.5	black	46	17				
N-PA150653776-009	020		PPAT, UMPS				white	rectangle	gene	0.5	black	46	17	5471,,7372			
N-PA150653776-11			dihydrofluorouracil				white	circle	compound	0.5	black	46	17				
N-PA150653776-010			DPYS		white	rectangle		white	rectangle	compound	0.5	black	46	17	1807		
N-PA150653776-13			fluoro-beta-ureidopropionate				white	circle	compound	0.5	black	46	17				
N-PA150653776-011			UPB1		white	rectangle		white	rectangle	compound	0.5	black	46	17	51733		
N-PA150653776-09			fluorouracil				white	circle	compound	0.5	black	46	17				
N-PA150653776-012	021		UPP1, UPP2				white	rectangle	gene	0.5	black	46	17	7378,,151531			
N-PA150653776-09			fluorouracil				white	circle	compound	0.5	black	46	17				
N-PA150653776-013			SLC22A7		white	rectangle		white	rectangle	compound	0.5	black	46	17	10864		
N-PA150653776-09			fluorouracil				white	circle	compound	0.5	black	46	17				
N-PA150653776-014	022		ABCC4, ABCC2				white	rectangle	gene	0.5	black	46	17	10257,,9429			
N-PA150653776-09			fluorouracil				white	circle	compound	0.5	black	46	17				
N-PA150653776-015	014	023	ABCC3, ABCC4, ABCC5				white	rectangle	gene	0.5	black	46	17	8714,,10257,,10057			
N-PA150653776-09			fluorouracil				white	circle	compound	0.5	black	46	17				
N-PA150653776-016			SLC29A1		white	rectangle		white	rectangle	compound	0.5	black	46	17	2030		
N-PA150653776-02			floxuridine				white	circle	compound	0.5	black	46	17				
N-PA150653776-001			TK1		white	rectangle		white	rectangle	compound	0.5	black	46	17	7083		
N-PA150653776-18			fluoro-beta-alanine				white	circle	compound	0.5	black	46	17				

Tabla Anexos 8. Archivo ATT Fluoropirimidinas

"7083"
"6240"
"978"
"1548"
"1890"
"1066"
"1806"
"83549"
"5471"
"1807"
"51733"
"7378"
"10864"
"10257"
"8714"
"2030"
"6241"
"8824"
"7371"
"7372"
"151531"
"9429"
"10057"

