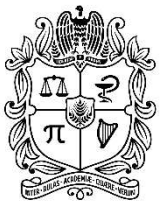


UNIVERSIDAD
NACIONAL
DE COLOMBIA

Evaluación comparativa de herramientas predictivas *In Silico* de variantes de cambio de sentido en genes de interés en farmacogenética: análisis bioinformático y poblacional para el gen *DPYD*

Diego Aeljandro Saldaña Peñaloza

Universidad Nacional de Colombia
Facultad de Medicina, Maestría en Genética Humana
Bogotá, Colombia
2024

Evaluación comparativa de herramientas predictivas *In Silico* de variantes de cambio de sentido en genes de interés en farmacogenética: análisis bioinformático y poblacional para el gen *DPYD*

Diego Aeljandro Saldaña Peñaloza

Tesis investigación presentada como requisito parcial para optar al título de:

Magister en Genética Humana

Director:

Daniel Hernán Mahecha Lopez, MD, M. Sc.

Codirector (a):

Mauricio Rey Buitrago BQ, M. Sc, PhD y Carlos Castro Rojas QF, M. Sc,PhD

Grupo de investigación: Genética clínica

Línea de investigación: Genética y cáncer

Universidad Nacional de Colombia

Facultad de Medicina, Maestría en Genética Humana

Bogotá, Colombia

2024

“La medicina es como una lenta obra de albañilería … Somos afortunados si en el plazo de una vida podemos poner un solo ladrillo”

El médico, Noah Gordon

Declaración de obra original

Yo declaro lo siguiente:

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

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

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

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

Diego A. Saldaña Peñaloza

Fecha 15/01/2024

Agradecimientos

Este documento no es solamente el resultado del trabajo realizado durante mi estancia en la maestría, sino, que adicionalmente es fruto del esfuerzo realizado por mis padres para que yo pudiera tener una educación de calidad, por lo cual mi gratitud será eterna, ya que sin ellos, tal vez mis sueños no serían hoy una realidad.

Agradezco también, la compañía que Andrés me ha dado durante este proceso, brindándome apoyo incondicional y convirtiéndose en una parte importante en mi vida, siempre motivándome a seguir adelante a pesar de cada obstáculo. A mi pequeña bola de pelos, Bruce, quien llegó para el inicio de la maestría y me ha dado días de muchas alegrías.

A mi mentor, el Dr. Daniel Mahecha, quien me brindó todo su conocimiento para que yo pudiera entrar en el bello mundo de la biología computacional. Sin su apoyo, este proyecto y mi desarrollo profesional no hubiera sido el mismo.

Al Dr. Carlos Castro y Dr. Mauricio Rey, quienes desde el inicio de la maestría realizaron el acompañamiento para que yo pudiera aprender e generar un particular interés por la farmacogenética, gracias por su paciencia y apoyo.

A cada uno de los docentes de la maestría; en cada una de las clases, seminarios y en sus consejos, vi su amor hacia la genética y tengo esto como un ejemplo en lo que quiero en mi vida como genetista.

A la Dra. Silvia Maradei y el Dr. Jorge Diaz, quienes me han brindado su apoyo incondicional y bajo su tutoría he podido desarrollar mi carrera como genetista.

Y finalmente a Biotecgen S.A.S, sus directivas y cada uno de sus colaboradores, quienes no solamente me permitieron realizar mi investigación con los datos disponibles en la unidad de analítica de datos, sino que también me han hecho sentir como un miembro más de esta gran familia.

Resumen

Evaluación comparativa de herramientas predictivas *in silico* de variantes de cambio de sentido en genes de interés en farmacogenética: análisis bioinformático y poblacional para el gen *DPYD*.

Gran parte de la toxicidad en pacientes oncológicos tratados con fluoropirimidinas se debe a la pérdida de la actividad enzimática de la dihidropiridina deshidrogenasa, causada por variantes en el gen *DPYD*, por lo cual diversos estudios han propuesto la genotipificación al inicio del tratamiento con estos fármacos. Este estudio buscó determinar la eficacia de diferentes algoritmos *In Silico* de predicción de variantes de cambio de sentido nocivas en el gen *DPYD*, con el propósito de proponer un flujo de evaluación basado en herramientas de anotación con alta sensibilidad y especificidad en la predicción de variantes de interés en farmacogenética en la población colombiana. Para lo cual se planteó la evaluación comparativa de herramientas de anotación *In Silico* en el gen *DPYD* basadas en los hallazgos de una revisión sistemática de alcance, adicional a un análisis descriptivo estructural, la búsqueda de variantes conocidas y así como de variantes no reportadas previamente en un conjunto de datos de secuenciación de exoma de un laboratorio de biología molecular en la ciudad de Bogotá. A partir de la revisión sistemática se seleccionaron los algoritmos BayesDel addAF, BayesDel noAF, Eigen, Eigen-PC, SIFT, MetaSNP, Mutation Assessor, Revel y Provean, los cuales fueron evaluados en 137 variantes en el gen *DPYD*, encontrando que las herramientas con mejor rendimiento fueron PROVEAN, Revel y MetaSNP. En el análisis poblacional, se encontró que, en general, la frecuencia poblacional de variantes conocidas como nocivas, incluyendo *DPYD**2A, era menor al 1%, lo cual es inferior a lo reportado para poblaciones caucásicas, y la de mayor frecuencia fue HapB3. Se identificó la variante c.1127A>C, la cual por herramientas de anotación podría ser nociva, sin embargo, se deben realizar estudios adicionales para confirmar el efecto de la variante. En conclusión, a pesar de que este es un primer acercamiento en el análisis computacional para identificar variantes en el gen *DPYD* en población colombiana, se debe profundizar en los hallazgos reportados en esta investigación, lo cual podría permitir una aplicación de flujos de análisis en farmacogenética acordes a las características poblacionales colombianas.

Palabras clave: Dihidropirimidina deshidrogenasa, valor predictivo de las pruebas, mutación de cambio de sentido, análisis *In Silico*, farmacogenética

Abstract

Comparative evaluation of *in silico* predictive tools of missense variants in genes of interest in pharmacogenetics: bioinformatics and population analysis for the *DPYD* gene.

Much of the toxicity in cancer patients treated with fluoropyrimidines is due to the loss of the enzymatic activity of dihydropyrimidine dehydrogenase, caused by variants in the *DPYD* gene. Therefore, various studies have proposed genotyping before the initiation of these drugs. This study aimed to determine the efficacy of different in silico algorithms for predicting deleterious missense variants in the *DPYD* gene, with the purpose of proposing an *in silico* evaluation flow based on annotation tools with high sensitivity and specificity in predicting variants of interest in pharmacogenetics in the Colombian population. For this aim, a comparative evaluation of in silico annotation tools in the *DPYD* gene was proposed based on the findings of a rapid systematic review, in addition to a structural descriptive analysis, a search for known variants, as well as a search for previously unreported variants in the database of a molecular biology laboratory in Bogotá. Based on the systematic review, the BayesDel addAF,BayesDel noAF, Eigen, Eigen-PC, SIFT, MetaSNP, Mutation Assessor, Revel, and Provean algorithms were selected and evaluated on 137 variants in the *DPYD* gene, finding that the tools with the best performance were PROVEAN, Revel, and MetaSNP. In the population analysis, it was found that, in general, the population frequency of known harmful variants, including *DPYD**2A, was less than 1%, which is lower than reported for Caucasian populations, and the most frequent was HapB3. The variant c.1127A>C was identified, and, according to annotation tools, could be harmful; however, additional studies should be conducted to confirm the effect of the variant. In conclusion, despite being an initial computational analysis to identify variants in the *DPYD* gene in the Colombian population, further investigation is required to validate the findings of this research. This could lead to the development of analysis workflows in pharmacogenetics tailored to the characteristics of the Colombian population.

Keywords: Dihydropyrimidine dehydrogenase, predictive value of tests, missense mutation, *In Silico* analysis, pharmacogenetics.

Contenido

Lista de figuras

Lista de gráficas

Pág.

Lista de tablas

Pág.

Lista de abreviaturas

Abreviatura	*Término*
5-FU	5-fluorouracilo (5-Fluoropirimidina-2,4-diona)
ACC	Exactitud (Acurracy)
ACMG	American College of Medical Genetics
AUC	Área bajo la curva
ADN	Ácido desoxirribonucleico
ARN	Ácido ribonucleico
CPIC	Consorcio de Implementación de Farmacogenética Clínica
DHFU	Dihidrofluorouracilo
dNTPs	Desoxiribonucleótidos o deoxiribonucleótidos
ddNTPs	Dideoxiribonucleótidos
DPD	Enzima dihidropirimidina deshidrogenasa
DPYD	Gen dihidropirimidina deshidrogenasa
DPYS	Dihidropirimidinasa
Esp	Especificidad
FA	Frecuencia alélica
FAD/FMN	Flavín mononucleótido
FBAL	Fluoro-β-alanina
Fe	Hierro
FdUDP	Fluorodesoxiuridina difosfato
FdUTP	Fluorodesoxiuridina trifosfato
FN	Falso negativo
FP	Falso positivo
FUMP	5 fluorouridina monofosfato
FUDP	Fluorouridina difosfato
FUTP	Fluorouridina trifosfato
FUPA	Fluoro-β-ureidopropionato
HWE	Equilibrio de Hardy Weinberg
Kb	Kilobases
kDa	Kilodaltons
MCC	Coeficiente de correlación de Matthews
NADPH/NADP	Nicotinamida-Adenina Dinucléotido fosfato
OMS	Organización Mundial de la Salud
OPTR	Orotato fosforribosil transferasa
pb	Pares de bases
RAM	Reacción adversa a medicamentos
ROC	Característica Operativa del Receptor
RR	Ribonucleótido reductasa
S	Azufre
Sen	Sensibilidad
SNV	Variante de único nucleótido
TK	Timidina quinasa
TP	Timidina fosforilasa
TS	Timidilato sintasa
UPB1	β-ureidopropionasa
VCF	Variant Call Format

VN	Verdadero negativo
VP	Verdadero positivo
VPN	Valor predictivo negativo
VPP	Valor predictivo positivo

Introducción

Con la publicación de la metodología propuesta por Sanger y colaboradores para la secuenciación del ADN a partir del uso de ddNTPs y la actividad natural de las ADN polimerasas en 1977 (1) y la posterior aparición de otras tecnologías para la secuenciación, se ha ampliado la información disponible sobre la función biológica del ADN y con esto se ha profundizado en la investigación sobre las implicaciones de la variación en el ADN humano. Con el uso metodologías para conocer la secuencia del ADN se ha logrado no solamente identificar las funciones de regiones específicas del ADN y correlacionar con la aparición de enfermedad en el ser humano, sino que, enmarcándose dentro de la farmacogenética y farmacogenómica, esto ha permitido establecer posibles dianas terapéuticas en diversas enfermedades y analizar la interacción entre fármacos y el genoma humano, postulando posibles genes modificadores así como biomarcadores de respuesta y toxicidad a fármacos específicos (2,3).

Con la determinación de la secuencia del ADN humano, se han logrado identificar los tipos de variación en el genoma humano, como las variantes de único nucleótido (SNV), que implican el cambio en un solo nucleótido del ADN, siendo las más frecuente en presentación; se estima que estas variantes pueden encontrarse, en promedio una vez cada 300 a 400 pares de bases (4). Las variantes de cambio de sentido, dentro del grupo de SNV, son de particular interés debido a su impacto en la secuencia proteica, ya que pueden alterar la función de la proteína (4). Por esta razón, el análisis de estas variantes no solamente se ha propuesto para establecer las bases moleculares de enfermedades monogénicas, sino también como biomarcadores de riesgo y en el análisis de marcadores de progresión en enfermedades complejas. Desde el punto de vista computacional se han desarrollado diversos enfoques para su análisis, surgiendo herramientas basadas en el alineamiento de secuencias múltiples con evaluación de conservación de los residuos, herramientas basadas en el análisis de las características de la secuencia junto con el impacto a nivel de la estructura en la proteína y los meta-scores que se basan en la ponderación de un efecto a partir de la generación de un puntaje compuesto (5).

La necesidad de ampliar el conocimiento disponible acerca del genoma humano, evaluar el impacto de la presencia de variantes en el ADN y el desarrollo de la medicina de

precisión con la individualización del tratamiento médico, especialmente en pacientes con cáncer, se han desarrollado herramientas que han permitido la correlación de los hallazgos derivados de la secuenciación del genoma con el desarrollo de enfermedad, la anotación de variantes genéticas de interés y la postulación de posibles biomarcadores (6,7); encontrándose que a partir de datos de farmacogenómica se podría explicar más del 80% de la variabilidad en la eficacia y seguridad de los fármacos entre los pacientes expuestos, postulando la importancia de las variantes raras en la aparición de efectos adversos relacionadas con el uso de fármacos y describiéndose más de 240 farmacogenes relacionados con las reacciones adversas a medicamentos (RAM) (8).

Las fluoropirimidinas son fármacos antimetabolitos de interés dentro de la medicina de precisión ya que son usados ampliamente como agentes anticancerígenos en monoterapia o en combinación con otros fármacos o terapias en esquemas de tratamiento principalmente para tumores sólidos (9). Sin embargo, se ha encontrado que hasta un 40% de los pacientes expuestos presentan alguna RAM relacionada con el uso del medicamento, asociadas principalmente con defectos en el catabolismo de las pirimidinas, principalmente en la enzima dihidropirimidina deshidrogenasa (DPD). Por esta razón, es de especial interés el tamizaje de la funcionalidad de esta enzima, no solamente a nivel fenotípico sino también en la identificación de variantes en el gen de la dihidropirimidina deshidrogenasa (*DPYD)* que puedan impactar en la función de la proteína (9–11).

Se ha planteado la importancia del ajuste de dosis del 5-fluorouracilo (5-FU) y sus derivados, donde grupos como el *Dutch Pharmacogenetics Working Group* (DPWG) y el *Clinical Pharmacogenetics Implementation Consortium* (CPIC) han desarrollado guías que recomiendan el tamizaje genotípico previo al inicio del tratamiento (12,13). Sin embargo, las recomendaciones de ajuste de dosis actuales son aplicables sólo cuando se identifican 4 variantes específicas (c.190511G>A, c.1679T>G, c.2846A>T y c.1129–5923C>G (rs75017182, HapB3)) (12) para las cuales hay suficiente evidencia de su relación con la toxicidad y el uso de las fluoropirimidinas. En consecuencia, se presentan dificultades para la toma de una decisión clínica cuando se identifican otras variantes durante la secuenciación del gen que se consideren nuevas o con poca evidencia.

Por esta razón, se planteó una investigación en la cual a partir del uso de herramientas computacionales, se realizó un análisis comparativo de diversos algoritmos *In Silico* de

anotación, planteando identificar los predictores con mayor sensibilidad y especificidad para detectar variantes nocivas en el gen *DPYD*. Adicionalmente, se realizó la búsqueda de variantes nuevas o previamente reportadas en una muestra de datos de secuenciación de exoma de pacientes colombianos en un laboratorio de biología molecular en la ciudad de Bogotá. Todo esto con el fin de generar información frente a la presencia de variantes accionables en el gen *DPYD* en individuos colombianos y avanzar en la generación de información acerca de alternativas del tamizaje de pacientes de riesgo con el uso de fluoropirimidinas bajo el contexto colombiano.

Justificación y planteamiento del problema

La variabilidad en la respuesta a los fármacos derivada del componente genético individual se estima en un 20 al 30%, lo cual también afecta la incidencia y severidad de las RAM (14). Hasta el 30% de las RAM con ingreso hospitalario son causadas por medicamentos con anotaciones en farmacogenética, en muchos casos con indicación de genotipificación preventiva (14), por lo cual durante los últimos años ha aumentado el interés en generar evidencia frente a la aplicación de estas medidas para prevenir la aparición de toxicidad relacionada con el uso de fármacos, especialmente en enfermedades de alto costo como el cáncer.

Dentro de los fármacos de uso frecuente en oncología están las fluoropirimidinas, las cuales tienen un amplio uso en el manejo de tumores sólidos, especialmente del tracto gastrointestinal, mama y tejidos blandos de cabeza y cuello, generalmente con adecuada respuesta y tolerabilidad durante su uso. Sin embargo, se han reportado pacientes con eventos de toxicidad grave y potencialmente mortales durante los primeros ciclos de tratamiento, lo que supone la necesidad de cambio de esquema de manejo y retrasos en el tratamiento con generación de costos adicionales que impactan en el pronóstico y desenlace final de la enfermedad (15). Por este motivo, se ha planteado la necesidad de identificar biomarcadores de respuesta y toxicidad con el uso de estos fármacos, siendo el gen *DPYD* y su producto proteico, la enzima DPD, uno de los más estudiados. Esta enzima es determinante durante la inactivación de las fluoropirimidinas para su posterior eliminación, por lo cual su actividad funcional se relaciona con las concentraciones de metabolitos del catabolismo de estos fármacos y la aparición de la toxicidad (9,16).

Se estima que aproximadamente entre el 10-30% de los pacientes tratados con fluoropirimidinas presentan eventos de toxicidad severa (grado 3-4) dependiendo del régimen de tratamiento utilizado y aproximadamente el 0,1-1% de los pacientes presentan mortalidad derivada del uso de estos fármacos (17,18), sin embargo, los datos de incidencia de la toxicidad y eventos adversos están poco documentados en muchos centros médicos y no son de notificación obligatoria en general en los centros médicos, por lo cual los datos de farmacovigilancia disponibles frente a estos eventos pueden no ser acordes a la realidad. Se estima que el 20-60% de las toxicidades graves asociadas al uso

de las fluoropirimidinas se deriva de defectos de la actividad de DPD y un 20-30% de las toxicidades graves tempranas se relacionan con la presencia de variantes nocivas en el gen *DPYD* (18), teniendo en cuenta una prevalencia de la deficiencia de DPD del 3% al 5% en pacientes de origen europeo y del 8% en pacientes de Origen africano (17). Adicionalmente el Instituto Nacional de Salud de Estados Unidos estima un aproximado de 1.300 muertes año atribuibles a la deficiencia del DPD (17).

Se ha estimado que el genotipado preventivo de *DPYD* y el ajuste de dosis en portadores de la variante *DPYD**2A reduce del 73 % al 28 % el riesgo de toxicidad, calculándose una reducción de 5 veces en la prevalencia de toxicidad gastrointestinal (17), por lo cual se han desarrollado guías con recomendaciones clínicas acerca del uso de la genotipificación previa al inicio de tratamiento con fluoropirimidinas para la disminución del riesgo de aparición de toxicidad mediante el ajuste de dosis, donde se plantea la identificación de las variantes *DPYD**2A (c.1905+1G>A, IVS14+1G>A), *DPYD**13 (c.1679T>G), c.2846A>T y c.1236G>A (en desequilibrio de ligamiento con c.1129–5923C>G) para clasificar a los pacientes en metabolizadores normales, metabolizadores intermedios y metabolizadores lentos y, según el puntaje de riesgo, aproximar la mejor dosis de acuerdo al riesgo de aparición de toxicidad (12,13). No obstante, a pesar de la utilidad de las recomendaciones, su aplicación de forma generalizada es limitada dada la escasa validación de su utilidad en poblaciones diferentes a la caucásica. Esto se debe a la presencia de variantes diferentes a las establecidas en las guías en otras poblaciones, a menores frecuencias alélicas de las reportadas de las variantes recomendadas para genotipificar en las poblaciones no caucásicas y la deficiencia en las recomendaciones sobre la conducta ante nuevas variantes encontradas durante la secuenciación masiva del gen *DPYD*.

Para el caso de Latinoamérica son pocos los estudios poblacionales que informen de la presencia de las variantes de interés en el gen *DPYD*, encontrando datos principalmente en bases de datos poblacionales generales como el caso de 23andMe que ha informado una frecuencia de DPYD*2A de 0,26% y de p.D949V de 0,43% en población hispana-latina (17). Sin embargo, en estudios realizados teniendo en cuenta la heterogeneidad poblacional latinoamericana se ha encontrado que las variantes predictivas documentadas en poblaciones europeas (c.1905+1G>A, c.1236G>A y c.2846A>T) son extremadamente bajas en población trihíbrida ecuatoriana y en amerindios amazónicos las frecuencias poblacionales de variantes en farmacogenes incluyendo a *DPYD* son muy distintas a las

evidenciadas en africanos y europeos, por lo tanto, un cribado genético con estas variantes no sería informativo en pacientes oncológicos (19,20). Para el caso colombiano, en un estudio de 509 individuos se encontró que las variantes con mayor prevalencia son los alelos *DPYD**9 (69,8%) y *DPYD**5 (21,6%), sin embargo, en PharmVar y CPIC se informa que estas variantes no generan cambio en la función proteica (21).