Tabla Anexos 9. Genes list Fluoropirimidinas

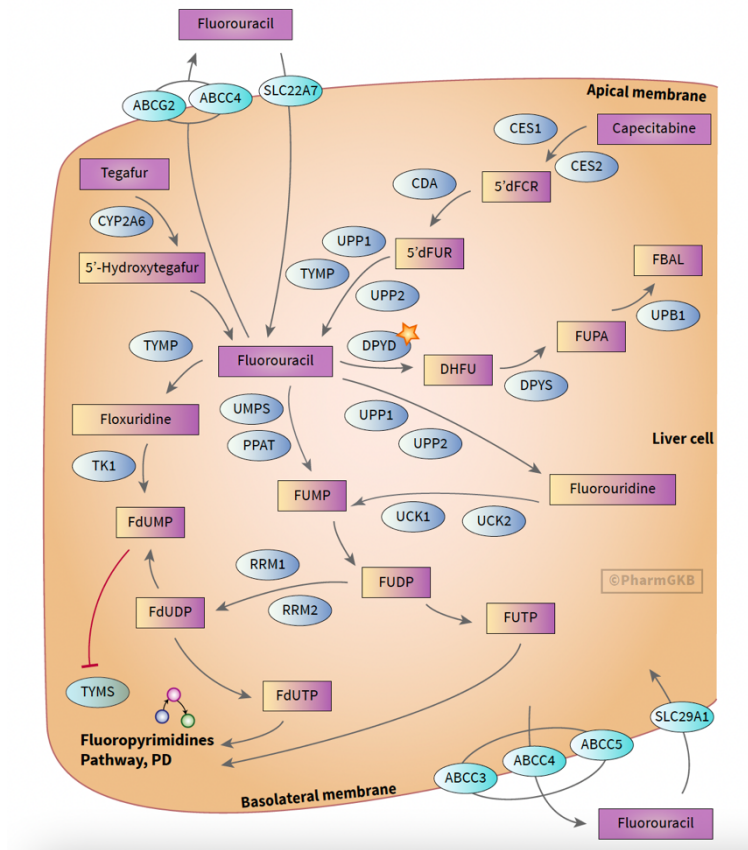


Figura Anexos 1. Ruta Fluoropirimidinas PharmGKB

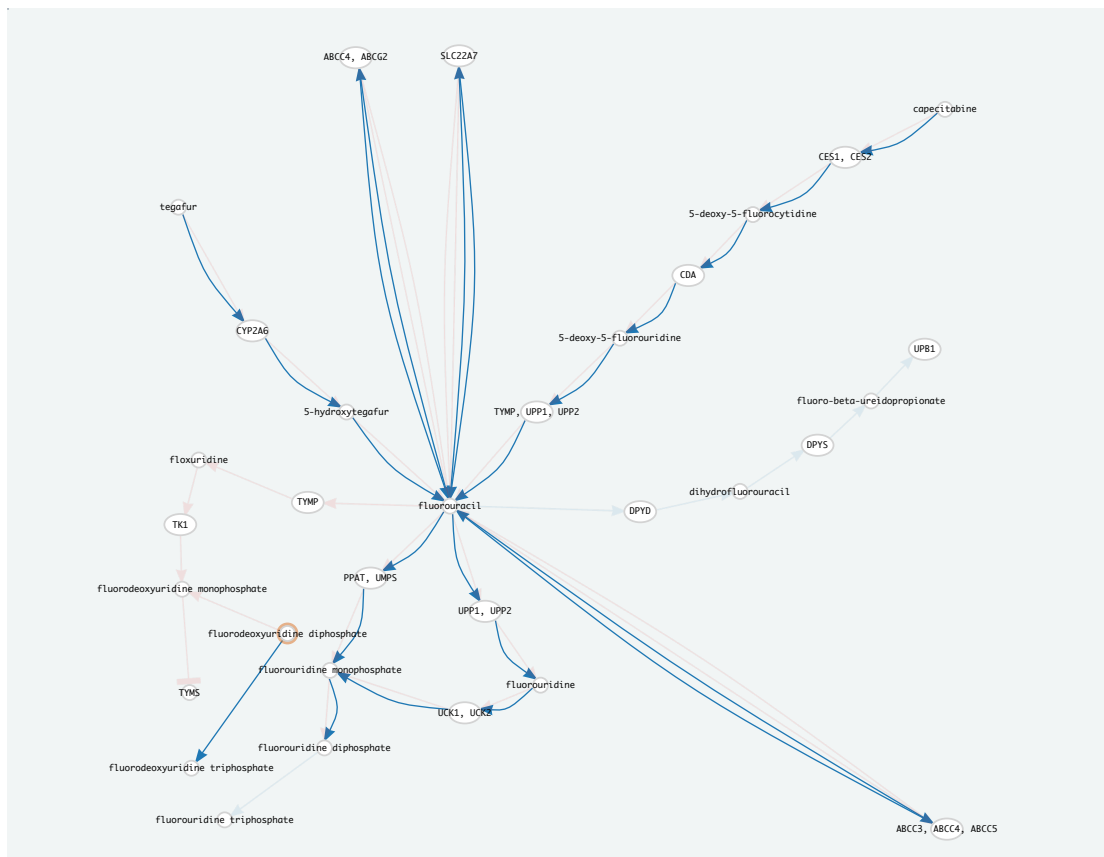


Figura Anexos 2. Fluoropirimidinas, circuito fluorodeoxiuridina trifosfato

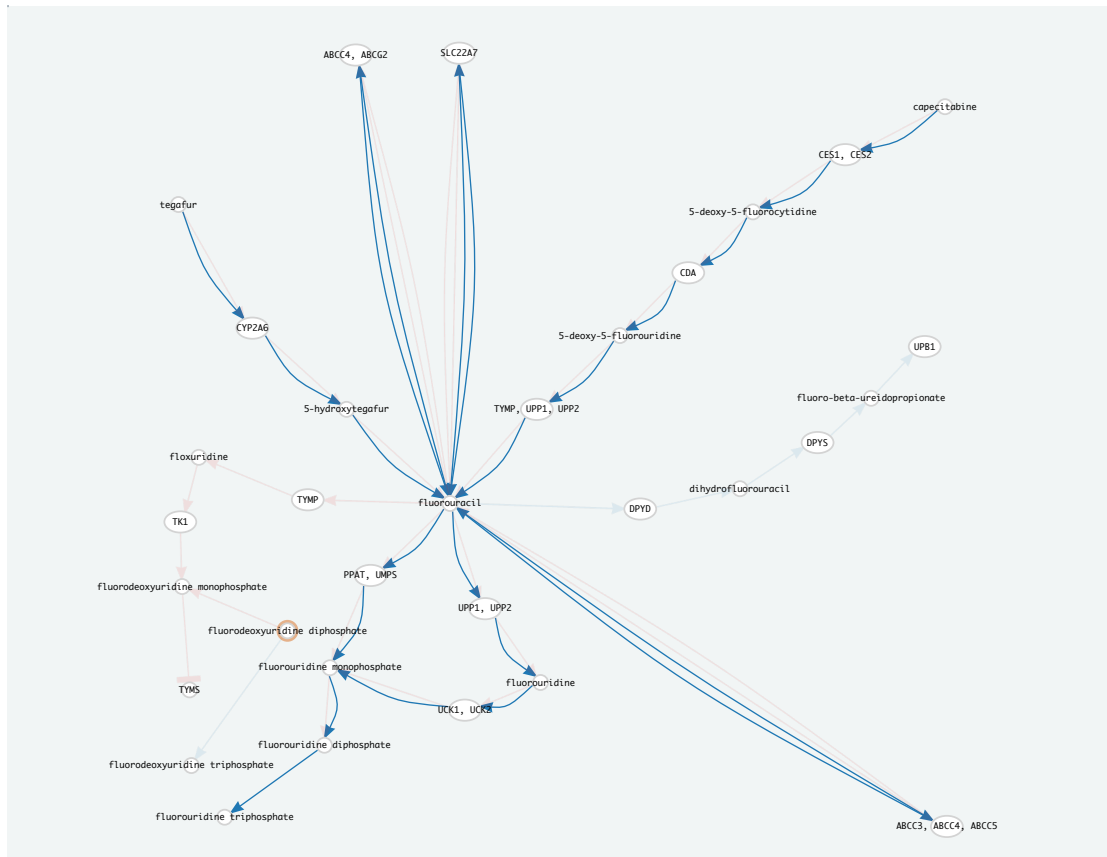


Figura Anexos 3. Fluoropirimidinas, circuito fluorouridina trifosfato

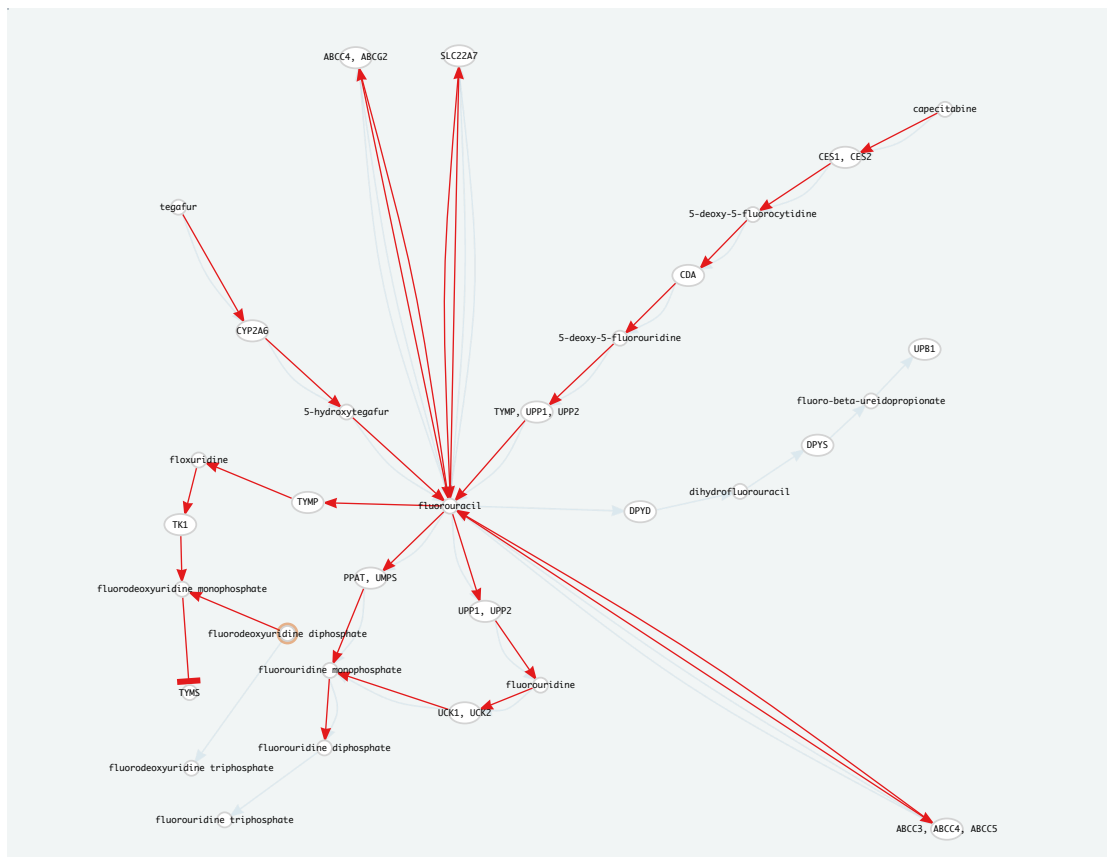


Figura Anexos 4. Fluoropirimidinas, circuito TYMS

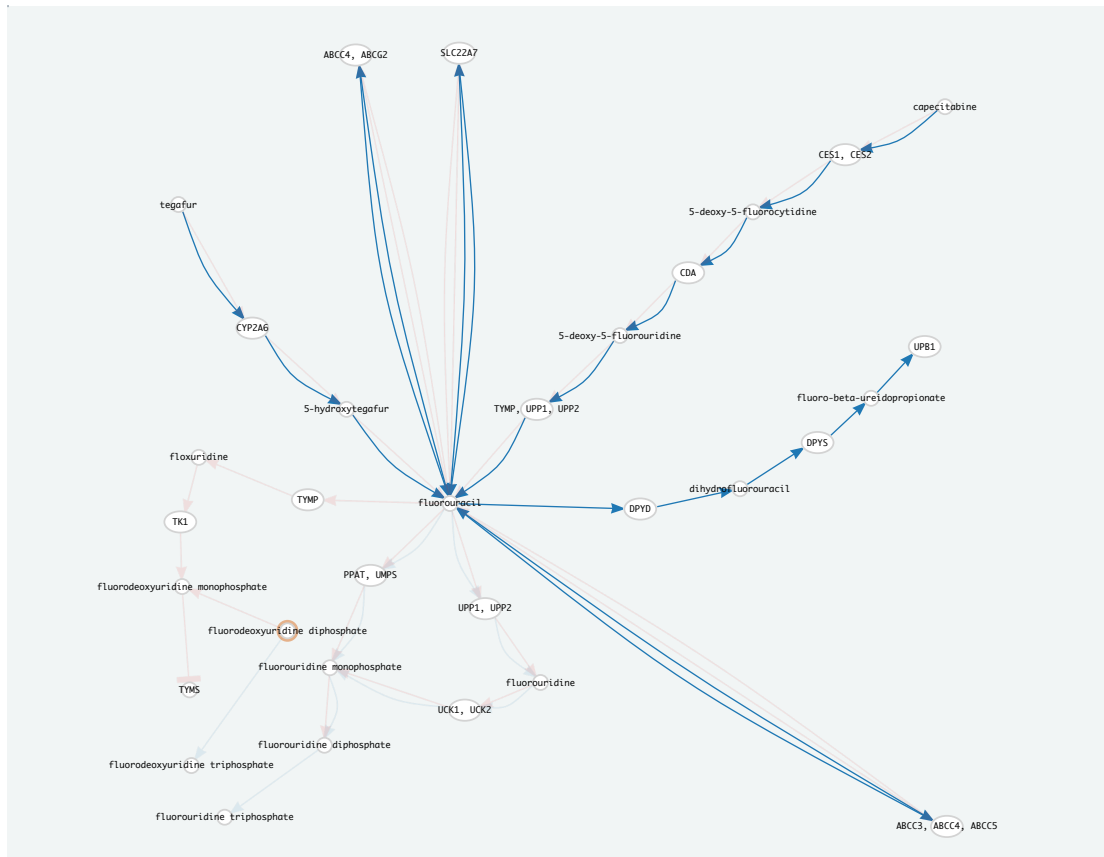


Figura Anexos 5. Fluoropirimidinas, circuito UPB1

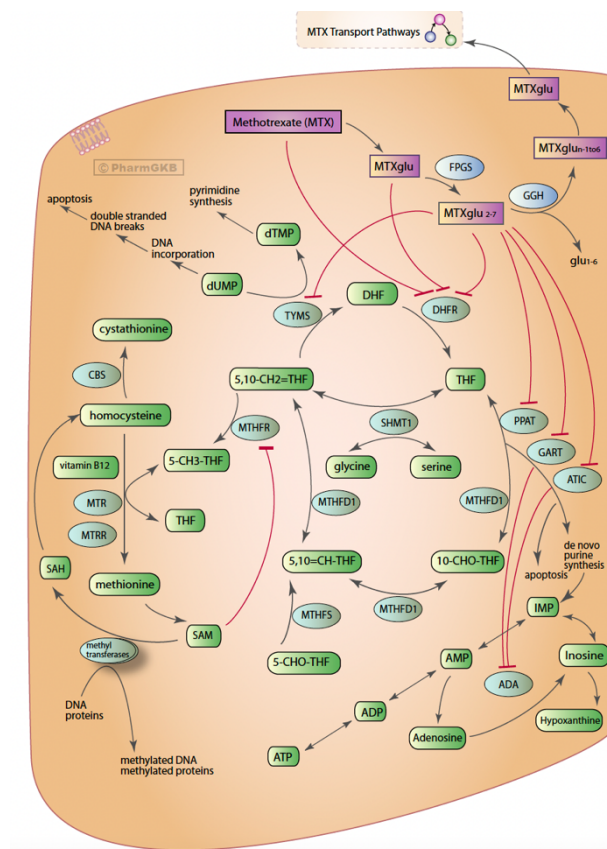


Figura Anexos 6. Figura Anexos 1. Ruta Metotrexate PharmGKB

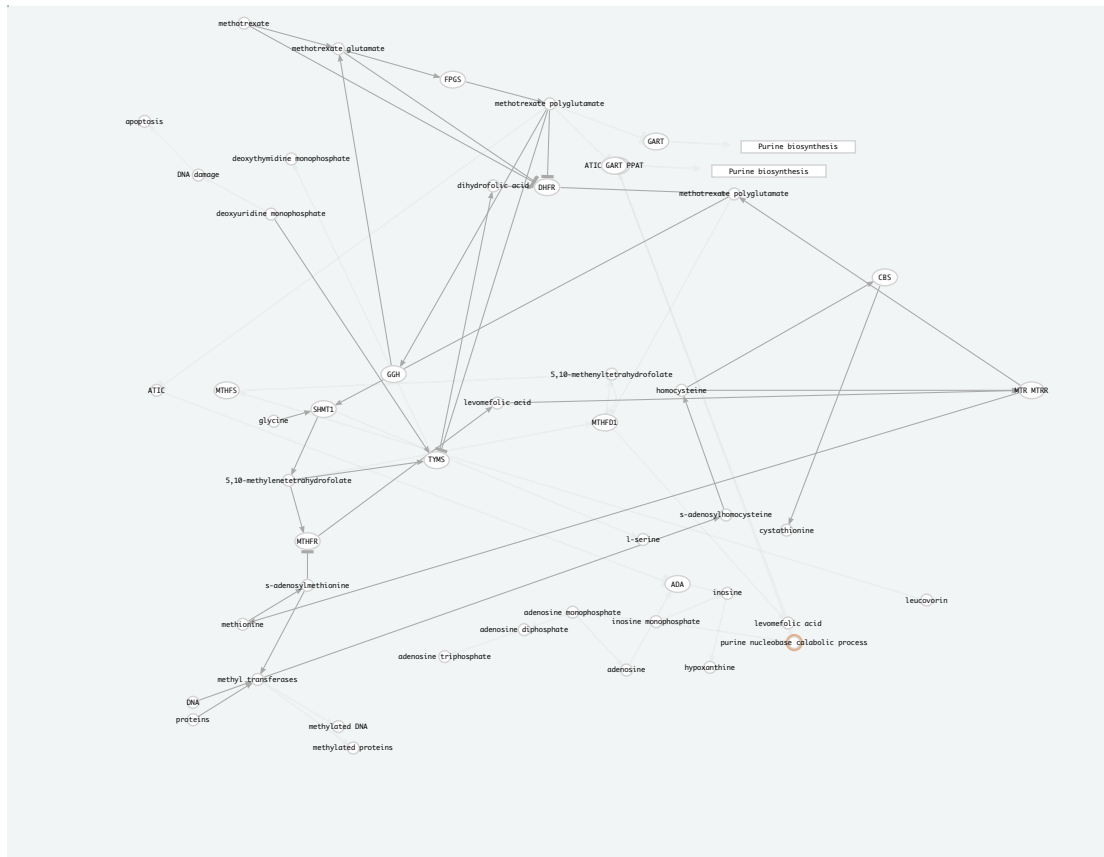


Figura Anexos 9. Metotrexate, circuito cistationina

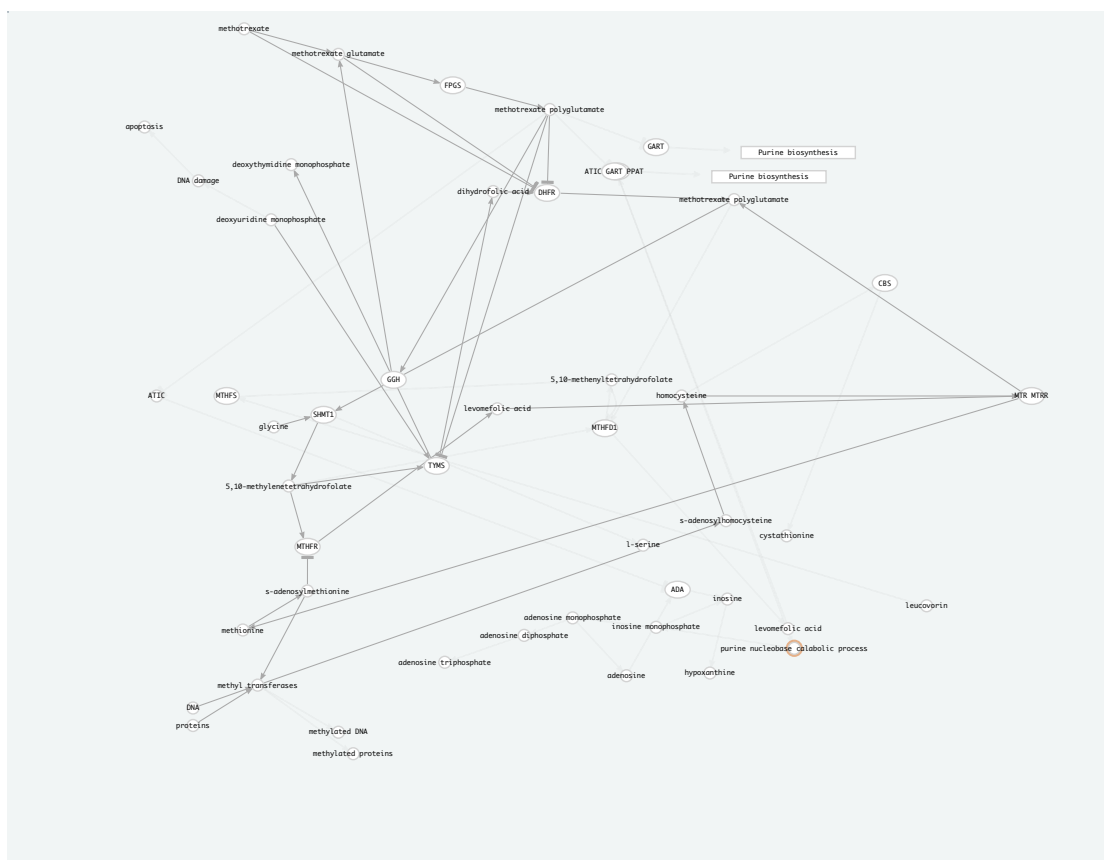