Actualmente existe una gran preocupación frente a las implicaciones derivadas de la toxicidad asociada al uso de fármacos, reportándose que las reacciones adversas a medicamentos son aproximadamente el 6% de los ingresos hospitalarios y el 9% de los gastos hospitalarios, donde, según el tipo de paciente y sus comorbilidades, el riesgo de RAM puede ser diferente, teniendo mayor frecuencia en los trastornos cardiovasculares (20-30%), el cáncer (25-30%) y los trastornos del sistema nervioso central (10-20 %) en los países desarrollados (14). Por tanto, desarrollar protocolos que permitan la implementación generalizada de herramientas en medicina de precisión se ha convertido en una prioridad en la práctica clínica, donde uno de los genes de interés es el gen *DPYD*.

Dada la dificultad de la aplicación del tamizaje planteado en las guías de ajuste de dosis basado en genotipo en *DPYD* derivado de las diferencias en las frecuencias poblacionales de las variantes de interés, así como la posibilidad de encontrar nuevas variantes no caracterizadas, se plantea la necesidad de identificar herramientas *In Sillico* de anotación con alta sensibilidad y especificidad en identificar variantes potencialmente nocivas con uso a nivel clínico. Lo anterior no solamente se justifica por el impacto positivo que tiene la identificación de pacientes con mayor riesgo de toxicidad, sino también por los beneficios que tiene el análisis *in silico* vs el análisis funcional en laboratorio, donde el último presenta altos costos y representa una gran carga frente al tiempo y necesidades requeridos para la validación a nivel clínico o *In Vitro* (22,23).

Adicionalmente, en el contexto colombiano, con el incremento de la información derivada de la secuenciación del ADN por su uso en el diagnóstico médico, permitido por el modelo actual de salud, plantear un análisis inicial de la información genómica a nivel computacional podría suponer una ventaja en la reducción de costos en investigación e implementación de la medicina de precisión en el país, ya que al disponer de datos de secuenciación, agrupar las variantes de acuerdo a la predicción de su efecto, permitiría analizar solamente las variantes potencialmente nocivas o con una predicción incierta,

reduciendo así el número de variantes que requieren una validación bioquímica adicional de su efecto, contrario a un enfoque basado solamente en el análisis de laboratorio donde se requiere el análisis individualizado para cada variante indiferente de su efecto, y actualmente no se dispone para su uso clínico, como si lo es el análisis de datos derivados de secuenciación. También, este tipo de enfoque permitiría el desarrollo de herramientas robustas con aplicación en la clínica, y al momento de implementar modelos basados en genotipo, no existiría un conflicto frente a quien asumiría los costos de dichas intervenciones, ya que actualmente a se encuentra incluidas dentro de lo beneficios ofrecidos por el sistema de salud colombiano.

Por lo cual se buscó identificar algoritmos de predicción de impacto *in silico* que sean precisos en identificar variantes en el gen *DPYD* con un efecto deletéreo en el producto del gen, con lo cual se propuso un flujo de evaluación de variantes nuevas encontradas durante la secuenciación masiva del gen, siendo este el primer paso en establecer un mecanismo de análisis de variantes de alto interés con el cual se podrían identificar pacientes con riesgo de toxicidad y definiendo grupos de riesgo a partir de estudio de genotipo, lo cual se podría plantear como un alternativa para definir pacientes que puedan ser llevados a fenotipificación posterior a la categorización a partir de genotipo, bajo la aplicación de un modelo mixto de ajuste de dosis basado en fenotipo y genotipo, como ya ha sido previamente descrito en la literatura (24,25).

Por todo lo anterior, se desarrolló una investigación en la cual se planteó identificar algoritmos de predicción *In Sillico* precisos que pudieran ser usados en la identificación de variantes en el gen *DPYD* accionables desde el punto de vista farmacogenético, lo cual se considera podría ser la piedra angular al momento de establecer protocolos de evaluación de variantes nuevas en genes de interés en farmacogenética en el contexto colombiano; lo cual se plantea dado el acceso libre de estos algoritmos de anotación, la disponibilidad de tecnologías para la secuenciación en el país, la posibilidad del análisis poblacional de datos disponibles en laboratorios y las actuales necesidades en salud pública en Colombia.

Marco teórico

3.1 Medicina personalizada, farmacogenética y toxicidad relacionada con el uso de fármacos en cáncer.

Las reacciones adversas son definidas como “una reacción apreciablemente dañina o desagradable que resulta de una intervención relacionada con el uso de un medicamento; prediciendo el peligro de la administración futura y justifican la prevención, o el tratamiento específico, o la alteración del régimen de dosificación, o la retirada del producto”, siendo la reacción adversa un evento nocivo y no intencionado que constituye una importante causa de morbilidad y mortalidad (26,27); la aparición de los efectos adversos se asocian frecuentemente con errores en la prescripción, uso inadecuado incluyendo el abuso, y reacciones en función de la farmacología del fármaco, donde se debe tener en cuenta la dosis del fármaco, el curso temporal de la reacción y factores de susceptibilidad relevantes (factores genéticos, patológicos y otras diferencias biológicas) (26).

Para el caso del cáncer, los fármacos quimioterapéuticos son agentes altamente tóxicos, con brechas de seguridad estrechas y con requerimientos de dosis altas, por lo cual se ha estimado una incidencia aproximada de presentación de algún evento adverso del 50% entre los pacientes tratados (28). Sin embargo, a pesar que se ha estimado que el 7% de los ingresos hospitalarios en la población adulta general y hasta el 20% de los ingresos hospitalarios en pacientes mayores se deben a eventos de toxicidad terapéutica, son pocos los reportes frente a estos eventos en población oncológica a pesar de estar bien documentadas (29), por lo cual han surgido herramientas como la Common Terminology Criteria for Adverse Events planteada por el National Cancer Institute de los Estados Unidos, permitiendo una descripción estandarizada de la severidad de los eventos de toxicidad y así un intercambio óptimo de información sobre la seguridad de las terapias antineoplásicas y el tratamiento (30), con lo cual también se ha planteado la identificación de biomarcadores genómicos enfocados en la diferencia entre la severidad de los eventos adversos entre pacientes y posibles predictores del riesgo.

Con el uso de la secuenciación del ADN de forma masiva se ha logrado ampliar el conocimiento de la relación del genoma humano con la respuesta a diversos fármacos, lo

que ha impactado en la práctica clínica con la aplicación de enfoques basados en la personalización y precisión del tratamiento. En oncología, se busca la identificación de marcadores tanto de origen germinal, como en el tejido tumoral que permitan generar nuevas alternativas de manejo a partir de la caracterización molecular de la enfermedad mediante biomarcadores clínicamente accionables (31,32). Junto con el aumento en la investigación frente a terapia personalizada, esfuerzos como The Cancer Genome Atlas (TCGA) e International Cancer Genome Consortium (ICGC) han brindado una visión del panorama genómico del cáncer llevando a una mejor comprensión de los mecanismos subyacentes de la enfermedad. Esto ha generado avances en la aparición de la terapia molecular dirigida, la creación a gran escala de biobancos de tejidos de modelos tumorales, nuevas alternativas de diagnóstico como la biopsia líquida y la evaluación de predictores de carcinogénesis dentro de la compresión de la evolución clonal y la heterogeneidad intratumoral (33).

Uno de los retos en la aplicación de la medicina de precisión en cáncer es que a partir de la cuantificación del riesgo individual se pueda establecer la dosis correcta del medicamento correcto, para el paciente correcto en el momento correcto, según los perfiles genéticos del cáncer y las firmas mutacionales asociado al genotipo individual; lo cual se ha ido logrando con el desarrollo de las diferentes ciencias ómicas y el aumento de la capacidad en el análisis y almacenamiento de grandes cantidades de datos (34). Una de las potenciales áreas de investigación en farmacogenética es la intervención en la prevención de los eventos adversos derivados de uso de fármacos, esto dado que se ha encontrado que hasta el 3,6% de los ingresos hospitalarios generales en Europa son debidos a RAM y el 10% de los pacientes hospitalizados presentan efectos adversos durante su estancia hospitalaria con una estimación de costos médicos directos de 150 000 billones de dólares al año en Estados Unidos (14,35).

El enfoque para la identificación de genes de interés en la prevención de aparición de toxicidad se basa en la clasificación de la respuesta farmacogenómica de acuerdo con los genes relacionados con la patogénesis de la enfermedad, el mecanismo de acción de los fármacos, con el metabolismo y transporte del fármaco y sus metabolitos, así, como el análisis de las vías moleculares y metabólicas asociadas con la aparición de los efectos pleiotrópicos (36), por lo cual se ha planteado la necesidad del análisis de grandes bases de datos frente a la interacción de los fármacos y la relación fármacos-genoma que

permitan identificar potenciales biomarcadores que puedan ser usados al momento de prevenir el aumento en la incidencia de RAM (14). El cáncer es de particular interés por la heterogeneidad de la respuestas de los pacientes a los agentes anticancerígenos y las dificultades frente al estrecho índice terapéutico, planteando la necesidad de definir la individualización de los tratamientos médicos (37).

En la búsqueda de la individualización del tratamiento diferentes grupos de investigación como el consorcio de Implementación de Farmacogenética Clínica (CPIC), el Grupo de Trabajo de la Asociación Real Holandesa para el Avance de la Farmacia y Farmacogenética (DPWG), la Red Canadiense de Farmacogenómica para la Seguridad de los Medicamentos (CPNDS) y la Red Nacional Francesa de Farmacogenética (RNPGx) han establecido diferentes pautas a partir de los datos disponibles en farmacogenética, donde se plantea evaluar la variación alélica de los farmacogenes de interés, evaluando alternativas frente a la dosificación basada en genotipo y fenotipo (38).

3.2 El caso del 5-FU y sus análogos

Las fluoropirimidinas son un grupo de fármacos antimetabolitos que afecta el metabolismo tisular de la timina, con un mecanismo de acción basado en la formación de pares de bases adenina-uracilo en el ADN, la inhibición de la timidilato sintasa (TS) e inhibición de la síntesis proteica, con indicación de uso en monoterapia o politerapia principalmente en tumores del tracto gastrointestinal (39,40). El primero de estos fármacos en ser descrito fue el 5-fluorouracilo (5-FU), el cual mostró una actividad anticancerígena al incorporarse en los carcinomas hepatocelulares en rata generando citotoxicidad sobre el tumor (40). A partir del éxito del efecto citotóxico del 5-FU, se planteó el diseño de fármacos derivados de la estructura del 5-FU que permitieran su administración de forma oral, lo cual conllevó el desarrollo de profármacos (Tegafur y capecitabina) estables frente al ácido gástrico, permitiendo su absorción a nivel de la mucosa del tracto gastrointestinal. Estos presentan un efecto igual al evidenciado con la molécula original al convertirse en 5-FU a nivel hepático, con mejoría en los tiempos de administración, lo que ha permitido su uso para el tratamiento del cáncer de colon en el entorno adyuvante, así como para el tratamiento de cáncer de mama y cáncer gástrico (41,42).

El metabolismo del 5-FU a nivel celular se ha estudiado ampliamente, lo cual ha permitido identificar diferentes proteínas involucradas no solo en su biotransformación sino también en su transporte. Se ha podido describir que el transporte de la molécula de 5-FU al interior de la célula depende de la familia de los transportadores ABC (ABCC3 y ABCB1), encontrando que variantes en los genes que codifican estas proteínas se asocian a resistencia a este agente antineoplásico, lo cual se ha evidenciado en los casos de resistencia en cáncer colorrectal (43,44). También se ha documentado que el transportador SLC22A7 está involucrado en la captación del 5-FU, por lo cual es objeto de estudio frente a la predicción de respuesta celular al tratamiento quimioterapéutico (45).

A nivel intracelular, el 5-FU es biotransformado por diversas enzimas implicadas en el metabolismo de las pirimidinas, donde un paso limitante de la velocidad del catabolismo del fármaco es la conversión de 5-FU a dihidrofluorouracilo (DHFU) por DPD estimándose que aproximadamente el 80% del fármaco que ingresa a la circulación es biotransformado por esta enzima (46,47). Posterior a la formación de DHFU, continúa el proceso de catabolismo de la molécula para su eliminación con la conversión del DHFU en fluoro-β-ureidopropionato (FUPA) y posteriormente en fluoro-β-alanina (FBAL) por la dihidropirimidinasa (DPYS) y la β-ureidopropionasa (UPB1) como enzimas implicadas en este proceso (46).

Uno de los principales efectos del 5-FU es la interrupción en la generación de algunos de los desoxinucleótidos intracelulares necesarios para la replicación del ADN, lo cual se deriva del uso a nivel intracelular que se da al 5-FU, donde este es convertido en 5 fluorouridina monofosfato (FUMP) por la orotato fosforribosil transferasa (OPTR), producto que es fosforilado en 2 ocasiones formando el fluorouridina difosfato (FUDP) y el fluorouridina trifosfato (FUTP), este último siendo incorporado al ARN o transformado en fluorodesoxiuridina difosfato (FdUDP) por la ribonucleótido reductasa (RR), para posteriormente ser fosforilado formando el fluorodesoxiuridina trifosfato (FdUTP), el cual se incorpora al ADN durante la replicación (42,46,47). La fluorodesoxiuridina (FdUMP) surge del 5-FU debido a su conversión indirecta por timidina fosforilasa (TP) y timidina quinasa (TK) o por desfosforilación del FdUDP. El FdUMP es el principal metabolito activo implicado en la inhibición de la actividad de la timidilato sintasa (TS), ya que está implicado en el mayor mecanismo citotóxico de este grupo de fármacos (42,46). Lo anteriormente descrito se esquematiza en la figura 1.

El efecto anticancerígeno de las fluoropirimidinas se asocia con la formación de sitios abásicos (generando daños en el ADN), la incorporación de FUTP al ARN reemplazando hasta un 50% del uracilo (lo que conlleva a una disrupción y procesamiento erróneo del ARN) (42,47–49) y con la inhibición de la timidilato sintasa (siendo el mayor efecto antitumoral del 5-FU), la cual tiene una gran importancia en la síntesis *de novo* de timidina. La inhibición de esta enzima produce un agotamiento de dTTP y un aumento de dUTP impactando de forma negativa la síntesis y reparación del ADN (50).

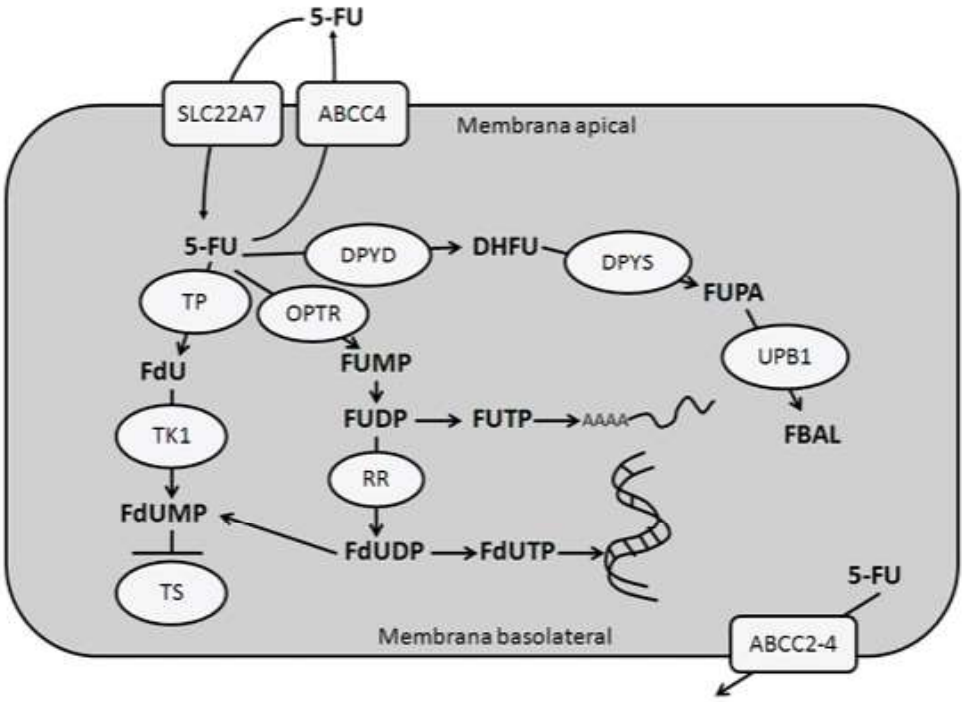


Figura 1. Tomado de: Castro. C et al. (2014) Metabolismo y mecanismo de acción del 5-FU en la célula tumoral.

Convenciones: timidina fosforilasa (TP), orotato fosforribosil transferasa (OPTR), dihidropirimidina deshidrogenasa (DPYD), la timidina cinasa (TK1), de 5-fluoro-2'-deoxiuridina-5'-monofosfato (FdUMP), timidilato sintasa (TS), 5-fluorouridina-5'-trifosfato (FUTP), ribonucleótido reductasa (RR) se da la formación 5-fluorouridina-2'-deoxiuridina-5'-trifosfato (FdUTP)

3.3 El gen *DPYD* y la enzima DPD

El gen *DPYD* se ubica en el locus 1p22 constando de 23 exones que abarcan una región de aproximadamente 150 kb de longitud siendo el exón 15 (69 pb) el más pequeño y el exón 23 (1404 pb) el más grande y para el caso de los intrones el tamaño ronda desde 1

kb hasta los 20 kb, dentro del gen se han descrito los dominios funcionales en la proteína DPD que incluyen sitios de unión de consenso putativos para NADPH/NADP y FAD/FMN, un sitio de unión de uracilo y agrupaciones (4Fe-4S) (figura 2) donde el motivo de unión de NADPH se encuentra en los exones 9 y 10, el dominio de unión de FAD / FMN en los exones 11 y 12, y el sitio de unión de uracilo en el exón 15 (51,52).

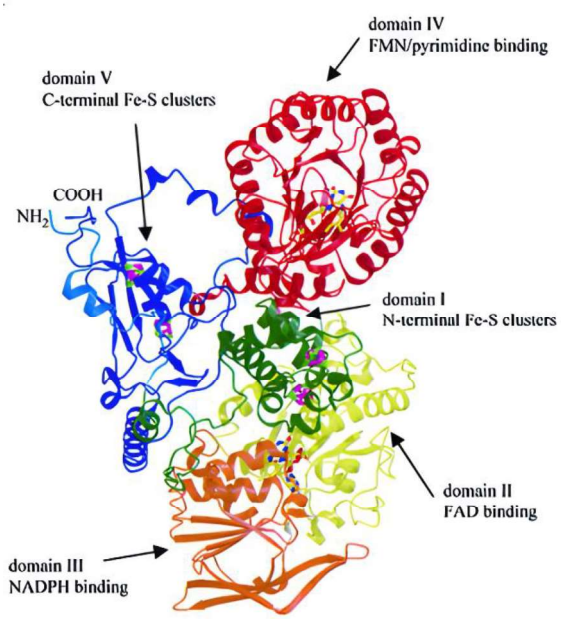


Figura 2. Tomado de: Dobritzsch. D C et al. (2001) Distribución de dominios y estructura de la proteína DPD del cerdo

A nivel del citosol, la dihidropirimidina deshidrogenasa cataliza la etapa limitante de la degradación de las pirimidinas; esta enzima se ha encontrado en una variedad de tejidos, sin embargo, la mayor actividad ha sido reportada a nivel hepático, donde se encarga de reducir las pirimidinas de uracilo y timina en una reacción dependiente de NADPH a las correspondientes 5,6-dihidropirimidinas. Durante el análisis por cristalografía, se ha evidenciado que DPD es un homodímero de 2×111 kDa donde cada subunidad de 1025 residuos de aminoácidos lleva un FAD, un FMN y cuatro grupos [4Fe-4S]. La enzima contiene sitios de unión separados para el cosustrato donante de electrones NADPH y las pirimidinas aceptoras de electrones respectivamente; el sitio de unión al sustrato es un bucle con una conformación abierta que permite estar expuesto al sustrato con un cambio conformacional cuando se une al sustrato llevando a que la enzima se pliegue y cierre el sitio activo (53,54).