Figura Anexos 10. Metotrexate, circuito deoxitimidina monofosfato

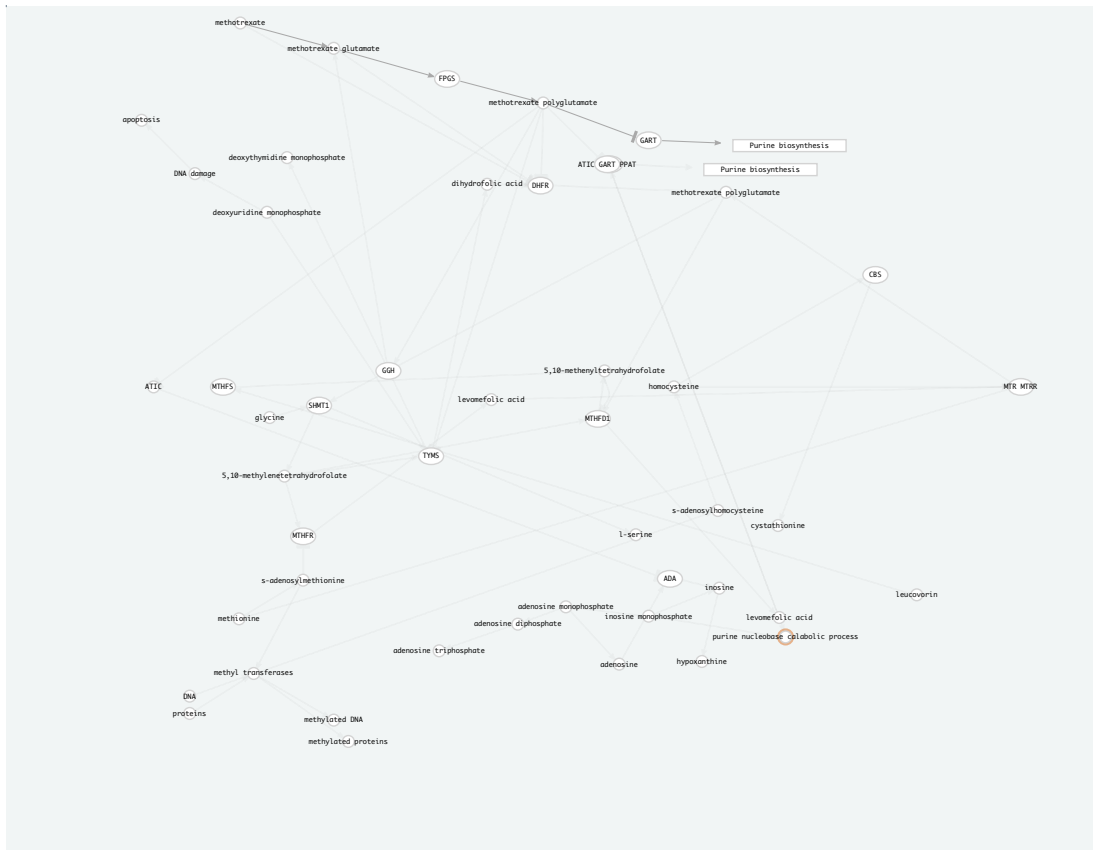


Figura Anexos 11. Metotrexate, circuito hipoxantina

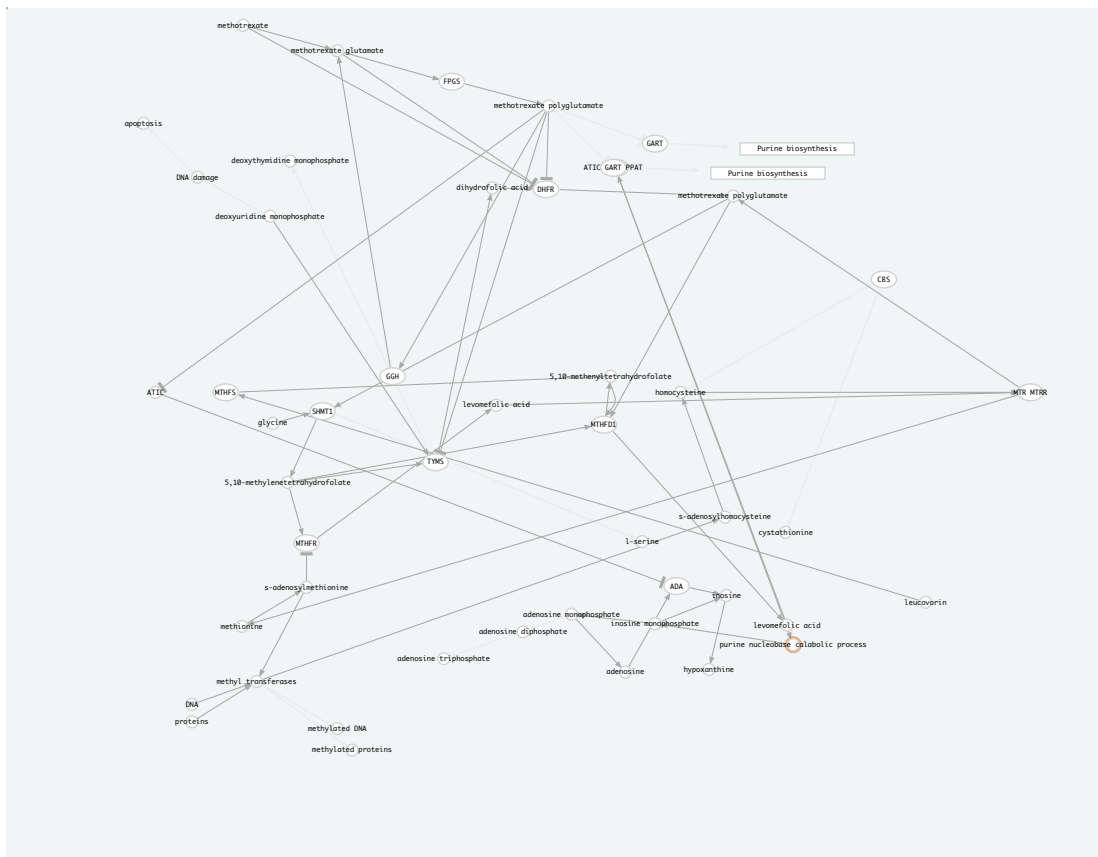


Figura Anexos 12. Metotrexate, circuito l-serina

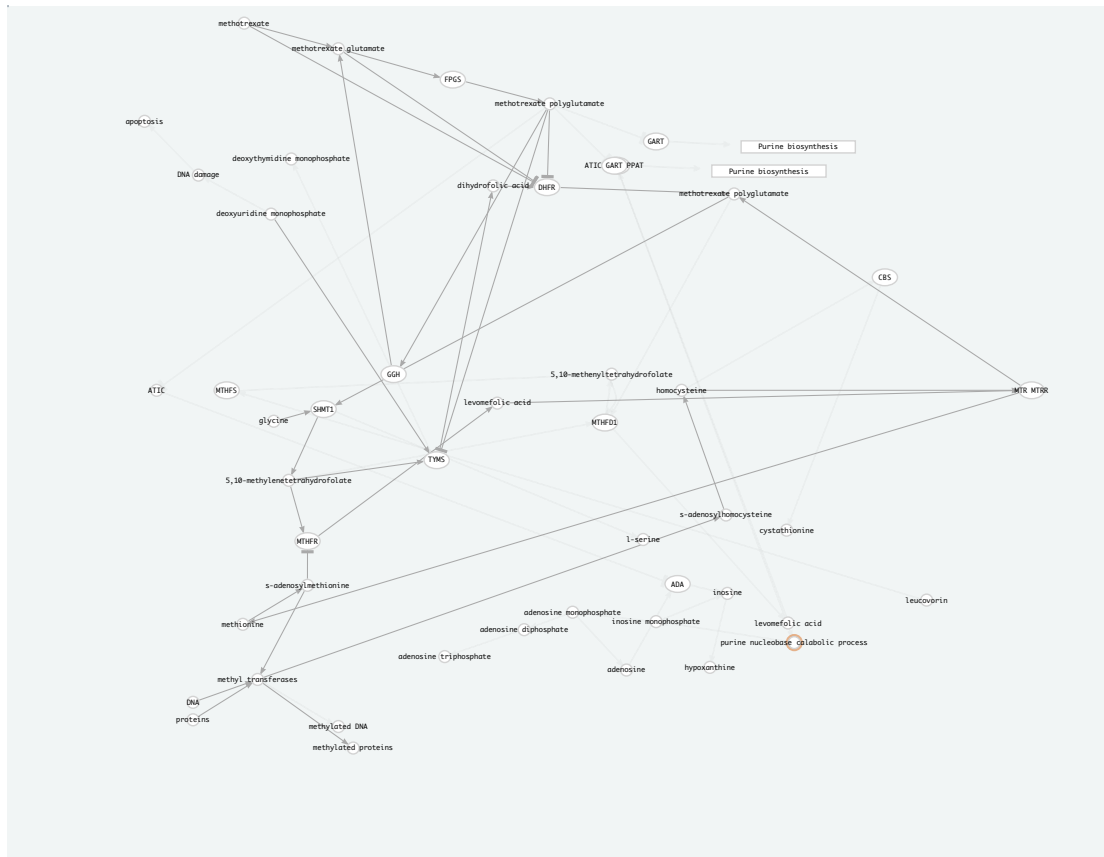


Figura Anexos 15. Metotrexate, circuito DNA metilado

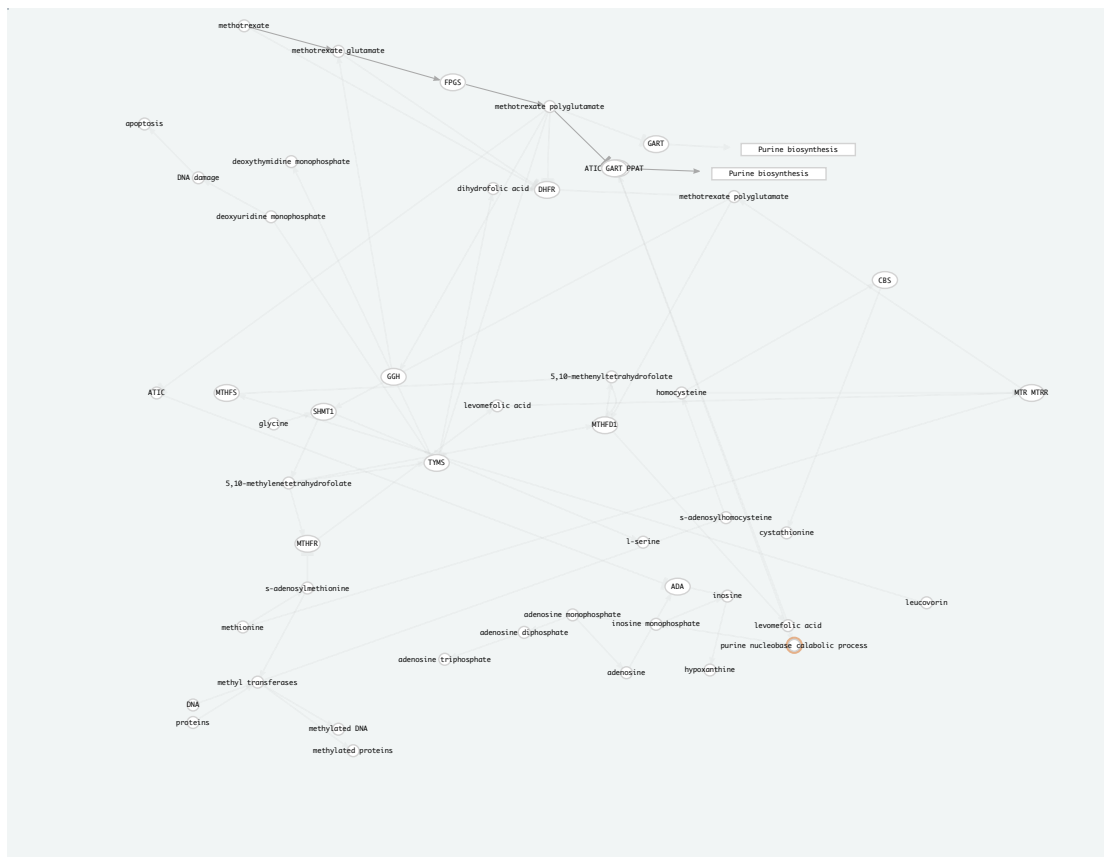


Figura Anexos 16. Metotrexate, circuito proteínas metiladas

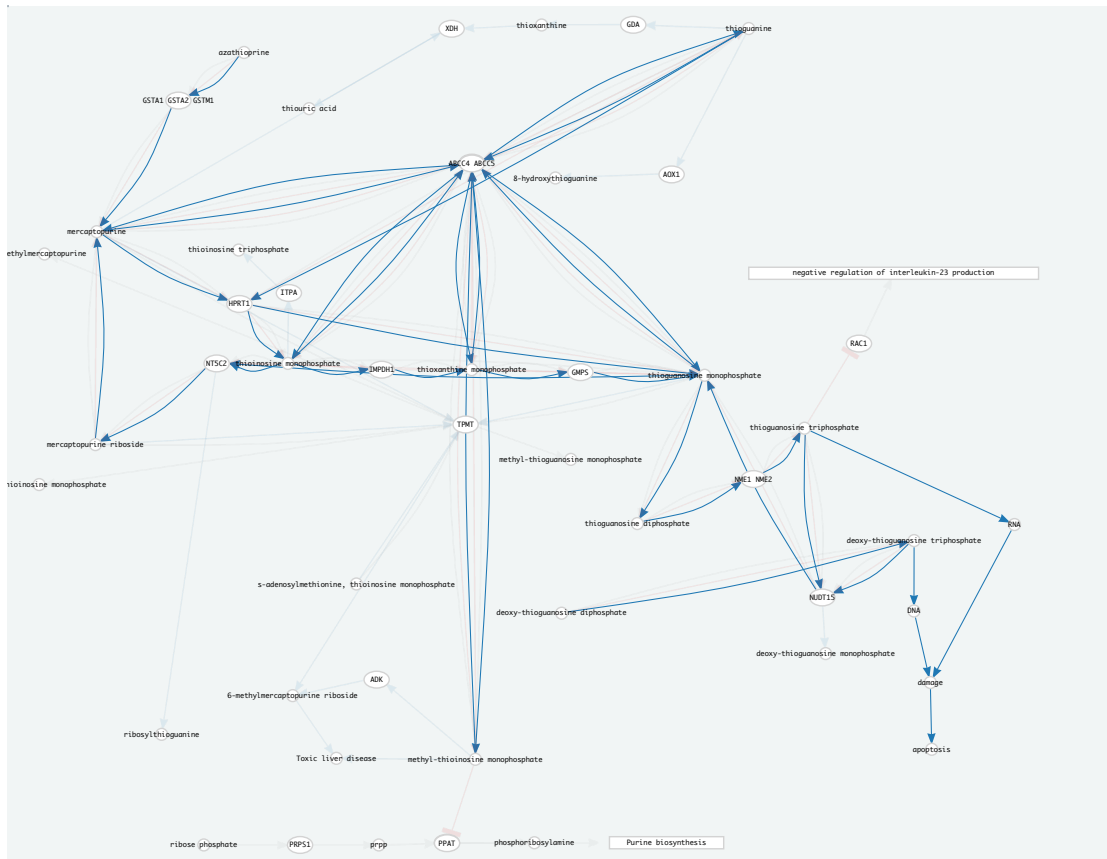


Figura Anexos 19. Tiopurinas, circuito apoptosis

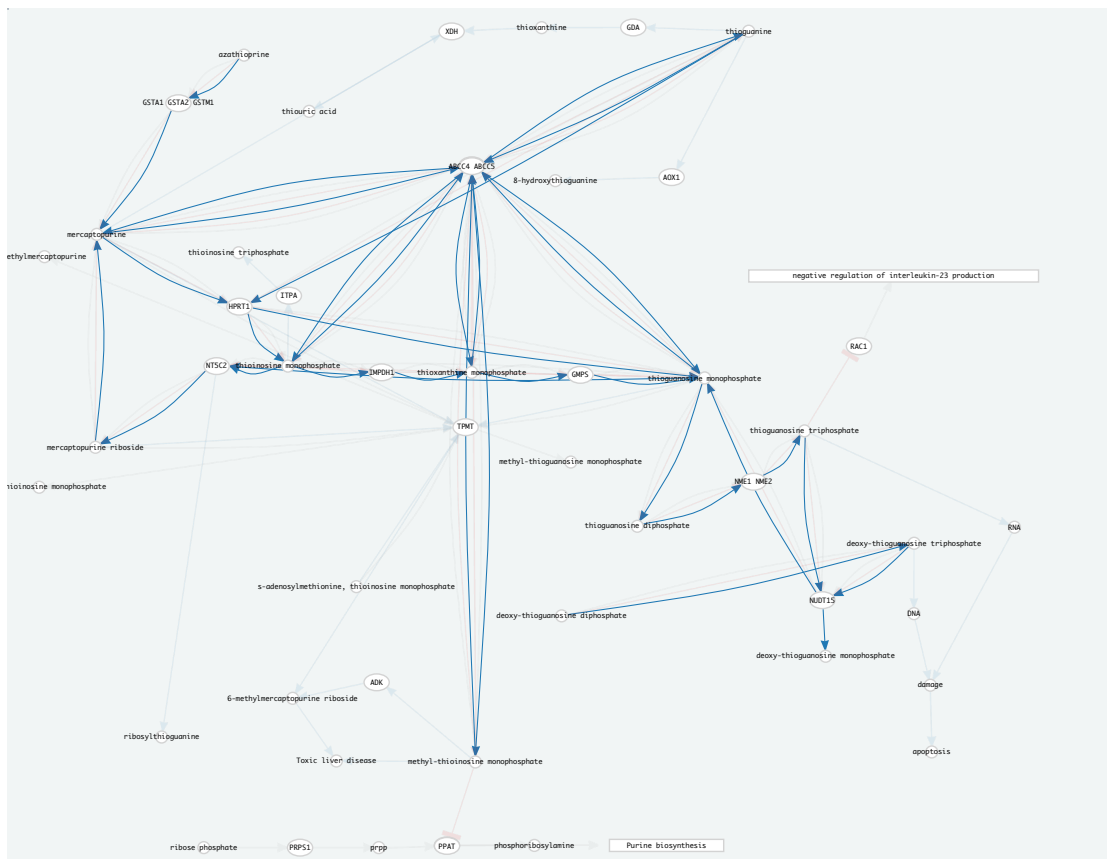


Figura Anexos 20. Tiopurinas, circuito deoxi-tioguanosina monofosfato

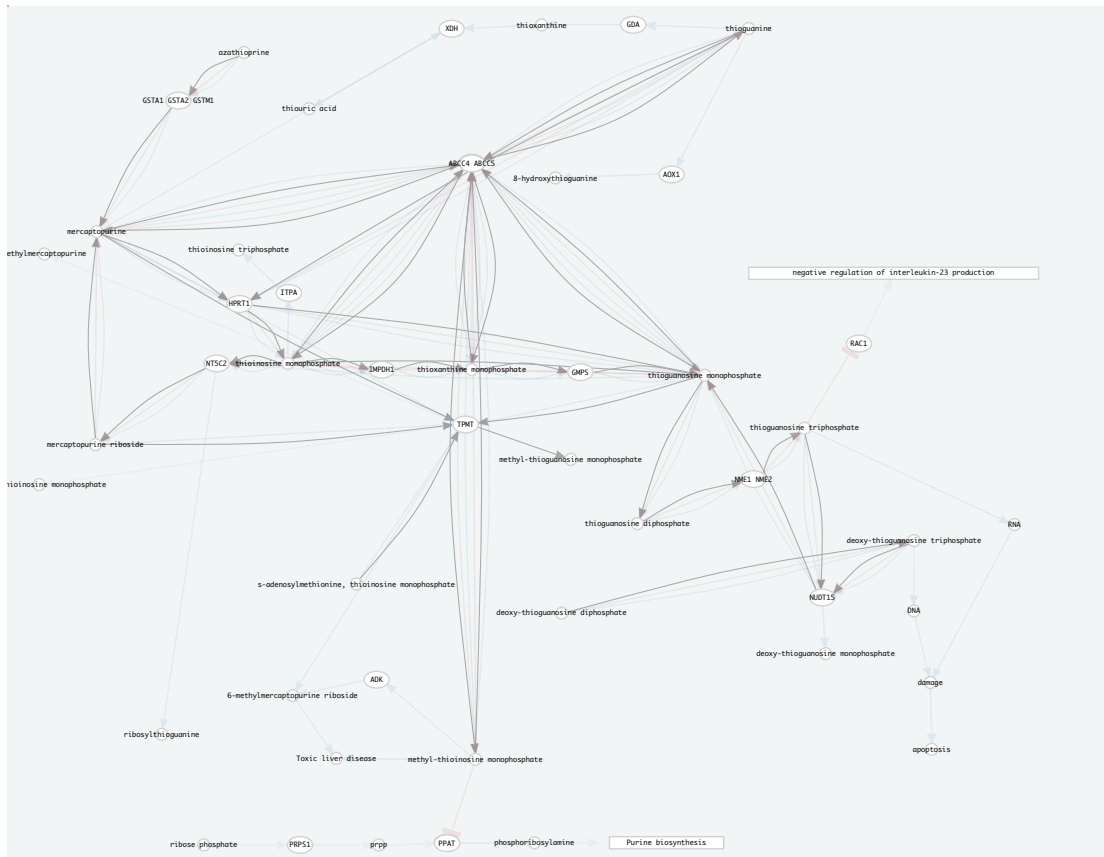


Figura Anexos 21. Tiopurinas, circuito metil-tioguanosina monofosfato

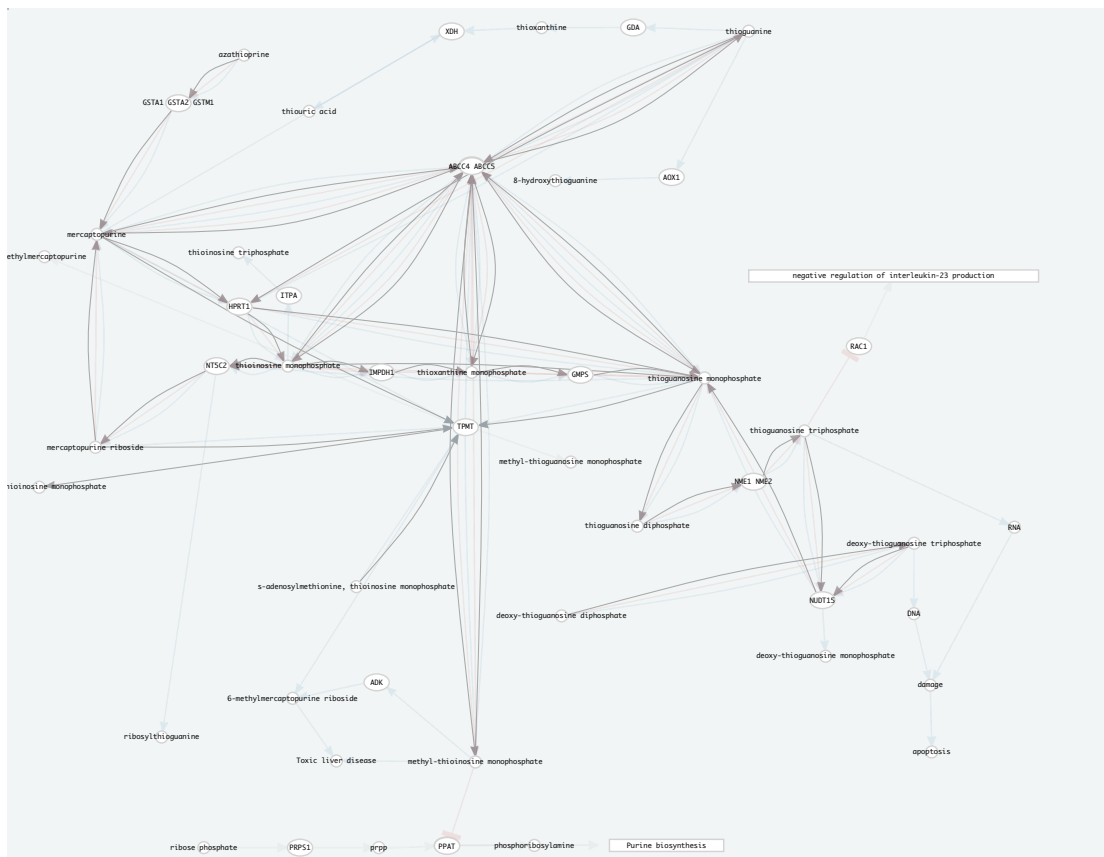