En la proteína, el dominio I contiene un grupo Fe-S (residuos 27-173) que en la estructura secundaria tiene una conformación α-helicoidal con un patrón de plegamiento similar a la enzima fumarato reductasa de la *Escherichia Coli* (figura 2); en ambas enzimas se ha descrito la presencia de residuos de cisteína que participan en la coordinación de los grupos Fe-S durante el proceso de óxido reducción, así como se ha encontrado similitud frente a la presencia de los ligandos de cisteína y los residuos que forman la capa hidrófoba en los grupos metales y la actividad funcional proteica (54).

Los dominios II y III son los encargados de la unión a FAD y NADPH (residuos 173-286, 442-524) y III (residuos 287-441). Han mostrado similitud con las flavoproteínas disulfuro oxidorreductasa, con una hoja β paralela central rodeada por hélices α, permitiendo de esta forma la generación plegamientos tipo Rossman para la unión de nucleótidos (54) (figura 2). El dominio IV de unión a FMN presenta similitud estructural a la dihidroorotato deshidrogenasa de clase A de la *Lactococcus lactis*, con una conformación en forma de barril dependiente de hebras β y hélices α con múltiples enlaces peptídicos que permiten esta conformación (54). Finalmente, el dominio V contiene el segundo grupo de unión Fe-S, el cual comprende los residuos 1-23 y 848-1025. Este dominio cuenta con secuencias de cisteína para unión a los metales y pliegues tipo hélices α y un pliegue antiparalelo tipo hoja β (54).

A nivel estructural los 1025 residuos de aminoácidos del monómero de DPD se pliegan para formar los 5 dominios de la enzima, donde los primeros 4 dominios (residuos 27 al 847) forman un núcleo alargado y el dominio V se encuentra separado del resto de dominios formando bucles (figura 2) (54). Para el caso del dominio I, este juega un rol importante en la formación de dímeros y la conexión de los cinco dominios, donde el núcleo V se empaqueta más estrechamente con los dominios I y IV permitiendo la transferencia de electrones al momento de la unión de NADPH y pirimidina (54)

A nivel estructural se ha planteado que las variantes de interés en farmacogenética relacionadas con la disminución o perdida completa de la actividad de la proteína DPD podrían estar relacionadas con la formación de la estructura tridimensional de la proteína, donde de las variantes descritas por ejemplo las mutaciones R235W y V995F interfieren directamente con la unión del cofactor y/o del sustrato donde el grupo guanidinio de la

arginina 235 forma un puente de hidrógeno con el átomo de O4 de FAD, y la presencia de una cadena lateral de triptófano probablemente debilita o previene la unión del anillo de isoaloxazina del cofactor. Para el caso de V995F la inserción de la cadena lateral más grande de fenilalanina en la posición 995, ubicada en vecindad directa de C996, un ligando de uno de los grupos Fe-S, tiene el mismo efecto para la unión/ensamblaje del grupo (54).

3.4 Toxicidad relacionada con las fluoropirimidinas e impacto de la genotipificación del gen *DPYD*

El 5-FU y sus derivados usualmente son usados para el tratamiento de una gran variedad de tumores sólidos por lo cual sus perfiles de seguridad y eficacia son bien conocidos, disponiendo de una amplia descripción de los efectos secundarios comunes y las toxicidades graves, lo cual suele suceder cuando los pacientes están sobreexpuestos a los medicamentos a través de una disfunción metabólica o una sobredosis (55). Dentro de su uso habitual en cáncer colorrectal, gástrico, de mama, tejidos blandos de cabeza y cuello así como en piel se ha descrito que entre el 15 al 30% de los pacientes sufre de toxicidad severa, donde se incluye aparición de diarrea, náuseas, mucositis, estomatitis, mielosupresión y neurotoxicidad, siendo amplio el grado de compromiso y aparición de los síntomas (9). En cuanto a la administración de 5-FU y Capecitabina, el 20% del fármaco no es catabolizado por DPD y aproximadamente el 1% al 3% de la dosis administrada se convierte en metabolitos citotóxicos; el 80% del fármaco se cataboliza rápidamente por la dihidropirimidina deshidrogenasa. Se ha reportado que la toxicidad relacionada con el uso de fluoropirimidinas puede estar relacionada con la sobresaturación de DPD, que conduce a niveles tóxicos de metabolitos citotóxicos lo cual puede darse por aumento de la disponibilidad del medicamento o por defectos metabólicos en la enzima (55).

La pérdida de la función de la proteína DPD y su relación con la presencia de variantes catalogadas como nocivas en el gen *DPYD* ha sido ampliamente descrita, donde se ha reportado que pacientes con fenotipos deficientes de DPD tienen un mayor riesgo de toxicidad al 5-FU y sus análogos (12,17,56). Encontrando que los individuos con una pérdida de la actividad de la enzima, ante la exposición al 5-FU o sus análogos presentan toxicidad grado 3 y 4 generada por la acumulación de los metabolitos tóxicos derivados de la biotransformación del medicamento en el organismo (10,47,57).

Las principales variantes de importancia clínica que generan pérdida de la actividad enzimática de DPD son c.1905+1G>A (*DPYD**2A), c.1679T>G (*DPYD**13), c.2846A>T y c.1129-5923C>G. Estas son causantes de cambios en la uracilemia y dihidrouracilemia asociados a la pérdida de actividad enzimática con aumento de toxicidad en pacientes portadores durante el tratamiento con 5-FU o sus análogos, por lo cual la Sociedad Europea de Oncología Médica recomienda realizar un cribado genético de estas cuatro variantes antes de iniciar el tratamiento con fluoropirimidinas y según el genotipo ajustar la dosis inicial. Sin embargo, las directrices se basan principalmente en hallazgos de población caucásica (17,57,58); estimándose que la frecuencia global para cada una de esas 4 variantes varía entre 0,02% - 0,96%. Se ha encontrado una mayor presencia en la población caucásica europea con una frecuencia de hasta 4,8% para *DPYD**2A frente a otras poblaciones como la del sur de Asia, con una frecuencia esperada del 0,001%. Lo anterior limita la aplicabilidad de las recomendaciones por las bajas frecuencias de las variantes de interés y el desconocimiento frente a otras posibles variantes de mayor frecuencia (17,58).

Tabla 1. Tomado y adaptado de: Castro. C et al. (2014) Variantes del gen *DPYD* asociadas con mayor riesgo de toxicidad y efecto enzimático.

Variante	Efecto en la enzima
DPYD*2A (c.1905+1G>A)	Actividad nula
DPYD (c.464T>A)	Actividad disminuida
DPYD*3 (c.1897delC)	Actividad nula
DPYD*13 (c.1679T>G)	Actividad nula
DPYD*4 (c.1601G>A)	Actividad nula
DPYD (c.2846A>T)	Actividad disminuida
DPYD (c.1129-5923G>C)	Actividad disminuida

Se han documentado otras variantes en el gen *DPYD*, sin embargo, aún no se ha logrado evaluar la disminución de la actividad en DPD. Por ejemplo, para el caso de c.85T>C y c.496A>G, se sabe que conducen a cambios de aminoácidos en la proteína resultante mostrando una aparente alteración en la función proteica en estudios *in vitro*. Sin embargo, no se ha podido recrear nuevamente ese cambio en la actividad enzimática (59). Otras de las variantes con importancia clínica se describen en la tabla 1. A pesar de que muchas investigaciones buscan variantes implicadas en la pérdida de actividad también se ha

documentado otras variantes como c.1601G>A (*DPYD**4) y c.2194G>A (*DPYD**6) en las que la evidencia sugiere que a pesar del cambio, este no genera una alteración en la estructura proteica final, lo que conlleva a una actividad enzimática conservada (13).

Hasta la fecha se han descrito diversas variantes que se asocian con pérdida de la actividad enzimática y por tanto un riesgo aumentado de aparición de toxicidad en pacientes tratados con fluoropirimidinas. También se han encontrado otras variantes que se describen como “sin significado clínico”, sin embargo, éstas se han documentado en cohortes caucásicas en las cuales se desestima el impacto que estas puedan tener por su frecuencia alélica, siendo un ejemplo c.577A>G la cual en población afroamericana y afrocaribeña tiene una mayor frecuencia, así como asociación con toxicidad severa. Sin embargo, en población europea no se encontró reducción de la actividad en DPD (58), por lo cual es necesario establecer las variantes en el gen que generan el riesgo de la toxicidad para cada población.

En las guías de ajuste de dosis de fluoropirimidinas para pacientes portadores de variantes en el gen *DPYD* que supongan una disminución de la actividad de DPD y un aumento de riesgo de aparición de toxicidad, se recomienda genotipificar variantes que cuenten con suficiente evidencia del impacto funcional. Sin embargo, a pesar del reporte de múltiples variantes en la literatura, establecer de forma clínica la actividad enzimática disminuida es difícil dada la baja frecuencia y no siempre es posible evaluarlas *In Vitro* (13,60). El requerimiento de la evaluación funcional deriva de la asignación de una puntuación de la actividad enzimática según la evidencia disponible de estudios *In Vitro* como clínicos, con lo cual, según la variante, el médico clínico podrá definir si administra la dosis completa, una dosis parcial del 75 al 50%, o emplea otro agente quimioterapéutico según la variante presente en el paciente (60).

A pesar de las dificultades para la validación de variantes potencialmente accionables, se ha encontrado que la genotipificación previo al inicio del tratamiento con fluoropirimidinas limita la aparición de complicaciones derivadas del tratamiento quimioterapéutico, como lo evidenció el trabajo realizado por Jolivet et al., quienes reportaron que una genotipificación para *DPYD**2A previa al inicio del tratamiento en pacientes que se iban a exponer a las fluoropirimidinas, disminuyó la frecuencia de toxicidad severa además del deterioro del estado funcional así como los retrasos en el reinicio del tratamiento posterior al episodio la

toxicidad, reduciendo los costos globales derivados de la intervención médica durante el tratamiento (61). Esto ha llevado a evaluar el posible impacto positivo con la reducción del riesgo y del impacto económico de la evaluación de los niveles de uracilo o el estudio de la presencia de variantes accionables en el gen *DPYD* como herramienta de medicina personalizada y de precisión (62).

3.5 Herramientas de anotación de variantes de cambio de sentido

Posterior a la secuenciación completa del genoma humano, una de las principales dificultades relacionadas con la interpretación de la variación en el código genético ha sido la asignación del efecto de las variantes documentadas en el producto del gen y su relación con la aparición de enfermedad. De los tipos de variantes que se pueden presentar en el genoma humano, una de las más problemáticas son las variantes en el cambio de sentido, lo cual se deriva del problema de establecer el efecto en la sustitución de un aminoácido en la proteína (63,64). Para lo cual actualmente se disponen de diversas alternativas experimentales que permiten establecer el efecto de estos cambios, sin embargo, por los requerimientos al momento de la evaluación del efecto de una variante de cambio de sentido, tal vez las herramientas de mayor uso son los predictores *in silico*.

Estas herramientas se basan en modelos de clasificación que evalúan la similitud entre las características de los aminoácidos y la información filogenética disponible frente a la posición del cambio, siendo una limitante de la evaluación del efecto de la variante, la falencia en la inclusión durante la evaluación de la complejidad de las diversas estructuras proteicas, la función biológica de la biomolécula y los requerimientos físico-químicos necesarios para la interacción con otras moléculas (65). Sin embargo, a pesar de las limitantes de los algoritmos de predicción, la evaluación del efecto de las sustituciones de aminoácidos es un primer paso en la identificación y priorización de variantes con un potencial efecto nocivo y con una posible relación con el desarrollo de enfermedad, donde la limitación humana para el análisis de un gran volumen de información y un panorama en el que se estima la presencia de 24000 a 40000 sustituciones de aminoácidos en el genoma de un individuo, justifican el uso de estas herramientas en el ámbito clínico, siendo válida la priorización de variantes a partir de la predicción de su efecto (63).

Con el uso de la secuenciación como herramienta diagnóstica en el ámbito clínico en la búsqueda de causas genéticas de enfermedades mendelianas y la necesidad de establecer la causalidad de las variantes identificadas con el fenotipo presentado por el paciente de interés así como el efecto de las variantes de cambio de sentido y la sustitución de un aminoácido se han ido desarrollando diferentes algoritmos, lo cuales a partir de diferentes abordajes establecen el efecto de estos cambios en el genoma (64,66). Algunas de estas herramientas son descritas en la tabla 2.

Tabla 2. Descripción de algoritmos de anotación de variantes de cambio de sentido

Herramienta	Descripción
SIFT (Sorting Intolerant From Tolerant)	Utiliza la homología de secuencia para predecir si una sustitución de aminoácidos afectará la función de la proteína, para lo cual supone que los aminoácidos importantes se conservarán en la familia de proteínas. Para predecir si la sustitución de un aminoácido en una proteína afectará la función de la proteína, SIFT considera la posición en la que ocurrió el cambio y el tipo de cambio de aminoácido (66).
PolyPhen-2	PolyPhen-2 utiliza características predictivas basadas en secuencias y estructuras que se seleccionan a partir de un algoritmo *greedy*. Dentro de las características predictivas se encuentra la selección de homólogos y la alineación múltiple con inclusión de ortólogos y parálogos (64)
Mutation Taster	Esta herramienta emplea un clasificador de Bayes para predecir el potencial efecto de una variante para el desarrollo de enfermedad. El clasificador de Bayes predice las consecuencias funcionales no sólo de las sustituciones de aminoácidos sino también de las alteraciones intrónicas y sinónimas, las mutaciones cortas de inserción y/o deleción (indel) y las variantes que abarcan los límites intrón-exón (67).
Mutation Assessor	Se basa en la predicción del impacto funcional de las mutaciones de cambio de sentido usando la evaluación de la conservación evolutiva de los residuos de aminoácidos en una alineación de secuencias múltiples de una familia de proteínas, su enfoque radica en la explotación de la conservación evolutiva en subfamilias de proteínas, que están determinadas por la agrupación de múltiples alineamientos de secuencias homólogas en el contexto de la conservación de la función general (68,69).
FATHMM	Esta herramienta es capaz de predecir los efectos funcionales de las mutaciones de cambio de sentido combinando la conservación de secuencias dentro de modelos ocultos de Markov (HMM), que representan la alineación de secuencias homólogas y dominios proteicos conservados, con "pesos de patogenicidad", que representan la tolerancia general de la proteína/dominio a mutaciones (70).
DANN	DANN entrena una red neuronal profunda que consta de una capa de entrada, una capa de salida de función sigmoidea y tres capas ocultas de 1000 nodos con función de activación tangente hiperbólica. Esta herramienta usa los mismos datos de entrenamiento que CADD, difiriendo de esta última en el abordaje usado para la predicción (71).
CADD	Esta es una herramienta para calificar el carácter nocivo de variantes de un solo nucleótido, así como variantes de inserción/deleción en el genoma humano. El agotamiento dependiente de anotaciones combinadas (CADD) es un marco que integra múltiples anotaciones en una métrica contrastando variantes que sobrevivieron a la selección natural con mutaciones simuladas (72).
MetaLR	Esta herramienta utiliza una regresión logística para integrar 9 puntuaciones de nocividad (SIFT, PolyPhen-2, GERP++, MutationTaster, Mutation Assessor, FATHMM, LRT, SiPhy y PhyloP) e información de frecuencia alélica para predecir la nocividad de variantes de cambio de sentido (73).

GERP++	Esta herramienta usa una herramienta de alineación del genoma humano y otras 33 especies de mamíferos, identificando con alta confianza más de 1,3 millones de elementos restringidos que abarcan más del 7% del genoma humano, con lo cual se aplica un enfoque de programación dinámica para predecir globalmente un conjunto de elementos restringidos clasificados por sus valores p y una estimación de tasa de falsos positivos concomitante (74).
BayesDel	Esta herramienta combina múltiples predictores de nocividad para crear una puntuación general. Esta herramienta incluye los predictores PolyPhen2, SIFT, FATHMM, LRT, Mutation Taster, Mutation Assessor, PhyloP, GERP++ y SiPhy (75).
REVEL	Se basa en un entrenamiento Random Forest de un conjunto de variantes, donde se predice la patogenicidad de variantes de cambio de sentido basado en una combinación de puntuaciones de 13 herramientas individuales: MutPred, FATHMM v2.3, VEST 3.0, PolyPhen-2, SIFT, PROVEAN, MutationAssessor, MutationTaster, LRT, GERP++, SiPhy , phyloP y phastCons (76).
VARITY	Esta herramienta es un enfoque de aprendizaje automático supervisado para crear modelos predictivos especializados utilizando ejemplos de entrenamiento con pesos diferenciales optimizados (77).
PROVEAN	Esta herramienta recopila un conjunto de secuencias homólogas y lejanamente relacionadas de la base de datos de proteínas NCBI NR utilizando BLASTP , para cada secuencia en el conjunto de secuencias, se calcula una puntuación delta utilizando la matriz de sustitución (78).
Eigen	Es un algoritmo de aprendizaje no supervisado que combina una variedad de predictores para formar grupos de posiciones de nucleótidos funcionales y no funcionales en el genoma. Para el caso de Eigen-PC, este se basa en la descomposición propia de una matriz de covarianza de anotaciones y utiliza un vector principal para ponderar las anotaciones individuales (79).
SNPs y GO	La herramienta recopila información derivada de la secuencia de proteínas, el entorno de secuencia local de la SNV, el perfil de secuencia de proteínas, las características derivadas de la alineación de secuencias y la función de las proteínas, lo que agrega un grado de complejidad al análisis funcional de los SNP (5).
PhD-SNP	Se basa en una máquina de vectores de soporte, donde el predictor se basa en una única SVM entrenada y probada en secuencia de proteínas e información de perfil (5).
Meta-SNP	Es un clasificador binario aleatorio basado Random Forest para discriminar entre nsSNV polimórficos y relacionados con enfermedades. Integra cuatro métodos existentes: PANTHER, PhD-SNP, SIFT y SNAP (5).

3.6 Predicción *In silico* y variantes missense en farmacogenes

Con el aumento en la cantidad de datos disponibles a partir de la secuenciación masiva del ADN, se ha requerido el desarrollo de herramientas de interpretación de la variación en el genoma humano, donde las variantes de único nucleótido (SNV), particularmente las variantes tipo de cambio de sentido -las cuales generan variación de un solo aminoácido- son particularmente interesantes por las implicaciones estructurales, funcionales y de estabilidad proteica (4). Esto ha llevado a un aumento de los esfuerzos en la descripción y predicción frente a su efecto en los últimos años, principalmente en los genes relacionados con cáncer, donde los hallazgos derivados del efecto de estas variantes podrían ser útiles dentro del desarrollo de nuevos fármacos y la búsqueda de biomarcadores (4).

La principal dificultad con el análisis *In Vitro* o *In Vivo* del efecto de una variable biológica son los costos y el tiempo requerido (80). En contraposición, el análisis *In Silico* permite el análisis de grandes cantidades de datos biológicos de diversas fuentes, integrando diferentes variables y estableciendo interacciones entre los objetos de estudio a pesar de los costos y los requerimientos de talento humano y de equipamiento, adicionalmente, el modelaje computacional se ha convertido en una herramienta fundamental para la selección de variantes que requieran esfuerzos adicionales en el análisis del efecto a nivel biológico mediante alternativas de laboratorio. Por esta razón, es de gran interés no solamente desde la genómica sino desde otras áreas de la biomedicina, como por ejemplo en la farmacología donde este tipo de metodologías ha permitido el progreso en la predicción de la interacción entre fármacos, interacciones entre el medicamento y la diana terapéutica, así como el descubrimiento de nuevas moléculas (81,82). Para el caso del análisis de variantes de cambio de sentido durante el análisis clínico de datos derivados de la secuenciación del genoma, se ha planteado la combinación de los datos disponibles frente a análisis funcionales *in vitro* o *in vivo* con los datos aportados por análisis *In Silico* al momento de establecer la implicación funcional de una SNV. Sin embargo, pocas veces se dispone de información relevante que permita establecer la relación causal de la variante con la enfermedad y los algoritmos de análisis *In Silico* no suelen ser específicos para una condición, siendo una limitante para su uso de forma determinística al momento de establecer relación de causalidad (81,83).