Figura Anexos 22. Tiopurinas, circuito metil-tioinosina monofosfato

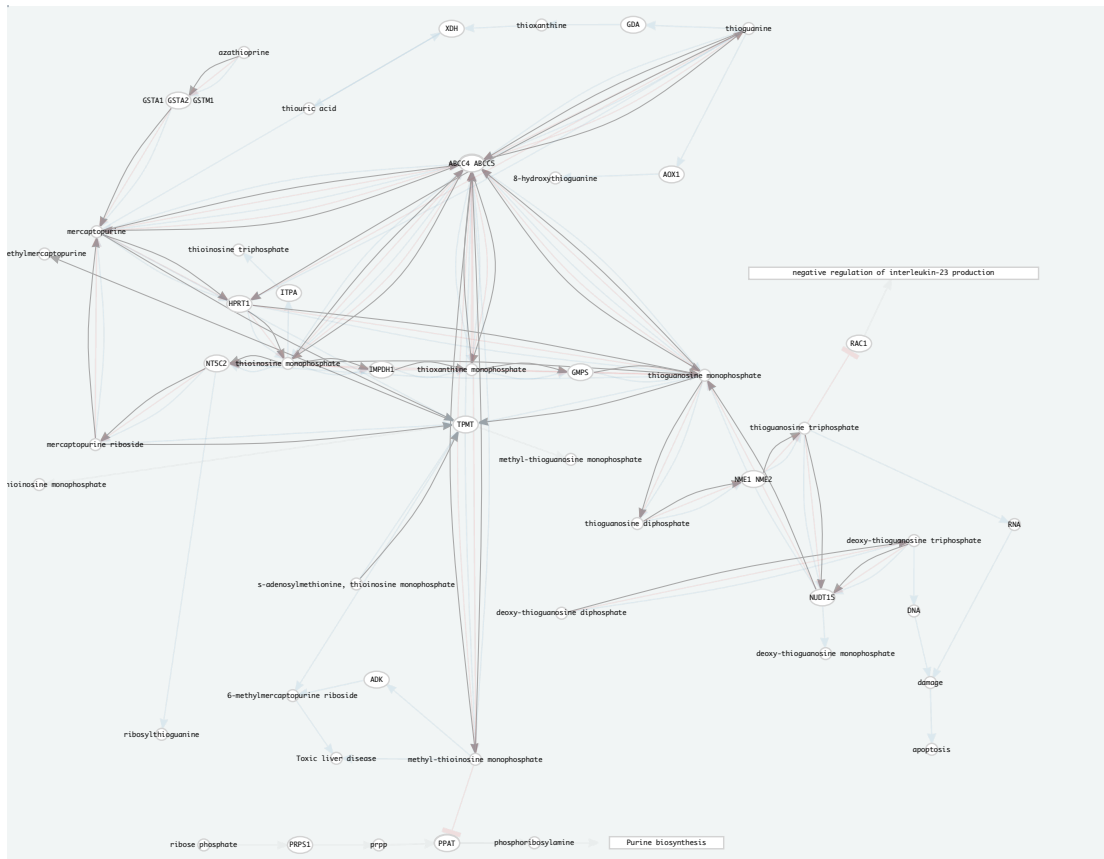


Figura Anexos 23. Tiopurinas, circuito metilmercaptopurina

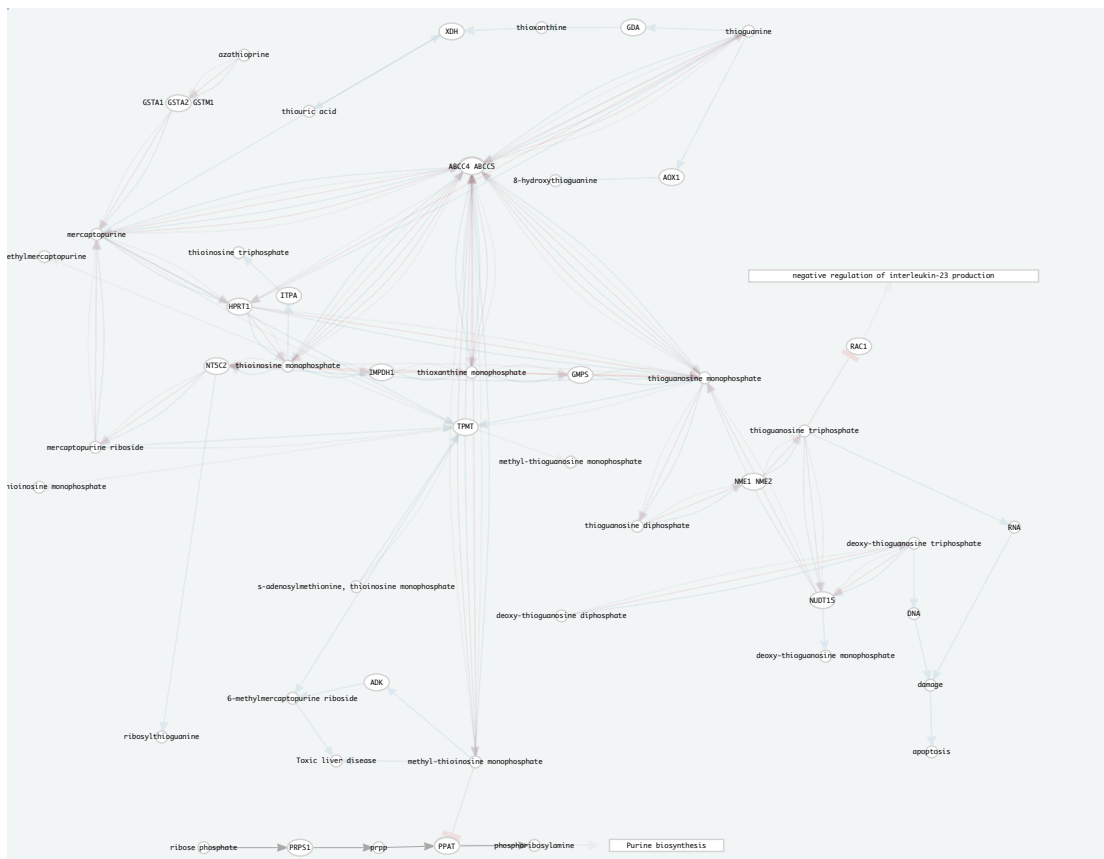


Figura Anexos 24. Tiopurinas, circuito fosforibosilamina

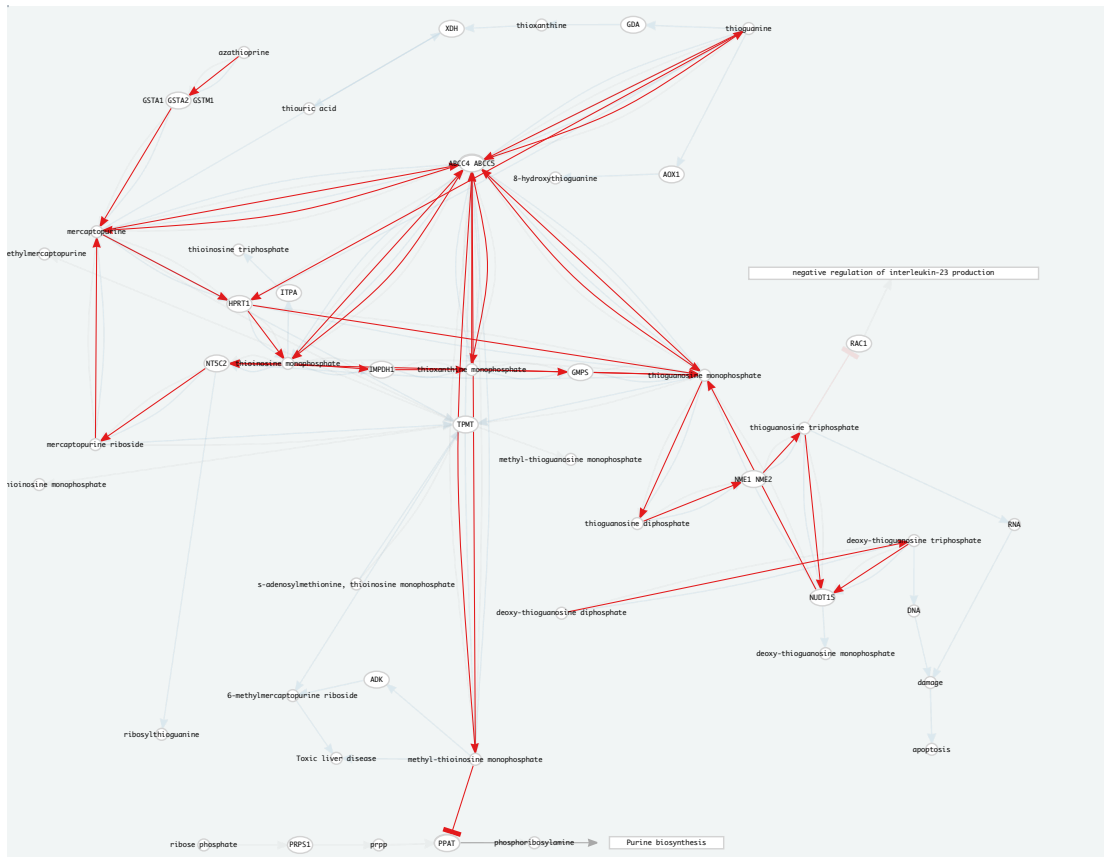


Figura Anexos 25. Tiopurinas, circuito PPAT

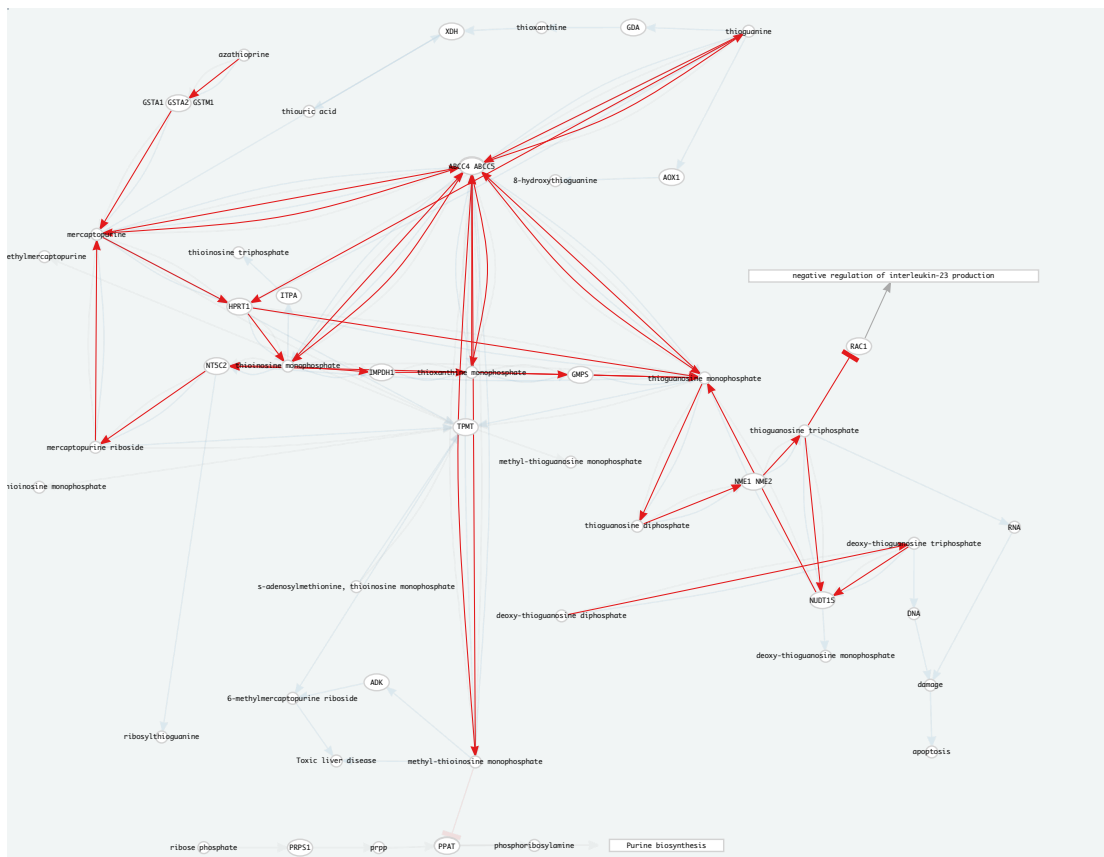


Figura Anexos 26. Tiopurinas, circuito RAC1

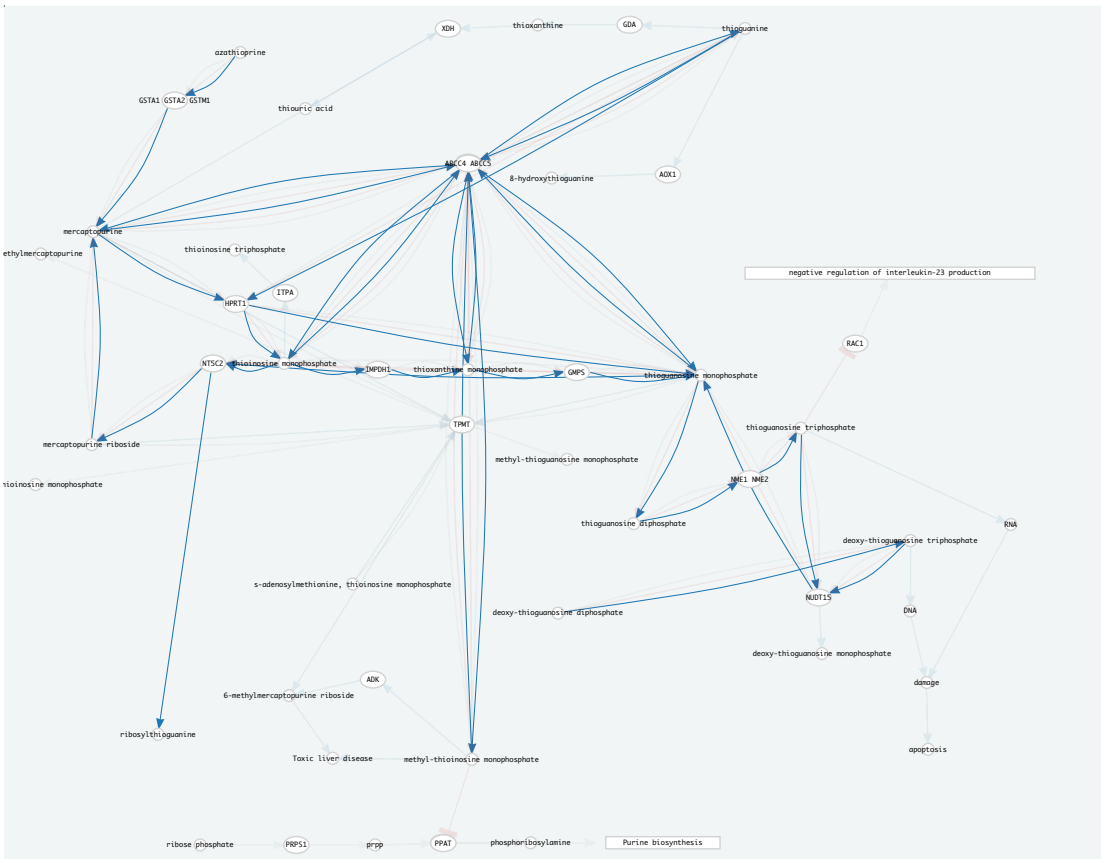


Figura Anexos 27. Tiopurinas, circuito ribosiltioguanina

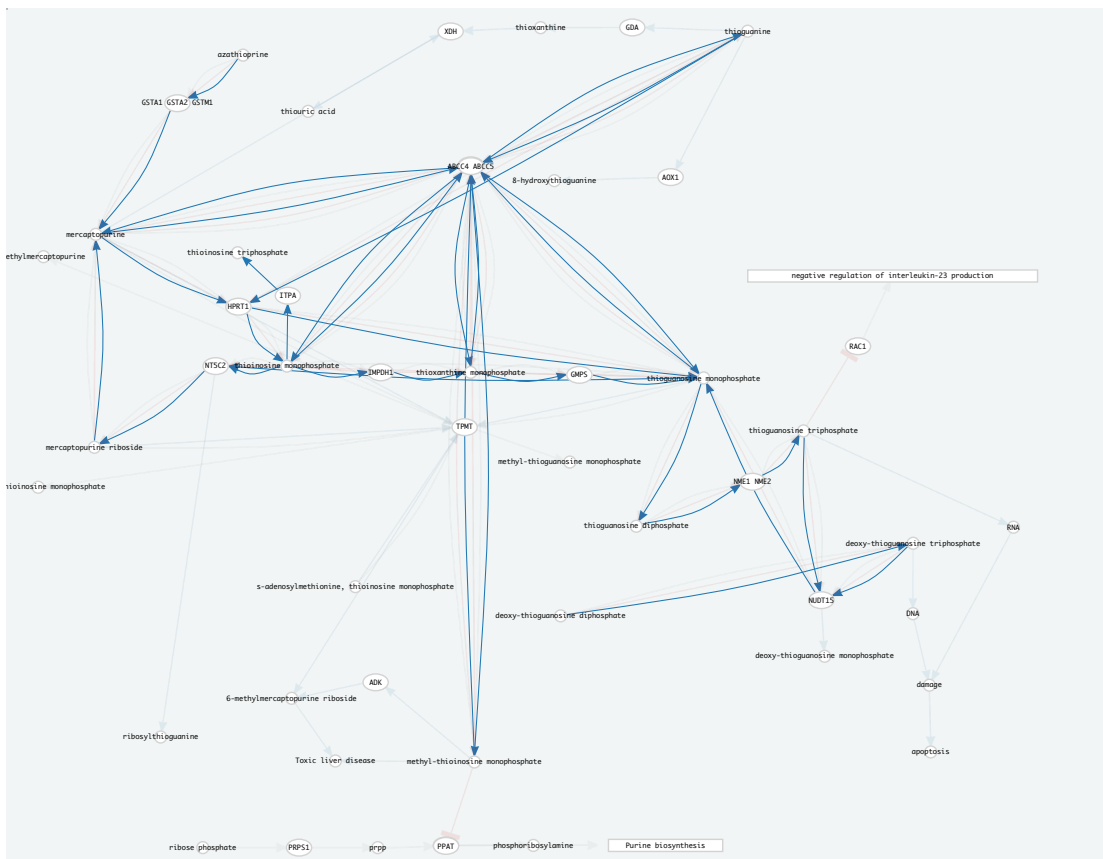


Figura Anexos 28. Tiopurinas, circuito tioinosina trifosfato

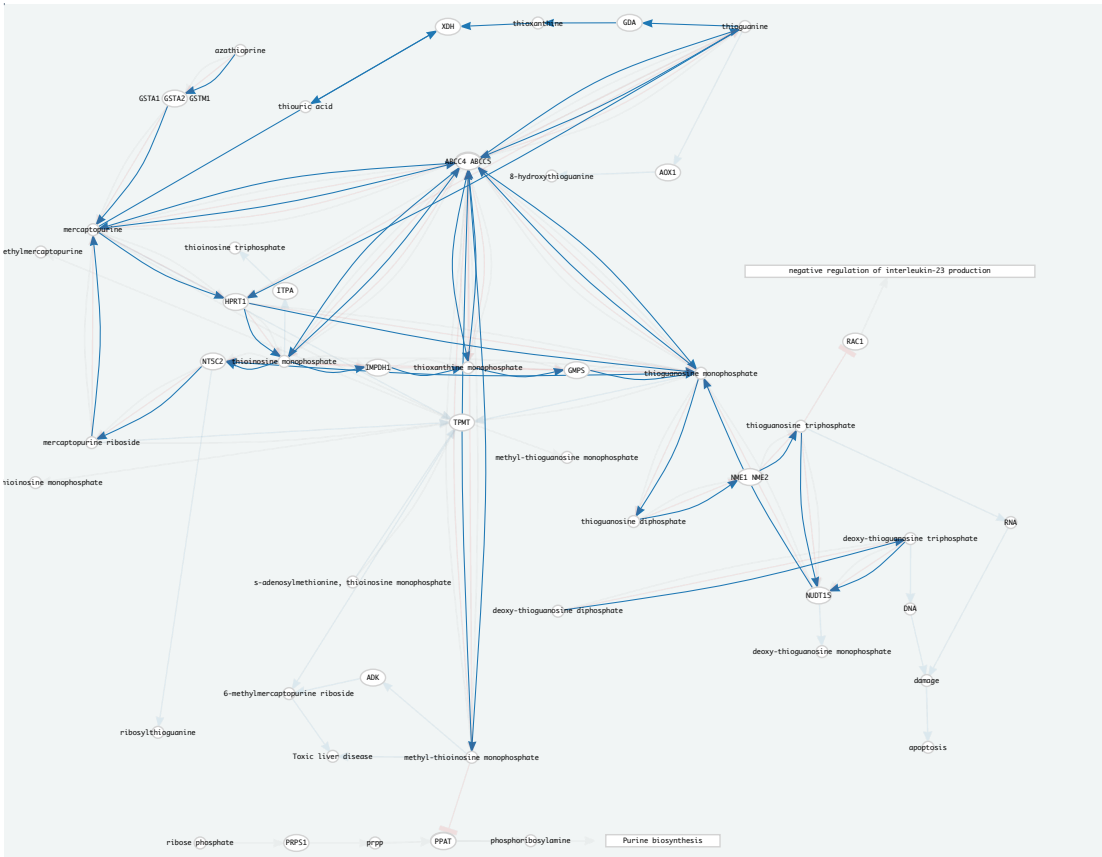


Figura Anexos 29. Tiopurinas, circuito ácido tiourico

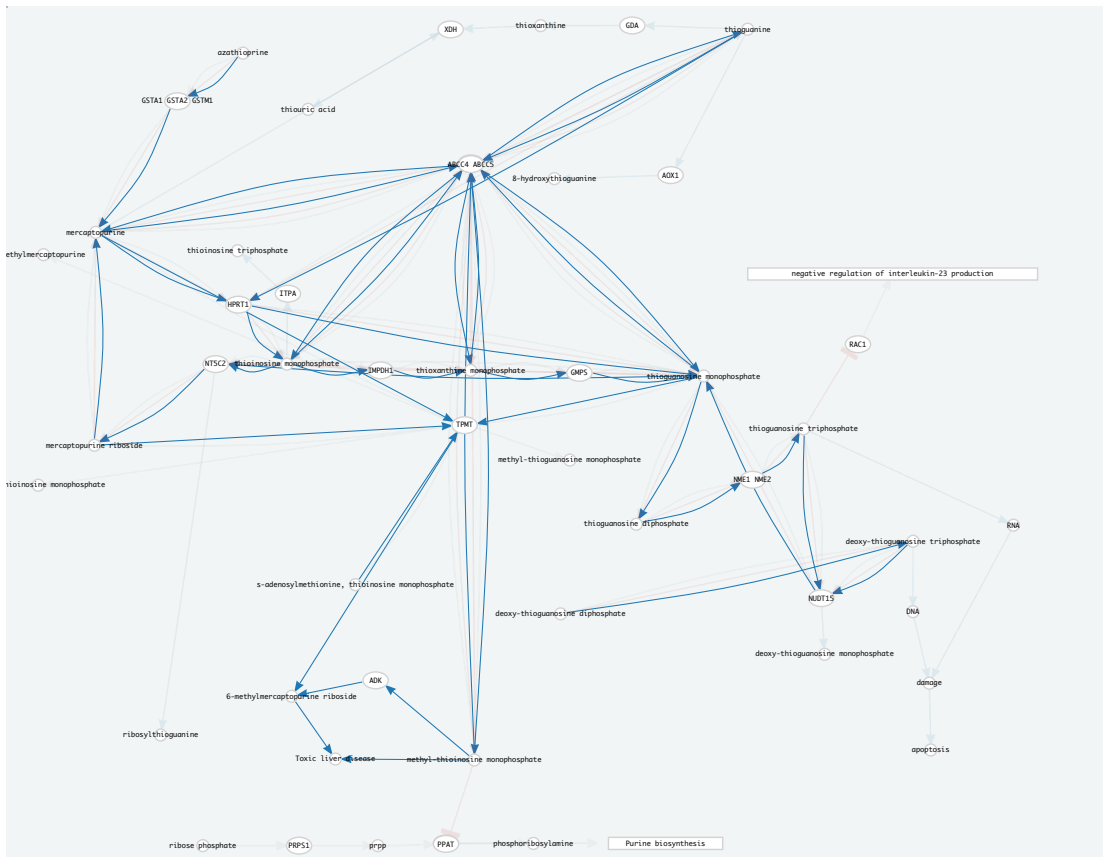