El análisis *In Silico* inicial del efecto de variantes de cambio de sentido se basaba en el análisis bioquímico teniendo en cuenta las bases teóricas de la conformación estructural de las proteínas, generando matrices con información fisicoquímica y del impacto de los cambios estructurales, que no necesariamente aportaban una evidencia sólida para una interpretación clínica. Sin embargo, en los últimos años se han desarrollado algoritmos basados en la comparación del efecto de la proteína en múltiples especies evaluando la aptitud evolutiva y pudiendo ser comparables con ensayos funcionales y el análisis integrado de múltiples parámetros para clasificar las sustituciones de cambio de sentido. Ajustados a las particularidades de cada condición y gen podrían constituirse como una herramienta fundamental de la valoración del impacto funcional de una variante (84).

Para el caso de los farmacogenes, a pesar de la abundancia de datos genómicos, la disponibilidad de variantes accionables en estos genes es limitada, lo cual es derivado de la necesidad de la caracterización basada en datos clínicos o experimentales para su uso en la práctica médica, planteando la necesidad del uso de algoritmos de predicción de forma generalizada que permitan establecer la probabilidad con precisión del efecto funcional de una variante, las características estructurales derivadas de la presencia de la misma así como las anotaciones frente a restricciones evolutivas, lo cual evidencia la necesidad de la valoración predictiva de cada uno de estos algoritmos de forma individualizada para cada farmacogen (80,85).

En los últimos años se ha desarrollado diversas herramientas predictivas que utilizan diversos conjuntos de datos para establecer el efecto de variantes de cambio de sentido, donde por ejemplo SIFT usa información de conservación, PolyPhen2 usa propiedades fisicoquímicas y análisis de estructura cristalina y MutPred integra una lista completa de cambios derivados de secuencia proteica, sin embargo, esto se enfoca en identificar efectos de variantes y su relación con la aparición de enfermedad, en contraposición al enfoque tradicional de estudio de efecto de variantes en farmacogenes donde se evalúa concentraciones de fármacos y parámetros biológicos relacionados con la proteína de interés (80). La preocupación en el desarrollo de herramientas de predicción de variantes en genes de interés en farmacología no ha sido tan grande como en el caso de evaluación de la patogenicidad en condiciones monogénicas, pero se ha planteado el uso de métricas de conservación desde la evaluación unidimensional y el análisis estructural con evaluación de interacciones para el caso de dinámica molecular, pero sigue siendo una limitante la variabilidad del efecto de estas variantes entre la población (86).

Una de las dificultades de la implementación generalizada de herramientas de anotación para la identificación de variantes de interés en farmacogenes, son los patrones de conservación evolutiva de estas y el efecto en la secuencia proteica, donde a diferencia de las que están relacionados con enfermedad, para el caso de los farmacogenes, la variación de relevancia no genera un defecto severo en la funcionalidad de la proteína que establezca un cambio frente a la conservación de la función biológica de la biomolécula (80). Se ha sugerido que una de las dificultades durante la evaluación de variantes en farmacogenes por algoritmos de anotación individuales, es una posible pérdida de la capacidad predictiva derivada de limitaciones de la herramienta en la correlación entre las

características propias de las enzimas implicadas en procesos metabólicos y la presencia de una variante en el gen, lo cual se puede explicar por el hecho de ser proteínas derivadas de genes menos conservados y un efecto nocivo sutil de las variantes en estudio, donde no se afecta la función biológica de forma severa, sino que la alteración se evidencia durante la exposición a un medicamento, alterando la farmacodinamia y farmacocinética esperada, sin embargo, la combinación de diversos anotadores podría aumentar los umbrales de predicción, lo cual puede ser prometedor en la identificación de pacientes de riesgo (87).

Actualmente el enfoque de identificación de variantes de riesgo en farmacogenética se basa en grandes estudios de asociación donde se busca evaluar la causalidad entre la presencia de una variante y una disminución de la actividad enzimática de la proteína relacionada, lo cual es factible a nivel de investigación pero ya en la práctica clínica generalizada es inviable debido a la gran cantidad de individuos que sería necesario analizar, los costos del proceso experimental y el volumen de datos necesarios para cada caso, por lo cual, la evaluación de herramientas computacionales constituye una estrategia que impactaría de forma positiva en el análisis de estas variantes, al permitir una agrupación inicial de las variantes y la priorización de los análisis de laboratorio de acuerdo con la predicción del efecto (88). Sin embargo, para su aplicación se debe realizar una evaluación de las diversas herramientas disponibles, dado que como primera limitante para su aplicación, es la forma en que están diseñadas para la identificación de efecto nocivo o neutro, donde el diseño del algoritmo puede incluir las variantes en farmacogenes en el grupo de variantes con un efecto neutro por la frecuencia poblacional y por el grupo de entrenamiento utilizado para estos algoritmos, pero es de anotar, que las herramientas que realizan una evaluación de conservación en farmacogenes asociados a enzimas podrían predecir un efecto funcional alterado (88).

Los enfoques computacionales para determinar el efecto las variantes tipo SNV proporcionan diferentes herramientas con los cuales se lograr definir los efectos patogénicos y determinar sus mecanismos moleculares subyacentes, lo cual es fundamental para la comprensión de las variantes de riesgo en los genes de interés en farmacología, donde en algunos casos con los hallazgos se logra comprender la estructura tridimensional a partir de simulación de dinámica molecular informando acerca del impacto de las mutaciones en el funcionamiento de la proteína (89). Donde un ejemplo del análisis

realizado para este tipo de genes son los resultados del análisis de variantes missense en el gen *CYP4F2* para el cual se han logrado hallazgos similares a lo reportado en estudios *in vitro* (89).

En otras aproximaciones reportadas en la literatura para el análisis computacional en variantes de interés en farmacología, se ha planteado como alternativa para la generación de datos derivado de las limitaciones por la privación de datos poblacionales y ausencia de estudios funcionales, el análisis de variantes mediante el genotipado de pacientes y el posterior análisis computacional mediante simulación unidimensional y estructural del producto proteico resultante de la presencia de la variante de interés, lo cual se ha realizado para el gen *CYP2C19*, logrando ampliar la información disponible acerca de las variantes de riesgo para este citocromo, el cual es relevante para el metabolismo de diversos fármacos (90). Para otras proteínas de la familia de los citocromos, este tipo de análisis no solamente ha permitido definir el impacto de diversos tipos de variantes en la función de las proteínas, sino que adicionalmente, ha permitido establecer posibles mecanismos asociados en diversos mecanismos de patogenicidad en diferentes trastornos (90).

Para el caso del gen *DPYD*, la mayoría de los estudios se basan en la identificación de nuevas variantes de riesgo a partir del análisis clínico y experimental, lo cual ha limitado el avance en la aplicación de la genotipificación de forma generalizada y limita el uso de la secuenciación de todo el gen para su análisis. Algunas investigaciones como la realizada por Sherestha et al. en 2018, han planteado el desarrollo de modelos de clasificación de variantes de cambio de sentido basado en aprendizaje automático, donde su modelo mostró una precisión del 85% en la identificación de variantes nocivas (91). Lo anterior podría ser el primer paso en establecer modelos mixtos de tamizaje, donde con la identificación de variantes potencialmente nocivas durante la secuenciación masiva, se puedan identificar pacientes de riesgo que puedan ser llevados a fenotipificación. Esto podría llevar al aumento de la disponibilidad de datos frente al efecto de las variantes de cambio de sentido y disminuir el riesgo de toxicidad al usar estos medicamentos, y constituye una oportunidad de abrir un espacio para la aplicación de medicina de precisión basada en uso de modelos multiparamétricos, siendo estos muchos más costo-efectivos como ya ha sido reportado para el tamizaje en *DPYD* (92).

Objetivos

4.1 Objetivo general

Determinar la capacidad predictiva de diferentes algoritmos *In Silico* de anotación de variantes de cambio de sentido en el gen *DPYD*, con el propósito de proponer un protocolo de evaluación *In Silico* basado en herramientas de anotación con alta sensibilidad y especificidad en la predicción de variantes de interés en farmacogenética en población colombiana.

4.2 Objetivos específicos

1. Establecer la validez, sensibilidad y especificidad de los algoritmos *In Silico* disponibles para la predicción del efecto de variantes genéticas de cambio de sentido con aplicación en genes de interés en farmacogenética a partir de una revisión sistemática de la literatura.
2. Evaluar la capacidad predictiva de distintos algoritmos *in silico* de anotación para detectar el efecto de variantes de cambio de sentido en el gen *DPYD*, utilizando variantes previamente caracterizadas con evidencia funcional y cuyo impacto en DPD es conocido, proponiendo un protocolo de evaluación de variantes de cambio de sentido en este gen de interés en farmacogenética.
3. Aplicar el protocolo propuesto en un conjunto de datos de muestras de secuenciación de exoma completo de un laboratorio de biología molecular de la ciudad de Bogotá para identificar variantes no reportadas previamente y establecer la frecuencia de variantes reconocidas como nocivas y reportadas en bases de datos de farmacogenética.

Metodología

5.1 Tipo de estudio

Se realizó un estudio observacional descriptivo farmacogenético, donde se evaluó la precisión en la predicción del efecto de variantes de cambio de sentido en el gen *DPYD* de diferentes algoritmos de anotación *in silico*, así como el análisis descriptivo del efecto estructural en la proteína DPD de estas variantes, para lo cual se usaron herramientas de modelado *in silico* de acceso libre. Adicionalmente se realizó la identificación de variantes conocidas como factor de riesgo en el gen *DPYD* para la aparición de toxicidad relacionada con el uso de fluoropirimidinas y se identificaron de variantes no reportadas previamente en un banco de datos de la unidad de analítica de datos de Biotecgen S.A.S.

5.2 Elección de algoritmos de anotación

Se realizó una revisión sistemática de tipo revisión rápida descriptiva de acuerdo a las recomendaciones de la guía práctica emitida por la Organización mundial de la Salud (OMS) (93) y las recomendaciones emitidas por el protocolo PRISMA para revisiones rápidas (94). La formulación de la pregunta de la revisión sistemática, así como la selección de los términos MeSH utilizados durante la búsqueda en las bases de datos se realizó bajo la metodología PICO ajustada para pruebas diagnósticas recomendada por PRISMA-DATA Statement, los términos MeSH así como la pregunta de investigación se enuncia en la tabla 3.

Tabla 3. Estrategia PICO y términos MeSH usados para la búsqueda en bases de datos

¿Cuáles son los algoritmos predictivos que presentan la mayor precisión en identificar variantes missense nocivas?				
Lenguaje natural	Acrónimos			
	Paciente	Prueba índice	Comparación	Resultado
	Missense Mutation	Algorithms	No se sugieren términos adicionales para esta sección	Predictive Value of Tests
MeSH	Amino Acid Substitution	Computer Simulation		Sensitivity and Specificity
		Computing Methodologies		
		Molecular models		

Inicialmente se planteó la búsqueda con restricción para artículos originales en los que se incluyeran análisis de genes de interés en farmacología, sin embargo, el número de artículos disponibles en las diferentes bases de datos fueron limitados por lo cual se amplió la búsqueda con la inclusión de artículos en los que se analizaran otros genes. Se realizó la búsqueda y recolección de artículos de interés en 2 fases: la primera con la inclusión de las palabras “Pharmacogenetics” y “Pharmacogenomics” asociadas al resto de términos MeSH dentro de una búsqueda específica y posteriormente una búsqueda general sin incluir estos términos. Para la traducción al español de los términos MeSH se usó el tesauro multilingüe DeCS/MeSH.

Se planteó como formula de búsqueda ('Computing Methodologies' OR 'algorithm' OR 'computer simulation' OR 'molecular simulation') AND ('missense mutation' OR 'amino acid substitution') AND ('sensitivity and specificity' OR 'predictive value') para la búsqueda general y en la búsqueda específica ('Computing Methodologies' OR 'algorithm' OR 'computer simulation' OR 'molecular simulation') AND ('missense mutation' OR 'amino acid substitution') AND ('sensitivity and specificity' OR 'predictive value') AND ('pharmacogenetics' OR 'pharmacogenomics'). En cada base de datos se ajustaron los patrones de búsqueda de acuerdo con las características propias de cada buscador. Se realizó la búsqueda en las bases de datos de PubMed, Scielo, Cochrane, Scopus y Embase ajustando la búsqueda a artículos publicados desde el 2013 al 2023, en idioma inglés y español.

Se realizó un filtrado inicial de los artículos encontrados mediante la evaluación de sus títulos y resúmenes. Para la selección de los artículos de interés, se consideró que el título debía definir claramente la población de estudio, mencionar la comparación de diferentes herramientas de predicción de variantes de cambio de sentido como tema central, y especificar el tipo de diseño experimental. En el análisis de los resúmenes, se excluyeron los artículos que no reflejaran claramente el enfoque metodológico de la investigación, priorizando investigaciones donde el diseño experimental buscara la evaluación comparativa de algoritmos de predicción de variantes de cambio de sentido, que no respondieran a la pregunta de investigación planteada bajo la metodología PICO, o que no proporcionaran suficiente información para evaluar la pertinencia del estudio para ser incluido en la revisión sistemática. Posteriormente, los artículos fueron filtrados de acuerdo con los siguientes criterios de inclusión y exclusión:

Criterios de inclusión:

- Estudios primarios que investigan el uso de algoritmos de anotación *In Silico* para identificar variantes missense nocivas en genes de interés biológico o clínico.
- Estudios que utilizaron muestras humanas, como datos de secuenciación genómica, cohortes clínicas o bancos de datos genéticos bien caracterizados.
- Estudios que informen métricas de evaluación cuantitativa, como sensibilidad, especificidad, valor predictivo positivo o negativo, área bajo la curva de característica operativa del receptor (ROC) u otras medidas de rendimiento.

Criterios de exclusión:

- Revisiones sistemáticas, metaanálisis, opiniones, comentarios y cartas al editor.
- Estudios realizados con algoritmos de anotación usados para evaluar variantes somáticas en cáncer.
- Estudios que no se centran en el uso de algoritmos de anotación *In Silico*.
- Estudios con datos incompletos o falta de información detallada, como deficiencia en la caracterización sobre los algoritmos utilizados, sus características o que no informen sobre su precisión y rendimiento.
- Estudios duplicados o que comparten el mismo conjunto de datos y resultados.

Finalmente, se realizó la evaluación de la calidad de la información de los artículos incluidos con el instrumento QUADAS-2 (95), las preguntas del instrumento fueron ajustadas de acuerdo a las recomendaciones emitidas por los autores de la herramienta de acuerdo a las necesidad de la revisión realizada, el instrumento se puede consultar en el anexo 1. Para la síntesis de resultados se realizó de forma narrativa donde se presentaron los principales hallazgos para cada investigación, adicionalmente se recopilaron las diferentes métricas de sensibilidad, especificidad y área bajo la curva para los algoritmos evaluados en cada artículo. No se realizó el análisis estadístico de los datos obtenidos por la variación frente a los genes estudiados y las métricas tenidas en cuenta en cada ensayo, por lo cual se consideró que generar un modelo de resumen estaría sesgado.

5.3 Elección de variantes de interés en el gen *DPYD*

Se realizó la búsqueda de variantes en el gen *DPYD* en la base de datos de acceso libre del Pharmacogene variation Consortium (PharmVar), donde se filtraron por tipo de variante y efecto esperado en el producto proteico, seleccionado las variantes de cambio de sentido y las que se encontraban anotadas con “decreased function”, “severely decreased”, “no function” y “normal function”. Adicionalmente se incluyeron las variantes reportadas en el estudio realizado por Shrestha. et al., 2018. Se construyó una tabla en la cual se incluyó el ID de cada variante, el dbSNP, posición genómica en genoma de referencia GRCh38 y GRCh37, cambio en el genoma y cambio a nivel de la proteína.

Para cada variante seleccionada se evaluó la correcta anotación de acuerdo con la nomenclatura recomendada para el reporte de variantes. Se tomó como prueba Gold Standard del efecto de las variantes en el gen *DPYD,* los estudios funcionales experimentales *in vitro* disponibles para cada caso; para establecer la validez de los estudios funcionales se siguieron las recomendaciones emitidas por el Clinical Genome Resource (ClinGen) (96) frente a la evaluación de estudios funcionales y su uso como evidencia de funcionalidad. Para cada estudio se aplicó el flujo de evaluación recomendado por la guía donde se identificaron: la evaluación experimental con uso de controles positivos y negativos, el número de réplicas realizadas para cada experimento, la evaluación de la actividad enzimática de DPD y su comportamiento frente al 5-FU o alguno de sus análogos. De acuerdo con el desarrollo experimental para cada estudio analizado, se asignó BS3/PM3 y el nivel de evidencia para designar los estudios que cumplieron los criterios de validez experimental recomendadas por el Clinical Genome Resource (96); los estudios que no cumplieron dichos criterios no se les asigno esta marcación y fueron excluidos del análisis. Las variantes en las que no se identificaron estudios con suficiente validez frente a su efecto, fueron excluidas del estudio. La asignación de BS3/PM3 se realiza como método para la designación de la validez experimental y asignación de la fuerza de evidencia, mas no como designación de patogenicidad de acuerdo con las recomendaciones emitidas por el American College of Medical Genetics en su guía para interpretación de variantes.

Para disminuir el sesgo derivado de la selección de la prueba de referencia solo se tuvieron en cuenta los estudios *in vitro* al momento de asignar el efecto de una variante en el

producto del gen, las variantes que disponían solamente de estudios clínicos o *in vivo* fueron excluidas del estudio. En la tabla construida adicionalmente se agregó el nivel de funcionalidad de acuerdo con el estudio *in vitro*, la clasificación reportada en PharmVar y la reportada en los estudios funcionales y la fuerza de la evidencia según las recomendaciones de ClinGen.

5.4 Generación de predicción y puntajes de las variantes por los algoritmos de anotación

Se descargaron de la base de datos de PharmVar los VCF disponibles para las variantes reportadas en el gen *DPYD*, se realizó la selección de los archivos de las 137 variantes incluidas en el estudio. Con la herramienta bcftools v1.10.2 (https://github.com/samtools/bcftools) (97) se generó un solo archivo VCF con las 137 variantes seleccionadas, donde se incluyó la posición genómica de acuerdo al genoma de referencia GRCh37 y GRCh38, nucleótido de referencia y alterno así como cambio en la proteína. Con el archivo VCF resultantes se usó la herramienta dbNSFP v4 (98) para obtener los puntajes de los algoritmos de predicción de efectos deletéreos Mutation Assesor, MutPred, Revel, Povean, BayesDel, SIFT y Eigen. Para el caso de Phd-SNP, SNPs&GO y Meta-SNP se usó la plataforma online disponible de Biomolecules Folding and Disease (https://snps.biofold.org/meta-snp/ y https://snps.biofold.org/snps-and-go/). Para las herramientas evaluadas online se usó la secuencia proteica disponible en UniProt Q12882, la cual corresponde a la proteína DPD humana.

Con todos los resultados de predicción se construyó una tabla en la cual se incluyó el ID de cada variante, la posición genómica para el genoma de referencia GRCh37 y GRCh38, nucleótido de referencia y alterno, el cambio en la secuencia de la proteína, el efecto funcional de acuerdo con lo documentado durante la elección de las variantes y los estudios disponibles funcionales y los resultados de la predicción para cada algoritmo de predicción de efecto deletéreo (puntaje obtenido e interpretación).

Para las herramientas evaluadas se usó como punto de corte los sugeridos por cada desarrollador, exceptuando la herramienta Revel, para la cual se tuvo en cuenta los puntos de corte sugeridos por el Clinical Genome Resource para su uso a nivel clínico.

Tabla 4. Umbrales usados para la anotación de las variantes analizadas

PREDICTOR	INTERPRETACIÓN
BayesDel	Las puntuaciones que genera el predictor oscilan entre -1,1 y 0,9. Se considera como umbral para predicción de un efecto nocivo un puntaje >0 (99).
Eigen	Las puntuaciones resultantes de Eigen y Eigen-PC varían de -3 a 2. Se considera como umbral para predicción de un efecto nocivo un puntaje >0 (79).
SNPs y GO	Para este algoritmo se relaciona la probabilidad de la relación de la variante con la enfermedad, se considera que si la probabilidad es mayor que 0,5 se considera como nociva y si es menor que 0,5, se predice como neutral (5).
PhD-SNP	El valor del índice de confiabilidad (RI) se evalúa a partir de la salida del SVMSVM (O) como RI = 20∗abs (O-0,5) (5).
SIFT	Las puntuaciones varían de 0 a 1, los valores cercanos a 0 indican una mayor fuerza frente a la predicción de un efecto dañino. Se considera como umbral para predicción de un efecto nocivo un puntaje <0,05 (99).
Meta-SNP	El valor del índice de confiabilidad (RI) se calcula a partir del resultado de Meta-SNP como: RI = 20∗abs (O-0,5) (5).
Mutation Assesor	Los umbrales establecidos para este predictor son: ≤0,8 = neutral, 0,8≤1.9 = low, 1,9≤3,5 = médium y >3,5 = high. Los desarrolladores sugieren que neutral y low se tomen como variantes toleradas (100)
REVEL	Para el caso de Revel el rango de umbral estimado para considerar una variante nociva es >0,644 (101).
PROVEAN	Para el caso de este predictor se tomaron como umbrales los informados por los desarrolladores,

> con un umbral para la clasificación de una variante ≥- 2.5 nociva y ≤- 2.5 neutra (100).

Se consideró para el análisis estadístico “1” como deletéreo y “0” como tolerado, para el caso de las variantes que fueran clasificadas como de significado incierto o el predictor indicara que no estaba disponible para esa variante, se asignó un valor de ”3”, estas variantes no fueron tenidas en cuenta al momento del análisis estadístico para el cálculo de las métricas de rendimiento de los algoritmos de anotación.

5.5 Cálculo de métricas de rendimiento

A partir de la tabla construida con la información acerca del efecto funcional y el análisis *in silico* para cada variante, se realizó el análisis de los datos en Google Colaboratory,v 0.0.1a2 donde se construyó a partir del lenguaje de programación Python el código para realizar el cálculo de las métricas de rendimiento y la generación de las curvas ROC. Para el cálculo de la sensibilidad (Sen), especificidad (Esp), valor predictivo positivo (VPP), valor predictivo negativo (VPN) y la exactitud (Acc) se usó como input la predicción del efecto por parte de las pruebas funcionales (tolerado o deletéreo) y el output emitido por cada algoritmo de anotación frente a la predicción del efecto (tolerado = 0 o deletéreo = 1); para los casos en los que el resultado emitido por la herramienta en evaluación no fuera en sistema binario, se ajustó el resultado cambiando los términos usados para el reporte del efecto para cada predictor, es decir, para los casos en los cuales solo se emitiera un valor numérico o una clasificación alfanumérica no binaria (neutral, low, médium, high), de acuerdo con los umbrales emitidos por los autores de cada herramienta se asignó “0” o “1” de acuerdo a la predicción, esto para poder generar uniformidad en los datos al momento del análisis estadístico. Con los datos uniformes se generaron las tablas de contingencia para cada herramienta de acuerdo como se evidencia en la figura 3, para lo cual se tuvo en cuenta (102):

- **Verdadero positivo (VP)**: Denota el número de muestras positivas clasificadas correctamente.
- **Verdadero negativo (VN)**: Denota el número de muestras negativas clasificadas correctamente.

- **Falso positivo (FP)**: Denota el número de muestras clasificadas incorrectamente como positivas.
- **Falso negativo (FN)**: Denota el número de muestras clasificadas incorrectamente como negativas.

	Prueba de referencia		
Prueba índice	Nociva	Tolerada	Total
Positiva	VP	FN	VP+FN
Negativa	FP	VN	FP+VN
Total	VP+FN	FN+VN	**N**

Figura 3. Análisis matemático de acuerdo con las tablas de contingencia

Para calcular las métricas de rendimiento se tuvo en cuenta las siguientes definiciones:

Exactitud (Accuracy): Es la relación entre las muestras clasificadas correctamente y el número total de muestras en el conjunto de datos de evaluación. Este valor está limitado a [0, 1], donde 1 representa predecir correctamente todas las muestras positivas y negativas y 0 representa no predecir correctamente ninguna de las muestras positivas o negativas (102).

$$ACC = \frac{TP + TN}{TP + FP + TN + FN}$$

Sensibilidad (Recall): Denota la tasa de muestras positivas clasificadas correctamente y se calcula como la relación entre las muestras positivas clasificadas correctamente y todas las muestras asignadas a la clase positiva. La recuperación está limitada a [0, 1], donde 1 representa una predicción perfecta de la clase positiva y 0 representa una predicción incorrecta de todas las muestras de clase positiva (102).

$$Rec = \frac{TP}{TP + FN}$$

Especificidad (Specificity): Denota la tasa de muestras negativas clasificadas correctamente. Se calcula como la relación entre las muestras negativas clasificadas

correctamente y todas las muestras clasificadas como negativas. La especificidad está limitada a [0, 1], donde 1 representa una predicción perfecta de la clase negativa y 0 representa una predicción incorrecta de todas las muestras de clase negativa (102).

$$Spec = \frac{TN}{TN + FP}$$

Valor predictivo positivo (VPP): Es la relación entre las muestras positivas clasificadas correctamente y todas las muestras clasificadas como positivas (102).

$$VPP = \frac{TP}{TP + FP}$$

Valor predictivo negativo: Es la relación entre las muestras negativas clasificadas correctamente y todas las muestras clasificadas como negativas.

$$VPN = \frac{TN}{TN + FN}$$

Curva ROC y área bajo la curva (AUC): La curva ROC es una gráfica en la cual se visualiza, organiza y seleccionan clasificadores en función de su rendimiento. Los gráficos ROC representan la compensación entre las tasas de aciertos y las tasas de falsas alarmas de los clasificadores (103), para los casos de las pruebas diagnósticas determinando la exactitud diagnostica de un test y su capacidad de diferencia entre sujetos sanos vs enfermos estimando la AUC, la cual refleja esa capacidad del test para discriminar los pacientes entre sanos y enfermos a lo largo de todo el rango de puntos de corte para la prueba (104). Para la generación de las gráficas para cada herramienta se usó la librería de Scikit-Learn v1.2.2 disponible para la generación de curvas ROC (105)

5.6 Análisis descriptivo estructural

Se evaluó el efecto estructural de cada variante seleccionada, para lo cual se usó la herramienta Online HOPE (https://www3.cmbi.umcn.nl/hope/about/) (106) y los datos disponibles en Missense 3D-DB (http://missense3d.bc.ic.ac.uk:8080/#referencespanel)

(107). Para el análisis con HOPE se usó la secuencia proteica disponible en UniProt Q12882, la cual corresponde a la proteína DPD humana. Se realizó un análisis descriptivo comparativo entre los cambios reportados a nivel de aminoácido (carga y tamaño), cambios en los contactos del residuo con estructuras cercanas, cambios en la estructura y predicción de alteración estructural entre las variantes conocidas como nocivas con las variantes toleradas.

5.7 Análisis poblacional

Se generó un listado de los VCF con las muestras disponibles en el banco de datos de la unidad de analítica de datos de Biotecgen S.A.S. Para la inclusión de las muestras se tuvo en cuenta:

Criterios de inclusión:

- Datos derivados de secuenciación de exoma completo.
- Consentimiento informado de Biotecgen S.A.S en el cual el paciente aceptó el uso de los datos derivados de la secuenciación para investigación.

Criterios de exclusión:

- No aceptación del uso los datos derivados de la secuenciación para investigación por parte del paciente.
- Datos derivados de secuenciación de genomas, paneles y genes únicos, para los dos últimos, solamente para los casos en los que la secuenciación se hubiera dirigido a genes específicos y no todo el exoma.
- Datos derivados de la secuenciación de muestras tumorales.

El listado de muestras se generó a partir de la línea de comandos en Bash con la ayuda del lenguaje AWK. Se produjo un documento en formato .txt que contenía la lista de muestras. Estas muestras se compararon con la base de datos disponible en la unidad de analítica de datos de Biotecgen S.A.S con la información de consentimiento informado y tipo de análisis realizado (secuenciación de genoma, exoma completo o panel de genes). No se excluyeron muestras por el origen de la muestra (somática o germinal), ya que la carpeta donde se encontraban almacenados los datos seleccionados no incluía datos derivados de la secuenciación de muestras tumorales o para estudios a nivel somático. Se

eliminaron las muestras en las que los pacientes no aceptaron el uso los datos derivados de la secuenciación para investigación, así como en las que no se realizó secuenciación completa del exoma. No se incluyeron el análisis de genoma, ya que en estos datos las profundidades esperadas podrían ser inferiores a las recomendadas para la identificación de variantes de interés en farmacología.

Luego, se utilizó la herramienta "akt - ancestry and kinship toolkit v0.3.9" (108) para generar una lista de muestras duplicadas y para identificar el parentesco entre las muestras. En el caso de muestras duplicadas, se eliminó una de las copias al azar. Para las muestras con parentesco, se tomaron decisiones basadas en si se trataba de dúos, tríos o familiares, y se seleccionó al azar una sola muestra para cada caso, conservando solo esa muestra en el análisis. Posterior a la eliminación de estas muestras se seleccionaron 1000 muestras para el análisis poblacional, la selección del n poblacional fue a conveniencia.

Con las 1000 muestras seleccionadas, se realizó la compilación en un solo VCF de todas las muestras en la región del gen *DPYD* (GRCh37:Chr1:97543100-98386715), para la identificación de variantes de interés se incluyeron las variantes usadas para el análisis de rendimiento de los algoritmos de anotación y el listado de variantes reportadas en PharmVar, por lo cual se realizó la búsqueda de las variantes usando la herramienta bcftools view (97) donde se comparó el VCF con los datos compilados y el VCF disponible en PharmVar con el listado de variantes reportadas para el gen *DPYD* (https://www.pharmvar.org/gene/DPYD) en el que están incluidas la variantes usadas para el análisis comparativo de los algoritmos de anotación; se filtraron las variantes de acuerdo a una profundidad >50 y calidad >20 según las recomendaciones para la evaluación de variantes en farmacogenes (109,110). Con el VCF resultante se usó la herramienta NGSEPcore v4.1 con su utilidad VCFDiversityStats (111) para el cálculo de las métricas poblacionales descriptivas (frecuencias alélicas y equilibrio de Hardy–Weinberg).

5.8 Identificación de nuevas variantes

Para la identificación de variantes no reportadas, se usó el VCF con la compilación de todas las muestras, se realizó el filtrado de las variantes con la herramienta bcftools view (97), se ajustaron los filtros para que al momento de la selección de las variantes se excluyeran las variantes presentes en el VCF descargado de Pharmvar o en PharmGKB.

Para la búsqueda de variantes nuevas, se permitió una mayor laxitud frente a los filtros aplicados para la inclusión de variantes, por lo cual solo se evaluó el estado de validación por parte del algoritmo de llamado de variantes, incluyendo variantes clasificadas como "PASS", no se incluyó la profundidad de acuerdo con las recomendaciones para la identificación de variantes en genes de interés en farmacología (109,110), por el carácter exploratorio de la búsqueda. Para obtener los parámetros de calidad de las variantes obtenidas, por medio de comandos AWK se obtuvieron las muestras en las cuales estaban presentes las variantes y usando bcftools v1.10.2 se obtuvieron las métricas de calidad de los respectivos VCF.

Considerando los hallazgos realizados durante la evaluación comparativa de las variantes, se consideró a PROVEAN como la principal herramienta para la identificación de variantes nocivas, por lo cual se planteó como única herramienta de predicción durante la construcción del protocolo de evaluación de las variantes en el gen *DPYD*. Sin embargo, para el análisis de nuevas variantes, al no disponer de un análisis fenotípico de las muestras, se realizó el análisis de las muestras seleccionadas en dos pasos.

En el primer paso se realizó la evaluación de las variantes por un grupo de selección, donde se incluyeron las herramientas BayesDel AddAF, Eigen y PROVEAN; se tomó esta decisión con alternativa para apoyar la predicción del efecto de las variantes evaluadas, considerando variantes potencialmente nocivas aquellas en las que al menos 2 herramientas de anotación (PROVEAN y otra herramienta) indicaran un efecto “nocivo”. Para el segundo paso (clasificación) se incluyeron las herramientas PROVEAN y MetaSNP, para este paso se consideró a MetaSNP como herramienta de soporte, por lo cual para las variantes en las que ambas herramientas indicaran un efecto nocivo, se consideró como altamente probable que la variante tuviera un efecto nocivo sobre el producto del gen.

Las variantes seleccionadas como nocivas fueron evaluadas a nivel estructural con HOPE y missense 3D y se realizó el cálculo de métricas descriptivas poblacionales con NGSEPcore (111).

Consideraciones éticas

De acuerdo con los principios establecidos en la declaración de Helsinki, la resolución número 2378 de 2008 y la resolución número 8430 de 1993 se consideró una investigación sin riesgo de acuerdo con los siguientes criterios:

- No se realizó ninguna intervención directa en seres humanos que implicara la administración de medicamentos, la toma de muestras o el seguimiento de pacientes, la presente investigación se basó en el análisis de datos biológicos disponibles en la literatura y resguardados en el banco de datos de la unidad de analítica de datos de Biotecgen S.A.S.
- El análisis para identificar las herramientas con mayor precisión durante la predicción de variantes missense en el gen *DPYD* se realizó con herramientas y datos libres disponibles en diversas bases de datos de farmacogenética y biología computacional.
- Para la aplicación la evaluación del flujo de análisis propuesto de las herramientas disponibles para la predicción de variantes missense en el gen *DPYD* así como la identificación de variantes reportadas previamente en la literatura, se usó una muestra de los VCF disponibles en el banco de datos de la unidad de analítica de datos de Biotecgen S.A.S, aplicando un modelo de minería de datos con acceso de los investigadores solamente a los datos disponibles del procesamiento delos datos crudos, los cuales se encuentran anonimizados y no permiten el acceso a información sensible, demográfica o clínica para cada caso. Adicionalmente solamente se usaron los VCF de los pacientes en los cuales Biotecgen S.A.S contaba con consentimiento informado para el uso de dichos datos.
- No se documentaron casos en los cuales en un VCF se identificarán 2 variantes potencialmente nocivas en el gen *DPYD* que pudiesen estar en homocigosis, por lo cual no se requirió notificar a Biotecgen los hallazgos individuales de la investigación para la evaluación de una condición monogénica.
- Al tratarse la investigación del análisis de variantes de potencial interés en un farmacogen, donde el fenotipo esperado para una variante determinada se relaciona con el uso de un fármaco particular, así como las recomendaciones de la guía de análisis de variantes planteada por la ACMG y sus recomendaciones frente

al reporte de hallazgos secundarios, donde para el caso del gen *DPYD*, este no se encuentra incluido dentro de los genes accionables de obligatoria notificación, no se reportó en ningún caso los hallazgos individuales.

Los VCF usados para el análisis del presente proyecto de investigación son propiedad de Biotecgen S.A.S y los autores en ningún momento dieron un manejo diferente al establecido por la empresa, la información usada para evaluar el rendimiento predictivo de diferentes herramientas de predicción *in silico* corresponderá a su respectivo autor y fue referenciado para cada caso. La propiedad intelectual de los datos analizados, así como los resultados derivados del presente proyecto son de cada uno de los autores que participaron en el proyecto, así como los productos derivados del proyecto de investigación.

Resultados

7.1 Descripción de métricas de calidad y artículos seleccionados

Se realizó la búsqueda de los artículos originales de acuerdo con la metodología planteada, en las bases de datos de PubMed, Scielo, Cochrane, Scopus y Embase encontrando un total de 359 artículos originales de los cuales posterior a filtrar de acuerdo con título, resumen, eliminación de duplicados y evaluación de calidad se obtuvieron un total de 5 artículos, el resumen del flujo de selección de los artículos se muestra en la figura 4.

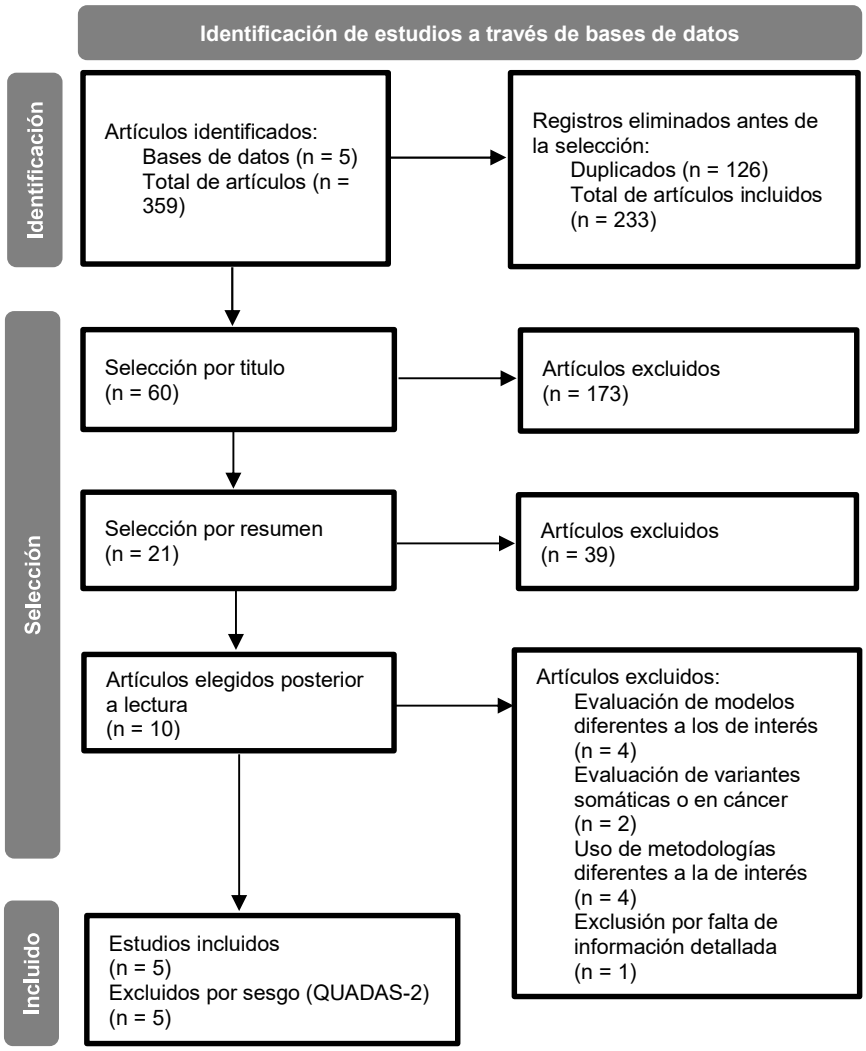


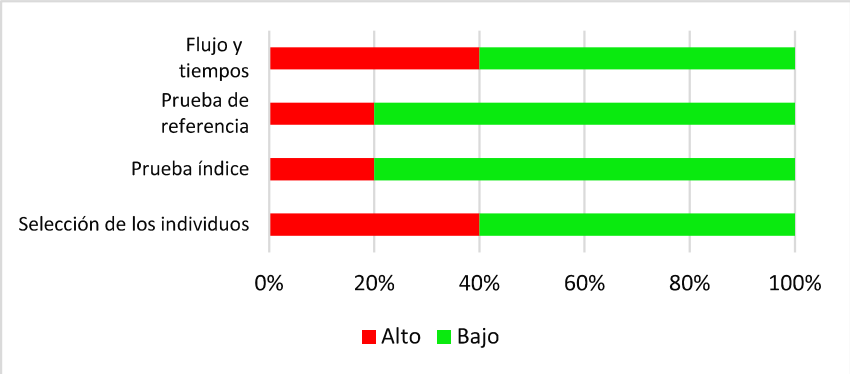
Figura 4. Flujo de selección de artículos de la revisión sistemática

Se realizó la evaluación del riesgo de sesgo para cada artículo original usando el instrumento de evaluación QUADAS-2, el cual se ajustó a las necesidades de la

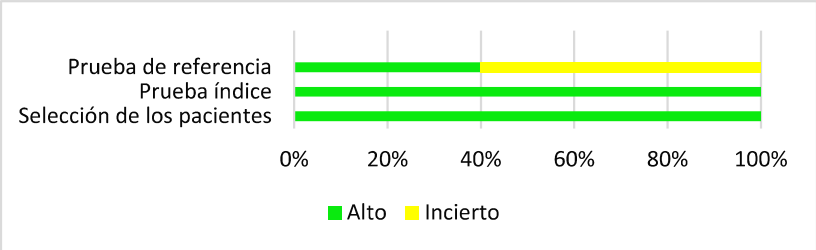
investigación y se presenta en el anexo 1. Se incluyeron artículos que evaluaran diferentes algoritmos de anotación y su precisión en la identificación de variantes nocivas; 2 de los 5 artículos incluidos para la revisión sistemática evaluaron herramientas de anotación *in silic*o en genes de interés en farmacología donde uno de ellos se enfocaba en el gen *DPYD*, el resto de los artículos se centraban en evaluación de herramientas de anotación en genes relacionados con condiciones monogénicas. El resumen de la evaluación de calidad y de la aplicación de estos artículos se evidencia en la tabla 5 y las gráficas 1 y 2.