Figura Anexos 30. Tiopurinas, circuito enfermedad hepática tóxica

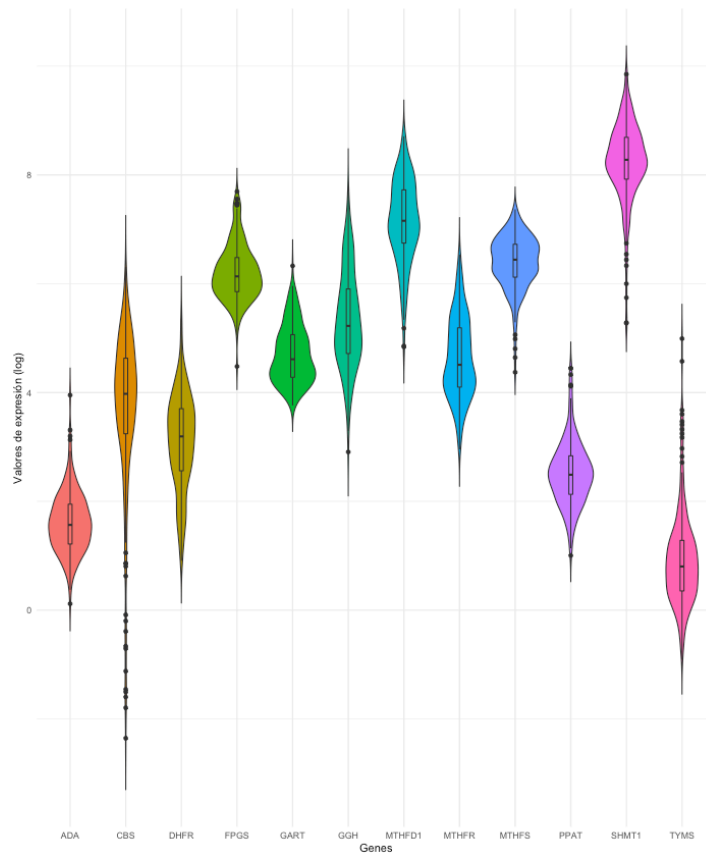


Figura Anexos 31. Violin plot. Análisis de expresión de genes relacionados con la ruta de Metotrexate en tejido hepático sano

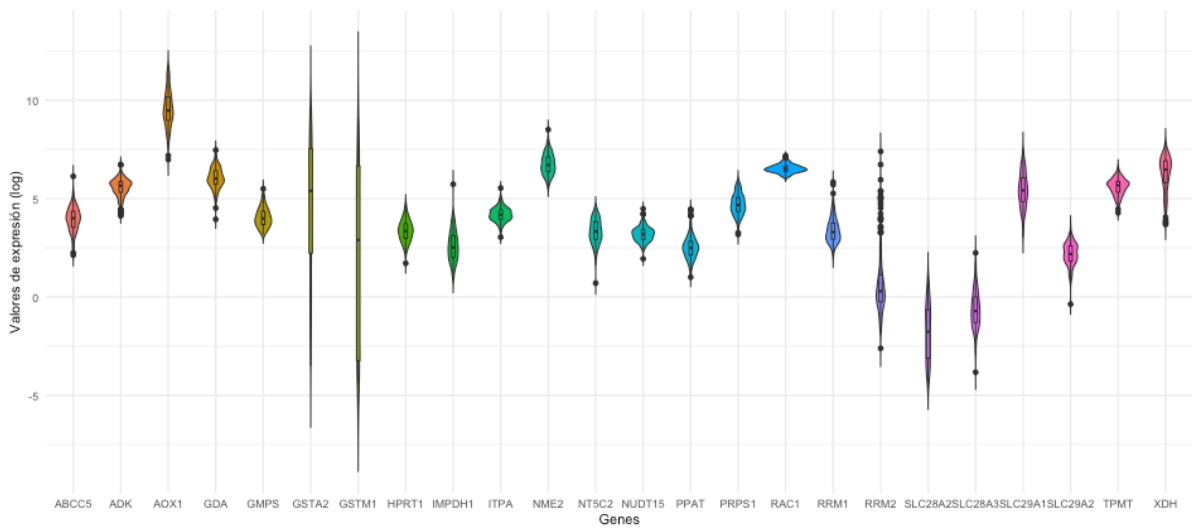


Figura Anexos 32. Violin plot. Análisis de expresión de genes relacionados con la ruta de las Tiopurinas en tejido hepático sano