Tabla 5. Resultados de evaluación de sesgo por la herramienta QUADAS-2 cara cada articulo

Articulo	Probabilidad de sesgos				Preocupación sobre la aplicabilidad de los resultados		
	Selección de los individuos	Prueba índice	Prueba de referencia	Flujo y tiempos	Selección de los pacientes	Prueba índice	Prueba de referencia
Shrestha et al, 2018	Baja	Alta	Baja	Baja	Baja	Baja	Baja
Rodrigues et al, 2015	Alta	Baja	Baja	Baja	Baja	Baja	Incierta
Tian et al, 2019	Alta	Baja	Alta	Baja	Baja	Baja	Incierta
Pshennikova et al, 2019	Baja	Baja	Baja	Alta	Baja	Baja	Incierta
Leong et al, 2015	Baja	Baja	Baja	Alta	Baja	Baja	Baja



Gráfica 1. Riesgo de sesgo de acuerdo con dominios de evaluación propuesto por QUADAS-2



Gráfica 2. Aplicabilidad de los datos en la revisión sistemática en el estudio propuesto

Como se presenta en la tabla 5 y en la gráfica 1, la principal causa de alto riesgo de sesgo se encuentra en la selección de las variantes y en el flujo y tiempos establecidos para los estudios. Para el primer caso se consideró que el riesgo de sesgo derivado de la obtención de variantes clasificadas como patogénicas/benignas de datos de bases no curadas como ClinVar podría estar presente por la ausencia de controles frente a los criterios usados para la asignación del efecto de la variante y la posibilidad de evidencia insuficiente para la asignación de una clasificación; también se consideró que un posible riesgo de sesgo podría aparecer durante la selección de variantes a partir de frecuencias alélicas en bases de datos poblacionales, donde se podrían caer en la falacia de la asignación de patogenicidad al tomar como criterio de clasificación la idea que las variantes raras son dañinas, donde para algunas condiciones autosómicas recesivas las variantes patogénicas pueden ser altamente frecuentes, existiendo ya recomendaciones frente al uso de estos datos para la asignación de clasificación (112).

Para el dominio “flujo y tiempos” se encontró que la aparición del riesgo de sesgo se daba por la inadecuada eliminación de variantes, encontrando que en algunos estudios las variantes que no podían ser evaluadas por alguna herramienta de anotación por no disponibilidad del algoritmo, era incluida dentro del cálculo de las métricas de rendimiento o se excluía del análisis sin realizar ajustes de desequilibrio durante la evaluación estadística de los modelos.

A nivel de la preocupación de la aplicabilidad, en general los estudios seleccionados respondían a la pregunta de investigación establecida para la revisión sistemática, sin embargo como se observa en la tabla 5 y la gráfica 2, tres artículos presentan una preocupación incierta en la prueba de referencia, lo cual se genera por las posibles fuentes del riesgo de sesgo mencionados y la selección de las variantes en bases de datos no curados sin el uso de una prueba de referencia claras como prueba *Gold standard* con la posibilidad de incluir variantes mal clasificadas, siendo un ejemplo de esto los estudios en los que se usaron variantes reportadas en ClinVar. Para el caso de las variantes disponibles en ClinVar, el reporte de la interpretación de la patogenicidad se realiza de acuerdo con los criterios emitidos por la ACMG, donde para muchos casos la asignación de patogenicidad o benignidad se puede realizar sin el uso de estudios funcionales, por lo cual se considera incierta la aplicabilidad de estos estudios por las dificultades que representa este tipo de asignación. En general en los artículos seleccionados los

investigadores plantearon una estructura de análisis que es acorde a la sugerida por QUADAS-2 para comparar pruebas diagnósticas, donde se aplicaron para cada variante los diferentes algoritmos de anotación y se calcularon las métricas de rendimiento de acuerdo con el conocimiento disponible del efecto causado en la proteína para cada variante. Sin embargo, en pocos casos se establecían con claridad las características necesarias para cumplir el criterio de selección frente a benignidad o patogenicidad.

Tabla 6. Principales descriptores de los artículos evaluados

Articulo	Gen de interés	Enfermedad de interés	Prueba de referencia	Prueba índice
Shrestha et al, 2018	*DPYD*	Riesgo de toxicidad relacionada con el uso de fluoropirimidinas	Cultivo celular con medición actividad DPD y sensibilidad 5-FU	PROVEAN, SIFT, PolyPhen, FATHMM, PhD-SNP, SNP&GO, Mutant Assessor y UMD-Predictor
Rodrigues et al, 2015	*UGT1A1*	Hiperbilirrubinemia no conjugada relacionada con el síndrome de Gilbert y el síndrome de Crigler-Najjar	Estudios *in vitro* o *in vivo* que incluyeron mutagénesis dirigida, estudios de expresión, ensayos de muestras de biopsia de hígado, administración de fenobarbital y patrón de bilis duodenal.	PANTHER, A-GVGD, SNP&GO, Xvar, SIFT, MAPP, PhD-SNP, HANSA, FATHAMM, SNAP, PMUT, Polyphen-2, SNPeffect, MutPred, CONDEL y MetaSNP.
Tian et al, 2019	*ATM, ATP7B, BRCA1, BRCA2, CFTR, COL3A1, FBN1, KCNH2, MLH1, MSH2, MSH6, MUTYH, MYBPC3, NF1, NSD1, RET, RYR2, SCN5A, TP53, TSC2, ACVRL1, AKAP9, ANK2, ANKRD11, APC, APOB, BARD1, BRIP1, CDH1, CDKN2A, CHD7, DHCR7, DMD, DNAAF1, DNAH11, DNAH5, DSG2, DSP, ENG, FBN2, FH, FLNA, GBA, GCK, KCNQ1, LDLR, LMNA, MEN1, MYH7, MYLK, NBN, NOTCH1, PALB2, PMS2, POLE, PTCH1, PTEN, PTPN11, SDHB, STK11, TNNT2, TSC1, TTN, TTR, VHL y VPS13B.*	Genes relevantes en ClinVar	Revisión en ClinVar con evaluación de variante por panel de expertos o evaluación por los siguientes remitentes: Ambry Genetics, Emory Genetics Laboratory, GeneDx, InSiGHT, InVitae y Sharing Clinical Reports Project.	CADD, MetaSVM, Eigen, REVEL, BayesDel, SIFT y PolyPhen.
Pshennikova et al, 2019	*GJB2, GJB6, GJB3*	Pérdida auditiva hereditaria no sindrómica	Estudio de segregación familiar con correlación entre genotipo-fenotipo	SIFT, FATHMM, MutationAssessor, PolyPhen, CONDEL, MutationTaster, MutPred, Align GVGD y PROVEAN
Leong et al, 2015	*KCNQ1, KCNH2, SCN5A*	Síndrome de QT largo	Caracterizados in vitro y/o estudios de cosegregación	SIFT, PolyPhen-2, PROVEAN, SNPs&GO, SNAP, Meta-SNP, PredictSNP

Dos de los artículos incluidos, los cuales se muestran la tabla 6 con sus principales descriptores, evaluaron herramientas de anotación en genes de interés farmacológico

(*DPYD* y *UGT1A1*). Sin embargo, solo el estudio que planteo un análisis comparativo de herramientas de anotación en el gen *DPYD* se centró en evaluar variantes de interés en farmacogenética, siendo particularmente interesante este estudio para la investigación desarrollada por ser el mismo gen de interés y por la presentación por parte de los autores de una herramienta diseñada por ellos para la evaluación de variantes en el gen y la aplicación de una prueba de referencia estándar para todas las variantes incluidas en el estudio. Para el caso del segundo gen, a pesar de estar disponible información para su uso en algunos medicamentos en pediatría, los autores se enfocaron en la identificación de variantes en hiperbilirrubinemia. Para el resto de los artículos incluidos, los genes de interés estudiados se relacionan con condiciones monogénicas. Para el estudio de Tian y colaboradores (99) se incluyeron 66 genes en los cuales estaban presentes algunos que condicionan riesgo para cáncer hereditario. A pesar de los criterios de inclusión este fue incluido dentro de la revisión por el uso de algoritmos de anotación que no fueron desarrollados para evaluar principalmente variantes en cáncer y el predominio de genes no relacionados con cáncer. En general se usaron diversas herramientas de anotación individuales, así como meta-predictores, siendo el único específico para el gen en estudio el presentado por Shrestha y colaboradores.

7.2 Descripción de hallazgos de revisión sistemática

Uno de los principales planteamientos de la necesidad de evaluación de herramientas de anotación in silico son las dificultades del análisis de variantes de interés en farmacología, por la falta de marcadores de toxicidad con el uso del fármaco para cada variante (91). En el estudio realizado por Shrestha y colaboradores (91) se evaluó la capacidad predictiva de ocho herramientas *In Silico* y un modelo desarrollado por los autores basado en un *random forest* donde se incluyen datos de actividad funcional de la proteína DPD, información estructural y cambios en las propiedades de los aminoácidos. Para este estudio se encontró que PolyPhen-2 (AUC 0,79; precisión 0,72), SIFT (AUC 0,76; precisión 0,73) y PROVEAN (AUC 0,77; precisión 0,72) fueron los que presentaron mejores métricas de rendimiento. Para el caso del modelo desarrollado por los investigadores (DPYD-Verifier) se encontró que el AUC fue de 0,84, siendo significativamente más alto que el resto de las herramientas y sugiriendo como regiones sensibles en la proteína los sitios de unión al sustrato, el sitio de unión FeS-I y el sitio de unión al FAD (91). El resumen de las métricas se encuentra en la tabla 7.

Tabla 7. Métricas de calidad reportadas para cada herramienta en los artículos evaluados en la revisión sistemática

	PROVEAN	SIFT	PolyPhen-2	PhD-SNP	SNP&Go	FATHMM	UMD-Predictor	Mutation Taster	Mutation Assessor	DPYD-Varifier	MAPP
Shrestha et al, 2018	S: 0,92 E: 0,49 AUC: 0,62	S: 0,80 E: 0,55 AUC: 0,63	S: 0,90 E: 0,40 AUC: 0,37	S: 0,71 E: 0,70 AUC: 0,70	S: 0,67 E: 0,59 AUC: 0,62	S: 0,78 E: 0,16 AUC: 0,35	S: 0,94 E: 0,33 AUC: 0,51	NE	S: 0,73 E: 0,52 AUC: 0,58	S: 0,73 E: 0,91 AUC 0,85	NE
Rodrigues et al, 2015	NE	SIFT Self S: 86,5 E: 44,4 AUC: 0,78 SIFT Orth S: 97,9 E: 80,0 AUC: 0,80	PolyPhen 2 Self S: 81,6 E: 30,0 AUC: 0,71 PolyPhen 2 Orth S: 88,9 E: 40,0 AUC: 0,75	S: 92,1 E: 70,0 AUC: 0,88	S: 92,1 E: 50,0 AUC: 0,83	S: 66,0 E: 40,0 AUC: 0,60	NE	NE	NE	NE	S: 79,0 E: 50,0 AUC: 0,73
Tian et al, 2019	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
Pshennikova et al, 2019	S: 0,67 E: 1,0 AUC: 0,833	S: 0,67 E: 1,0 AUC: 0,83	S: 0,67 E: 0,83 AUC: 0,750	NE	NE	S: 1,0 E: 0 AUC: 0,5	NE	S: 1,0 E: 0,33 AUC: 0,665	S: 1,0 E: 0,67 AUC: 0,83	NE	NE
Leong et al, 2015	AUC: 0,943	AUC: 0,715	AUC: 0,769	NE	AUC: 0,781	NE	NE	NE	NE	NE	NE
	CONDEL	**MutPred**	**SIFT/PolyPhen2**	**CADD**	**MetaSVM**	**Eigen**	**REVEL**	**BayesDel**	**SNEP**	**META-SNP**	**PredictSNP**
Shrestha et al, 2018	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
Rodrigues et al, 2015	v.1 S: 86,5 E: 50,0 AUC: 0,80 v.2 S: 84,0 E: 20,0 AUC: 0,71	S: 97,4 E: 70,0 AUC: 0,92	NE	NE	NE	NE	NE	NE	NE	NE	NE
Tian et al, 2019	NE	NE	S: 0,83 E: 0,923 AUC: 0,861	S: 0,926 E: 0,785 AUC: 0,871	S: 0,937 E: 0,950 AUC: 0,894	S: 0,943 E: 0,966 AUC: 0,901	S: 0,955 E: 0,797 AUC: 0,907	S: 0,957 E: 0,952 AUC: 0,908	NE	NE	NE
Pshennikova et al, 2019	S: 1,0 E: 5,0 AUC: 0,750	S: 0,33 E: 0,83 AUC. 0,583	NE	NE	NE	NE	NE	NE	NE	NE	NE
Leong et al, 2015	NE	NE	NE	NE	NE	NE	NE	NE	AUC: 0,627	AUC: 0,839	AUC: 0,603

E: Especificidad | S: Sensibilidad | AUC: Area bajo la curva | _Self: Alineación bajo configuración predeterminada | _Orth. Alineación especial con ortólogos

Rodrigues y colaboradores (5) evaluaron la precisión de 16 herramientas de anotación en la identificación de variantes en el gen *UGT1A1* relacionadas con la aparición de hiperbilirrubinemia, la herramienta que presentó un mejor rendimiento fue MutPred (Sen: 0,97, Esp: 0,70, AAC: 0,92). En general para este estudio todos los predictores aplicados presentaron adecuada sensibilidad (valor promedio de 0,829) pero bajas especificidades (valor medio 0,457) (5). Frente a la exactitud en la identificación de variantes potencialmente dañinas después de MutPred, los que le siguen con mejor precisión fueron SIFT Orth (AAC: 0,69), PhD- SNP (AAC: 0,62) y A-GVGD Orth (AAC: 0,48), la herramienta con el peor desempeño para este estudio fue FATHMM (AAC: 0,05) (5).

En el estudio realizado por Tian y colaboradores (99), se evaluó la precisión y el rendimiento diagnóstico de 5 metapredictores y 2 predictores individuales, para lo cual se usaron 4094 variantes en 66 genes accionables. De los algoritmos evaluados los que presentaron un mayor rendimiento diagnóstico fueron REVEL (AUC: 0,907) y BayesDel (AUC: 0,908). En general frente a los predictores individuales, se encontró que los metrapredictores tienden a evaluar mejor la patogenicidad de la variantes (99). En el caso de los datos presentados por Pshennikova y colaboradores (100) frente a la exactitud en la identificación de variantes nocivas en los genes *GJB2, GJB6* y *GJB3* frente a la aparición de hipoacusia no sindrómica, se encontró que los algoritmos que mostraron mayor sensibilidad y especificidad fueron SIFT (Sen: 0,67; Esp: 1,0) y PROVEAN (Sen: 0,67; Esp: 1,0). Para el caso de la precisión en la predicción el algoritmo que presentó las peores métricas fue FATHMM (AUC: 0,5) y los que presentaron mejor rendimiento fueron SIFT (AUC: 0,833) y PROVEAN (AUC: 0,83), seguidos por Mutation Assessor (AUC: 0,833) (100). Sin embargo, una limitante en este estudio fue el número de variantes evaluadas para cada uno de los genes de interés, lo que puede suponer un sesgo frente a las métricas de precisión calculada.

Finalmente, para el estudio realizado por Leong y colaboradores (113) se evaluaron 7 herramientas de anotación de forma individual o en combinación para la identificación de variantes dañinas o benignas en 3 genes relacionados con síndrome de QT largo. Para el caso del gen *KCNQ1* los algoritmos Meta-SNP, PROVEAN, PolyPhen-2 y SNPs&GO presentaron las puntuaciones de AUC más altos, sin encontrar diferencias entre las combinaciones de las herramientas *In Silico* individuales y los metapredictores (113). Para el caso del gen *KCNH2* los algoritmos que presentaron una mayor puntuación AUC en las

curvas ROC fueron PROVEAN, Meta-SNP, SIFT y SNPs&GO, sin encontrar al igual que en *KCNQ1* diferencias entre las combinaciones de los algoritmos individuales frente a los metapredictores. Para el gen *SCN5A* los predictores no presentaron un adecuado rendimiento predictivo como para los otros dos genes, pero a pesar del bajo rendimiento para este gen en general los mejores predictores fueron SIFT, Meta-SNP, SNPs&GO y PolyPhen-2 (113), las métricas generales de AUC calculadas para este estudio se muestran en la tabla 7.

En la tabla 7 se resumen los hallazgos presentados en cada una de las investigaciones seleccionadas para el desarrollo de la revisión sistemática rápida, donde los algoritmos con mejor rendimiento fueron PROVEAN, SIFT, PhD-SNP, SNP&Go, MutationAssessor, DPYD-Varifier, MutPred, Eigen, REVEL, BayesDel y META-SNP; siendo estos algoritmos en los que se reportó un mejor rendimiento teniendo en cuenta los valores del área bajo la curva (AUC). En general estos algoritmos presentaron una mayor sensibilidad que especificidad lo que podría plantear su uso en el tamizaje en la identificación de variantes potencialmente nocivas, sin embargo, por la variedad frente a la función biológica de los genes analizados y desarrollos experimentales usados, se debe tener cuidado frente a la extrapolación de estos datos a otros escenarios. Adicionalmente, se podría sugerir con los hallazgos presentados que para la identificación de variantes nocivas los metapredictores presentan un mejor rendimiento, lo cual puede estar relacionado con la inclusión de diferentes herramientas dentro de la predicción, donde a pesar de no incluir un análisis estructural, si conlleva el análisis desde diferentes aristas en el análisis unidimensional, teniendo en cuenta las implicaciones del cambio de residuo y sus implicaciones a nivel biológico.

7.3 Variantes seleccionadas para la evaluación comparativa de los algoritmos de anotación

Se identificaron 71 variantes en la base de datos de PharmVar y 69 variantes del estudio realizado por Shrestha y colaboradores (91). Las 71 variantes tomadas de la base de datos de farmacogenética se encontraban correctamente anotadas de acuerdo con las recomendaciones frente al uso de la nomenclatura para el reporte de variantes; de las variantes identificadas en el estudio de por Shrestha y colaboradores (91) fueron excluidas 3 variantes por ambigüedad frente a la nomenclatura reportada por los autores. Se

obtuvieron en total 137 variantes; para las 66 variantes tomadas de la investigación original se tuvo en cuenta el estudio funcional reportado por Shrestha. et al., 2018, para el caso de las variantes tomadas de PharmVar se tuvieron en cuenta los estudios originales de Offer y colaboradores (114,115). Como hallazgo interesante, en todas las investigaciones originales se siguieron los mismos lineamientos frente a la metodología experimental (114,115) y las validaciones realizadas para asegurar la veracidad de los hallazgos reportados. No se excluyó ninguna variante por evidencia insuficiente frente a su efecto a nivel de la proteína, las variantes documentadas se reportan en el anexo B.

7.4 Evaluación comparativa de los algoritmos de anotación

Los algoritmos de anotación en los que se encontró un mayor rendimiento de acuerdo con la revisión sistemática fueron: PROVEAN, SIFT, PhD-SNP, SNP&Go, MutationAssessor, DPYD-Varifier, MutPred, Eigen, REVEL, BayesDel y META-SNP; se planteó el uso de todo el conjunto de algoritmos para la evaluación de variantes seleccionadas, no obstante, se excluyeron 2 predictores del análisis. MutPred fue excluido, ya que este no se encontraba disponible para gran parte del conjunto de variantes, principalmente para las variantes conocidas como benignas, lo cual podría implicar un desequilibrio al momento de análisis estadístico. El segundo algoritmo excluido del análisis fue DPYD-Varifaier ya que este no está disponible para su uso y no se logró contactar a sus desarrolladores.

Los algoritmos PROVEAN y SIFT fueron las únicas herramientas en los que se encontró que para algunas variantes no se encontraba disponible la predicción o el resultado era incierto, SIFT para 1 variante tolerada y 1 variante nociva, para el caso de PROVEAN fueron 4 variantes toleradas las que no pudieron ser incluidas en el análisis; para ambos casos se excluyeron del análisis estadístico las variantes en las que no se logró establecer una predicción o esta fue incierta. Las métricas de rendimiento para las 137 variantes exceptuando el caso de PROVEAN (133 variantes) y SIFT (135 variantes) se reportan en la tabla 8 en el cual se incluye el resultado de la sensibilidad, especificidad, valor predictivo positivo y negativo, así como la exactitud diagnóstica para cada algoritmo evaluado. Para la herramienta BayesDel se realizó el análisis teniendo en cuenta la inclusión de las frecuencias alélicas poblacionales y sin estas, al igual que para la herramienta Eigen se tomó en cuenta su primera versión, así como su versión Eigen-PC.

Tabla 8. Métricas de rendimiento calculadas para cada herramienta evaluada

Predictores	Sensibilidad	Especificidad	VPP	VPN	Exactitud	MCC
Eigen-PC	0,933	0,304	0,396	0,457	0,511	0,267
Eigen	0,933	0,391	0,429	0,457	0,569	0,338
PROVEAN	0,933	0,477	0,477	0,477	0,632	0,395
BayesDel noAF	0,844	0,576	0,494	0,413	0,664	0,398
BayesDel addAF	0,822	0,674	0,552	0,402	0,723	0,466
SIFT	0,818	0,560	0,474	0,396	0,644	0,346
REVEL	0,778	0,739	0,593	0,380	0,752	0,490
MetaSNP	0,733	0,750	0,589	0,359	0,745	0,462
Mutation Assessor	0,733	0,500	0,418	0,359	0,577	0,222
SNP&Go	0,711	0,586	0,457	0,347	0,627	0,280
PhD-SNP	0,755	0,706	0,557	0,369	0,722	0,436

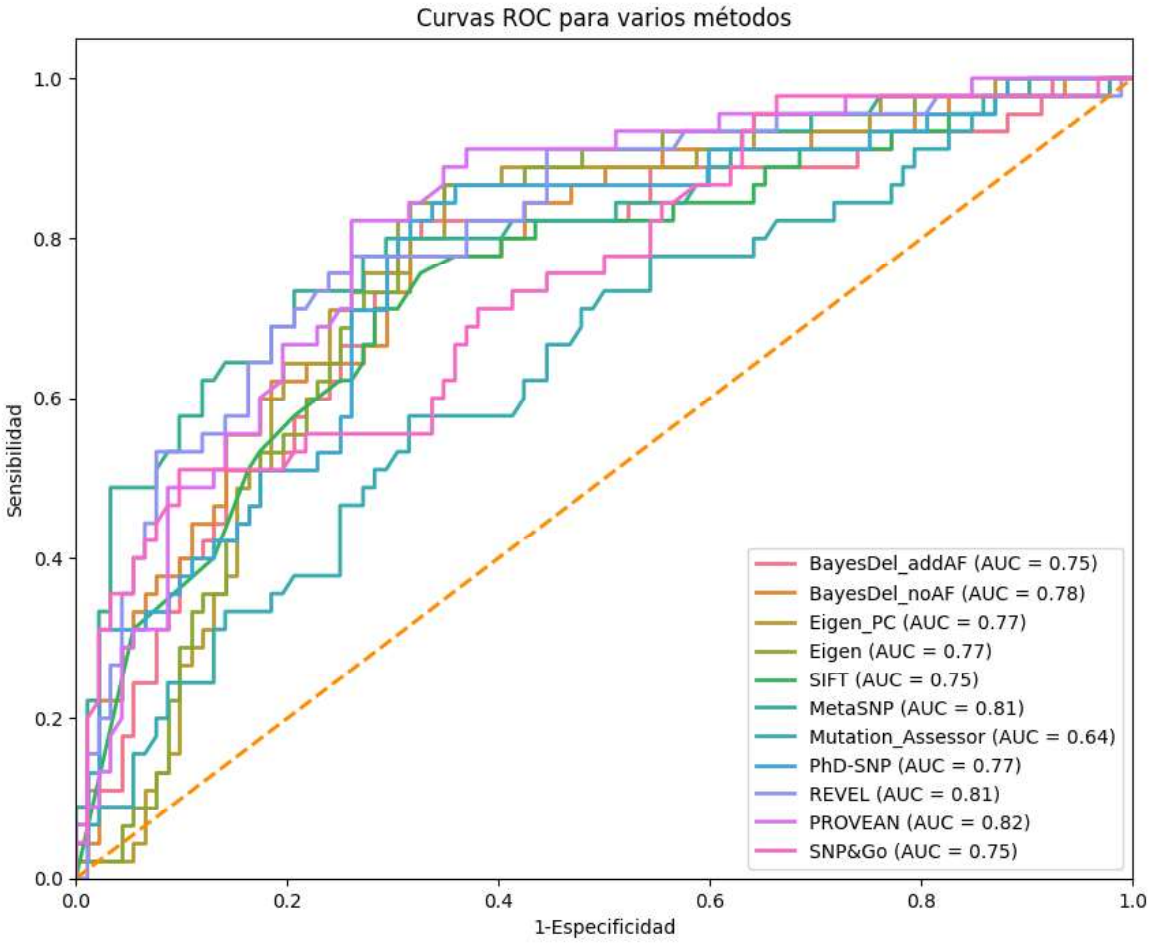
VPP: valor predictivo positivo | VPN: valor predictivo negativo | MCC: Coeficiente de correlación de Matthews | _addAF: uso de la versión que incluye en el análisis las frecuencias alélicas | _noAF: uso de la versión que no incluye en el análisis las frecuencias alélicas | -PC: uso de la versión que genera una descomposición de la matriz de covarianza

En términos generales, todas las herramientas exhibieron una mayor sensibilidad que especificidad. Eigen (incluyendo su versión -PC), PROVEAN y BayesDel noAF destacaron como las herramientas con la mayor sensibilidad. La especificidad en general fue baja para todas las herramientas, siendo Eigen (incluyendo su versión -PC) y PROVEAN las que mostraron una especificidad inferior en comparación con otras herramientas. En el caso de BayesDel noAF, su especificidad fue menor en comparación con su versión con la inclusión del análisis de las frecuencias alélicas. Esto sugiere que, dada la similitud en la sensibilidad entre las dos versiones de BayesDel, el rendimiento es mayor al incluir datos poblacionales.

Para todos los algoritmos los valores predictivos positivos y negativos tendieron a ser bajos (inferiores a 0,6), lo cual sugiere la limitación de este tipo de herramientas al momento de identificar verdaderos positivos o negativos, lo cual concuerda con los valores de exactitud que también evidencia una baja exactitud al momento de predecir el efecto de una variante missense. Para el caso de los coeficientes de correlación de Matthew se encontró que para todas las herramientas los valores son inferiores a 0,5 lo cual se relaciona con la limitación predictiva evidenciada con las otras métricas de calidad, pudiendo inferirse que las herramientas con los valores más cercanos a 0 (Eigen-PC y Mutation Assessor) los valores

predictivos pueden ser más debidos al azar que por una alta capacidad predictiva de la herramienta.

En la gráfica 3 se ilustra la curva ROC junto el valor de AUC para cada herramienta evaluada, donde en general se evidencia que los algoritmos de anotación *in silico* tienen un poder predictivo al momento de diferenciar las variantes nocivas de las neutras, sin embargo, su capacidad es limitada, donde las herramientas que presentaron mayor capacidad discriminatoria fueron PROVEAN, MetaSNP y REVEL, lo cual es acorde a los encontrado con los valores de MCC para estas herramientas.



Gráfica 3. Curva ROC con valores del área bajo la curva (AUC) para cada herramienta evaluada

Sin embargo, en términos de rendimiento general, las herramientas que sobresalieron fueron Eigen (S: 0,933, E: 0,391, ACC: 0,569, MCC: 0,338), PROVEAN (S: 0,933, E: 0,477,

ACC: 0,632, MCC: 0,395) y BayesDel addAF (S: 0,822, E: 0,674, ACC: 0,723, MCC: 0,466). Es importante destacar que, aunque Eigen-PC tuvo una sensibilidad cercana a la de Eigen, y en el caso de BayesDel noAF fue superior a BayesDel addAF, el resto de las métricas de rendimiento fueron inferiores, lo que sugiere un rendimiento general inferior.

En cuanto a los valores de AUC, las herramientas que demostraron un mejor desempeño fueron PROVEAN, REVEL y MetaSNP. A pesar de que REVEL y MetaSNP presentaron una menor sensibilidad y especificidad en comparación con BayesDel y Eigen (REVEL: S: 0,778, E: 0,739, ACC: 0,752, MCC: 0,490; MetaSNP: S: 0,733, E: 0,750, ACC: 0,745, MCC: 0,462), sus métricas de rendimiento fueron mucho más consistentes, lo que sugiere una mayor capacidad predictiva. Además, junto con PROVEAN, estas 3 herramientas demostraron un rendimiento superior en términos de exactitud, MCC y AUC en comparación con las demás herramientas, siendo en general PROVEAN la herramienta con el mayor rendimiento

Por lo cual se podría plantear una evaluación escalonada donde inicialmente se evalúen las variantes con herramientas altamente sensibles, con lo cual se puedan captar variantes de interés y posteriormente realizar una evaluación con un algoritmo con mayor precisión. Planteando el uso de BayesDel AddAF, Eigen y PROVEAN para la identificación de variantes de interés y PROVEAN, Revel y MetaSNP para la determinación del efecto.

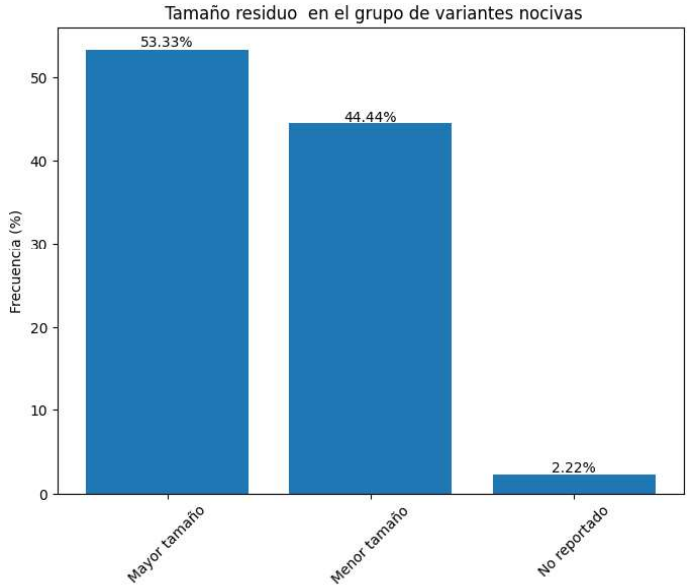
7.5 Hallazgos en el análisis estructural de las variantes

Para cada variante seleccionada se realizó en análisis con la herramienta HOPE (106) y missense 3D (107), para el caso de la información aportada por HOPE se evaluaron los hallazgos en la predicción frente al cambio en el tamaño del aminoácido, diferencias en la propiedad físico-química con el cambio de residuo, predicción en el cambio de las interacciones y predicción a nivel estructural (análisis del ensamblaje de la proteína, impacto en dominios y motivos, predicción de sitios activos y ligando y cambios a nivel de enlaces). Para el caso de la predicción realizada por missense 3D (evaluación de impacto en estructura terciaria, evaluación de sitio activo y predicción en cambios de solubilidad) solamente se tuvo en cuenta la predicción de daño o no a nivel estructural.

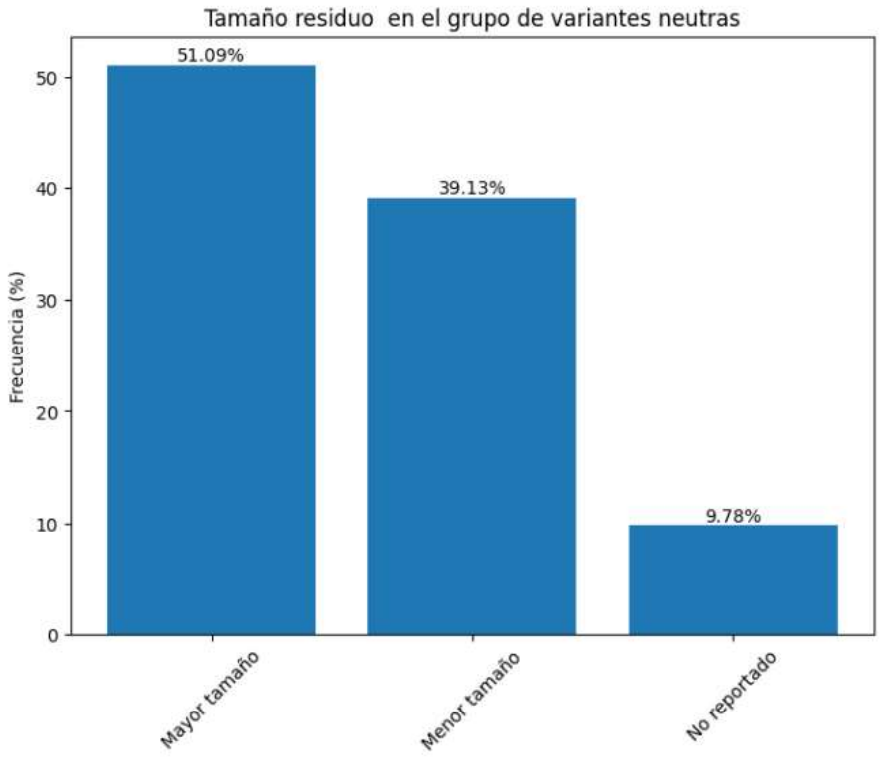
En la tabla 9 se resumen los hallazgos de acuerdo con la clasificación funcional de las variantes. Tanto para las variantes nocivas como neutras, predomina un reporte de cambio de tamaño del aminoácido (mayor o menor tamaño), siendo diferente en la predicción de interacción de la proteína relacionado con el cambio de residuo donde aumentó el número de variantes predichas como neutras. En el caso del análisis en los cambios en las propiedades fisicoquímicas entre el aminoácido de referencia y el alterno para las variantes nocivas predominó el reporte de un impacto nocivo, en el caso del grupo de variantes neutras los valores fueron cercanos. En la predicción en el impacto estructural tanto en HOPE como en missense3D se evidenció que fue menor la proporción de variantes con un reporte de daño para ambos grupos de variantes, sin embargo, para las variantes nocivas el número de variantes en las que se reportó un efecto dañino fueron cercanas para ambos predictores.

Tabla 9. Numero de variantes con predicción de cambios durante el análisis estructural

	Predicción de alteración estructural	Cambio en el tamaño del aminoácido	Cambio en las propiedades fisicoquímicas del aminoácido	Cambios en la interacción del residuo	Predicción de cambios a nivel estructural HOPE	Predicción de cambios a nivel estructural missense 3D
Nocivas	No	1	10	12	27	29
	Si	44	35	33	18	16
Neutras	No	8	43	41	66	80
	Si	84	49	51	26	10
	No disponible	-	-	-	-	2



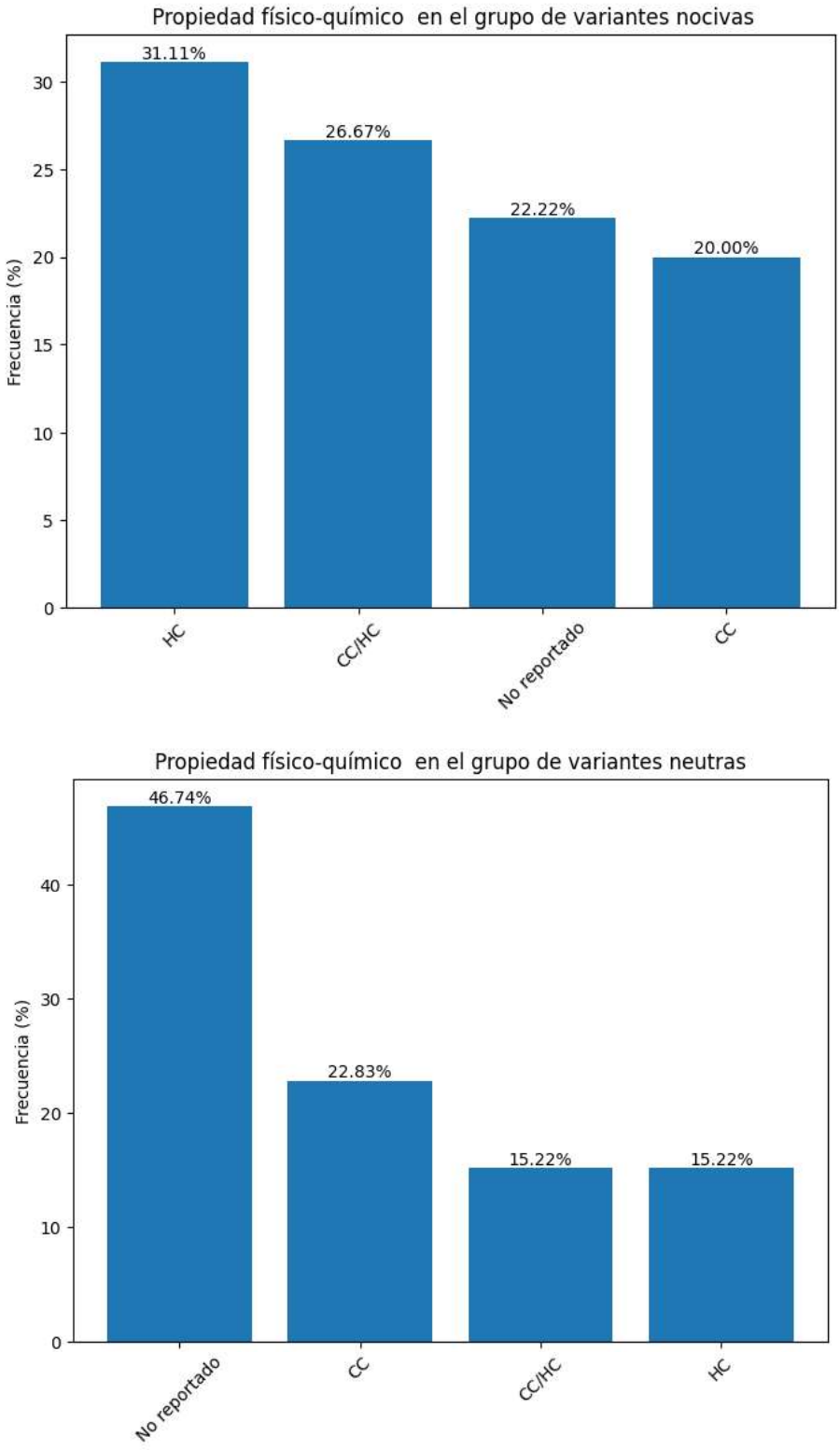
Gráfica 4. Frecuencias en el cambio de tamaño del aminoácido



Gráfica 5. Continuación: Frecuencias en el cambio de tamaño del aminoácido

Al evaluar cada característica del análisis estructural, en el cambio del tamaño del aminoácido (gráfica 4) se encontró que para ambos grupos de variantes predominó un cambio en el tamaño de aminoácido, siendo para ambos casos mayor el número de variantes en las que el nuevo residuo fue de mayor tamaño frente al *wildtype*. Para ambos grupos la frecuencia en la que el cambio de residuo tuvo un tamaño similar al de referencia fue menor al 10%.

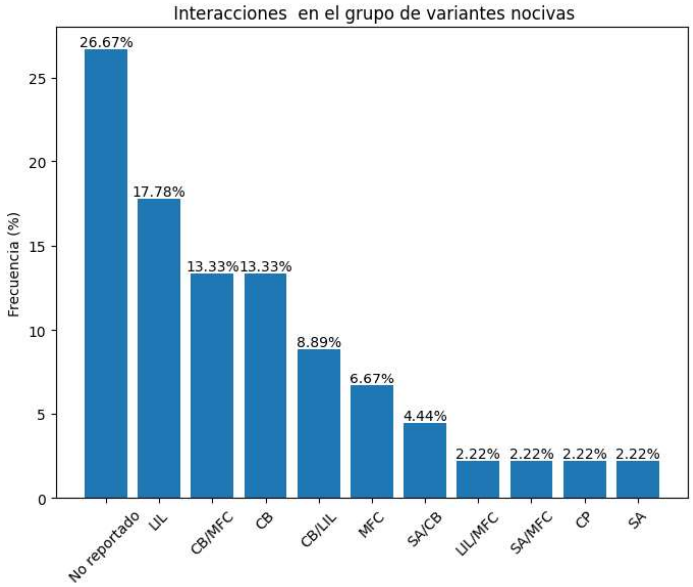
Para el caso del análisis en las propiedades fisicoquímicas (gráfica 5), en el grupo de SNV nocivas la proporción de variantes en las que se encontró un cambio en las propiedades fue mayor que en las que no, donde los residuos principalmente presentaban un cambio a nivel de la hidrofobicidad y anfipatía. En el grupo de SNV neutras la predicción en el cambio de las propiedades fisicoquímicas de los residuos alternos frente a los de referencia, la proporción de residuos alternos donde existía un cambio en las propiedades fueron cercanas a las que no se encontró un cambio.



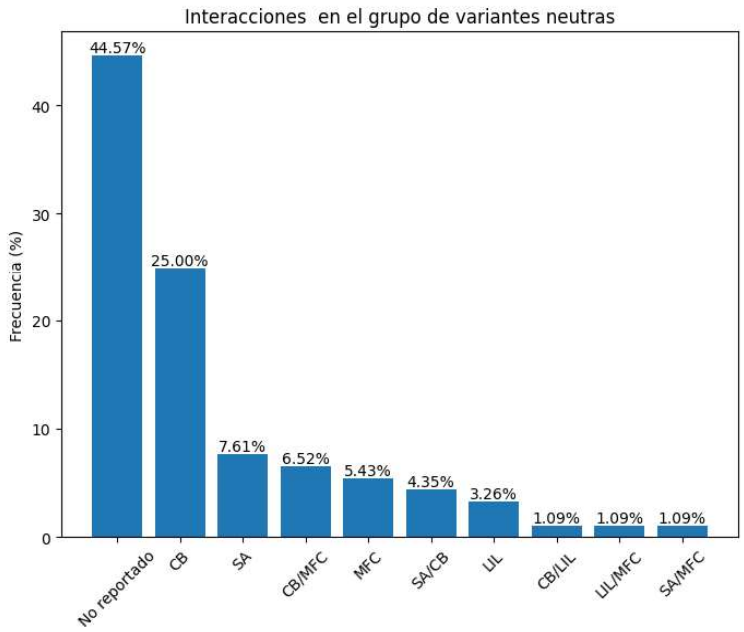
Gráfica 6. Descripción de cambio en propiedades fisicoquímicas en el cambio de aminoácidos CC: Cambio en carga eléctrica del aminoácido | HC: Cambio en hidrofobicidad y antipatía del aminoácido

Para el caso del análisis de los cambios en la interacción del residuo con otras estructuras o residuos cercanos (gráfica 6), se encontró que para el grupo de SNV nocivas fue mayor la proporción de variantes en las que se reportó algún cambio, encontrando que se afectaba principalmente la unión con un ligando o la funcionalidad de los sitios activos de la enzima, seguido por una alteración en la formación de enlaces y una alteración al momento de la formación de multímeros o la combinación de estas. Para el caso de las variantes neutras la proporción de variantes en las que se informó de un cambio en la interacción con las que no fueron cercanas, donde para el caso de las variantes en las que se informó un cambio, se encontró principalmente un cambio en la formación de enlaces y la estabilidad de la proteína derivada de la interacción con residuos cercanos. Para las variantes neutras fue menor la proporción de SNV que afectaban la unión con sustratos.

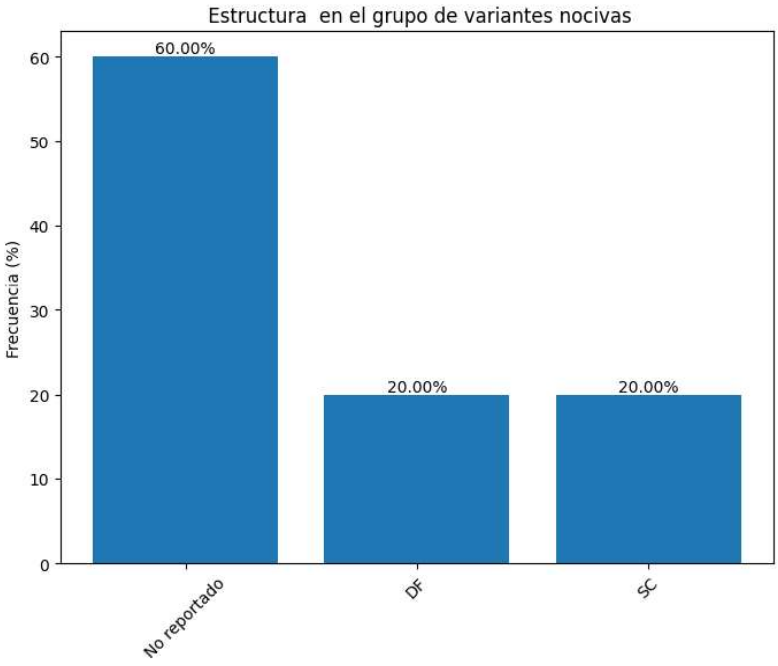
En la predicción del efecto estructural de las variantes, se encontró como principal hallazgo un reporte en el cual no se reportaba una alteración en la estructura de la proteína; para el caso de las variantes nocivas en las que se reportó un efecto en la estructura la proporción entre el cambio conformacional y una alteración de la función de un dominio fueron iguales, para el caso de las variantes neutras se encontró que principalmente se predecía un cambio en la conformación de la proteína.



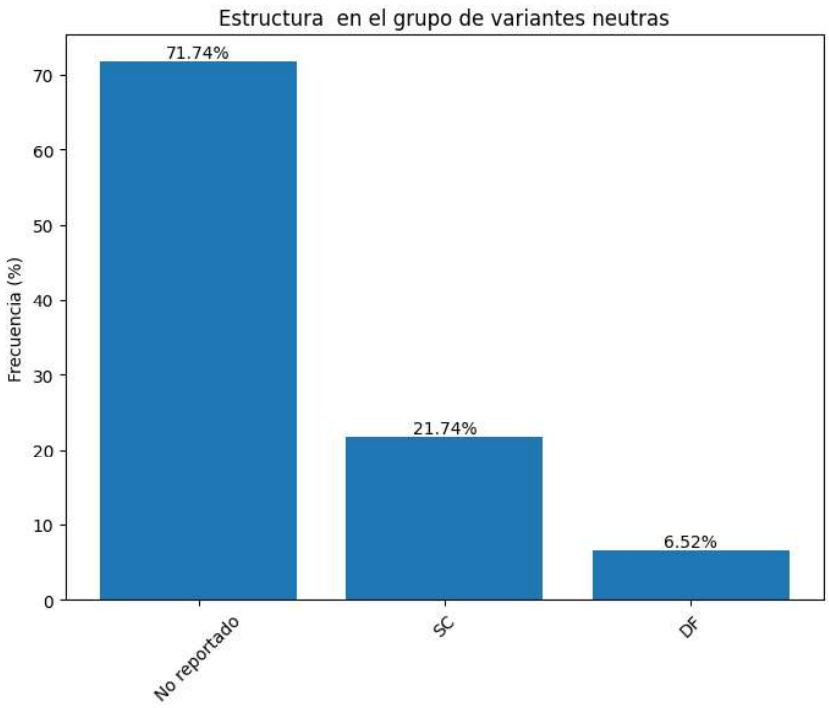
Gráfica 7. Descripción de cambios de interacción del residuo
CB: Cambio en el tipo de enlaces | LIL: Pérdida de la interacción con el ligando o sitio activo | SA: Alteración en la estabilidad por cambios en la interacción del residuo | MFC: Cambio en la formación de multímeros | CP: Residuo altamente conservado



Gráfica 8. Continuación: Descripción de cambios de interacción del residuo
CB: Cambio en el tipo de enlaces | LIL: Perdida de la interacción con el ligando o sitio activo | SA: Alteración en la estabilidad por cambios en la interacción del residuo | MFC: Cambio en la formación de multímeros | CP: Residuo altamente conservado



Gráfica 9. Descripción de los cambios en la estructura de la proteína predichos por HOPE
DF: Predicción de perdida función del dominio | SC: Predicción de alteración en conformación de la proteína



Gráfica 10. Continuación: Descripción de los cambios en la estructura de la proteína predichos por HOPE
DF: Predicción de pérdida función del dominio | SC: Predicción de alteración en conformación de la proteína

Durante el análisis de los datos reportados en missense 3D–DB se identificó que para ambos grupos de variantes, la proporción de variantes con predicción de un efecto nocivo fue menor que en las que no se predijo una alteración estructural, sin embargo, para el caso de las variantes nocivas el número de variantes con un reporte de daño fue mayor que para las neutras (35.5% vs 10.86%). Para las variantes neutras 2 de las SNV no se encontraron reportadas en la base de datos y no se logró la predicción del efecto en missense 3D ya que los modelos disponibles de la proteína (DPD) no incluyen estos residuos dentro del modelo.

Al momento de evaluar el efecto de las sustituciones de los residuos en la proteína DPD (disponibles en el anexo C) y su agrupación dicotómica de acuerdo con el efecto documentado en los estudios funcionales "Nocivas" o "Neutras", se encontró que para el caso de las variantes nocivas se predice principalmente la interrupción de enlaces disulfuro y puentes salinos, la contracción del volumen de la cavidad, cambios en la polaridad de los residuos, alertas de colisión y ángulos phi/psi no permitidos y para el caso de las variantes Neutras sustitución de residuos cargados por no cargados o modificaciones en

el volumen de la cavidad. Pudiendo proponer que para el caso de las variantes nocivas los cambios tienden a generar una mayor perturbación a nivel de la estructura de la proteína.

7.6 Propuesta de protocolo de análisis de variantes

Identificación de variantes de interés en el gen *DPYD*

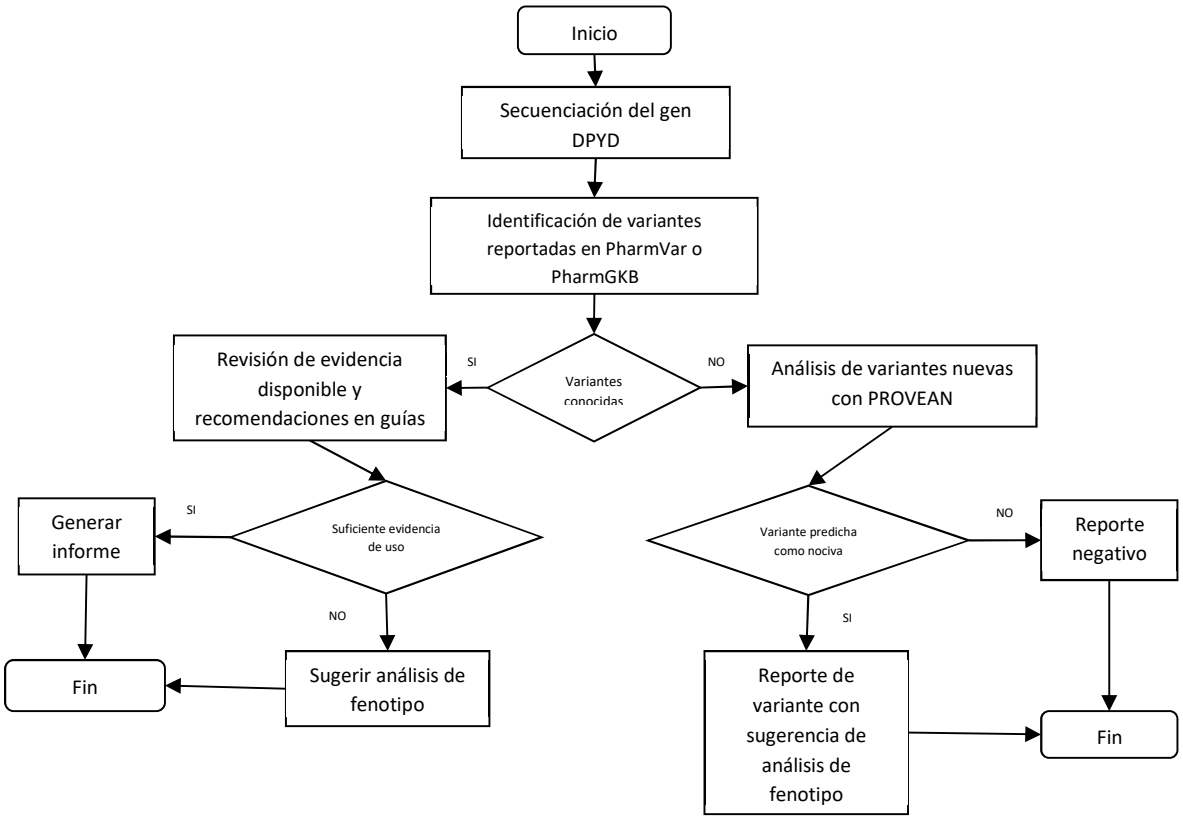


Figura 5. Propuesta de protocolo de análisis del gen DPYD

Con los hallazgos del análisis comparativo realizado, se construyó un protocolo de análisis de los datos derivados de la secuenciación del gen *DPYD*; en un primer paso se plantea realizar la búsqueda de las variantes conocidas como nocivas reportadas en las bases de datos de farmacogenética PharmVar o PharmGKB. En caso de documentar variantes ya reportadas en estas bases de datos, se sugiere realizar la revisión de las guías de ajuste de fenotipo y de acuerdo con los hallazgos, se deberá reportar la variante encontrada con la respectiva recomendación del ajuste de dosis, para los casos en los cuales no se cuente con recomendaciones, se deberá sugerir ampliar estudios con análisis de fenotipo. Para los casos en los que se documenten variantes no reportadas previamente, se indica el análisis de la variante con PROVEAN; para los casos en los que la herramienta indique un

efecto nocivo se deberá reportar la variante y sugerir el análisis de fenotipo. Para el resto de los casos el reporte deberá emitirse como negativo. El protocolo planteado se encuentra en la figura 5.

7.7 Análisis poblacional de muestras del banco de datos

Los VCF usados para el análisis poblacional fueron obtenidos del banco de datos de la unidad de analítica de datos de Biotecgen S.A.S; en el flujo de análisis de Biotecgen S.A.S se anota que la secuenciación fue realizada con enriquecimiento de los exones codificantes de los genes incluidos para el análisis de exoma completa (>20000 genes) a través de la tecnología de captura híbrida (Agilent SureSelect All Exon V6), la secuenciación fue realizada por medio de tecnología de secuenciación de Nueva Generación (Next Generation Sequencing technology) en Macrogen Inc., la profundidad promedio en general para cada experimento fue de 100X, el mapeo y llamado de variantes fue realizado usando el Genoma de referencia GRCh37 y de acuerdo al flujo de análisis bioinformático estipulado por Biotecgen S.A.S. En las muestras usadas para el análisis poblacional, se encontró que para el gen DPYD la profundidad promedio de secuenciación en las regiones exónicas fue de 79,68X, con profundidades promedio mínimas de 53,26X. Los promotores, regiones no traducidas y otras regiones no codificantes presentan bajas profundidades por las limitaciones propias de la metodología de secuenciación usada. La cobertura promedio para el gen fue de 69.68% y con cobertura completa para los exones codificantes del gen.

En la tabla 10 se consignan los datos de las frecuencias poblacionales de las variantes identificadas en el conjunto de 1000 muestras del banco de datos de la unidad de analítica de datos de Biotecgen S.A.S, en la cual cada variante se caracteriza por su ubicación de acuerdo con el genoma de referencia GRCh37, su identificador en la base de datos dbSNP, así como los nucleótidos de referencia y alternativo, su efecto reportado en PharmVar así como la frecuencia alélica. Para el caso de las 4 variantes que las guías sugieren ser tamizadas para el ajuste de dosis de acuerdo con fenotipo c.190511G>A (rs3918290, *DPYD**2A), c.1679T>G (rs55886062, *DPYD**13), c.2846A>T (rs67376798) y c.1129–5923C>G (rs75017182, HapB3), se encontró una frecuencia alélica de 0,0020 para la variante *DPYD**2A, de 0,0035 para la variante rs67376798 y de 0,0090 para la variante Hap3; la variante *DPYD**13 no fue identificada en este grupo de datos.

Tabla 10. Frecuencias poblacionales de variantes en el gen *DPYD* en la muestra de datos

Posición (GRCh37)	dbSNP	ID	Ref	Alt	Efecto PharmVar	Efecto estudios	AF	p-value HWE
1:97544543	rs114096998	-	G	T	N/A	Neutra	0,0030	0,2733
1:97544612	rs151074666	-	C	T	N/A	Neutra	0,0005	0,3171
1:97544695	rs145529148	-	T	C	N/A	Neutra	0,0010	0,1573
1:97547947	rs67376798	-	T	A	Disminuida	Nociva	0,0035	0,2801
1:97658762	rs199777072	-	C	T	N/A	Neutra	0,0155	0,3094
1:97770920	rs1801160	*DPYD**6	C	T	Conservada	Neutra	0,0310	1,0000
1:97915614	rs3918290	*DPYD**2A	C	T	Nula	No incluida	0,0020	0,2482
1:97915624	rs17376848	-	A	G	Conservada	No incluida	0,0825	0,2154
1:97981395	rs1801159	*DPYD**5	T	C	Conservada	Neutra	0,2555	0,3181
1:97981407	rs142619737	-	C	T	N/A	Neutra	0,0010	0,1573
1:97981421	rs1801158	*DPYD**4	C	T	Conservada	Neutra	0,0170	0,3101
1:98015291	rs72975710	-	G	A	N/A	Neutra	0,0015	0,2207
1:98039419	rs56038477	HapB3	C	T	Disminuida	No incluida	0,0090	0,3035
1:98039437	rs61622928	-	C	T	Conservada	Neutra	0,0055	0,2943
1:98058782	rs201785202	-	G	A	N/A	Neutra	0,0005	0,3171
1:98144726	rs45589337	-	T	C	N/A	Neutra	0,0045	0,2888
1:98165030	rs115232898	-	T	C	Disminuida	Nociva	0,0040	0,2850
1:98165091	rs2297595	-	T	C	Conservada	Neutra	0,0420	0,0259
1:98187098	rs200562975	-	T	C	N/A	Neutra	0,0050	0,2918

Las variantes con mayor frecuencia poblacional para el conjunto de datos fueron la variante rs1801159 (*DPYD**5) y rs17376848 con una frecuencia de 0,2555 y 0,0825 respectivamente, para ambos casos con un efecto conservado. Adicionalmente a las 3 variantes recomendadas por la literatura causantes de pérdida de función y de uso clínico, se documentaron dentro de los datos otras 2 variantes que tanto en Pharmvar como en los estudios funcionales tomados para el análisis comparativo se describen como nocivas, las cuales fueron rs67376798 y rs115232898 en ambos casos con una frecuencia alélica menor a 0,0040.

7.8 Nuevas variantes documentadas

Posterior a la búsqueda de las variantes no reportadas previamente se identificaron 4 variantes que no se encontraban previamente reportadas en bases de datos de farmacogenética las cuales se reportan en la tabla 11, de estas 4 variantes solamente 1 variante fue clasificada como nociva por los algoritmos de selección como de clasificación,

para el resto de variantes los predictores en general indicaron un efecto neutro exceptuando la variante S878R, para la cual BayesDel y MetaSNP indicaron un posible efecto nocivo, sin embargo, para el caso de MetaSNP el RI para la producción para esta variante fue de 1, lo cual sugiere que disminuye la probabilidad de que la predicción sea correcta. En la tabla 12 se reportan los parámetros de calidad para las variantes identificadas, donde la profundidad para la variante reportada como nociva fue de 17X con una profundidad para la posición de 42X. Para todas las variantes se encontraron adecuadas métricas de calidad dado por una calidad condicional del genotipo y un *likelihoods* genotípico en escala Phred con un nivel de confianza alto frente a la llamada de las variantes.

Tabla 11. Variantes no reportadas previamente (nuevas) identificadas en el conjunto de datos

Chr	GRCh37	GRCh38	Ref	Alt	Proteína	BayesDel addAF	Eigen	PROVEAN	REVEL	MetaSNP
1	97544668	97079112	G	A	T981I	N	N	N	N	N
1	97564177	97098621	A	C	S878R	D	N	N	N	D
1	98015239	97549683	T	A	E467D	N	N	N	N	N
1	98058775	97593219	T	G	E376A	D	D	D	D	D

_adAF: uso de la versión que incluye en el análisis las frecuencias alélicas | N: neutra | D: Nociva

Tabla 12. Parámetros de calidad de variantes no reportadas previamente

Chr	GRCh37	GRCh38	Profundidad posicion	Profundidad variante	GQ	PL
1	97544668	97079112	117	46	293	PASS:294,0,370
1	97564177	97098621	50	22	207	PASS:209,0,278
			26	9	99	PASS:202,0,479
1	98015239	97549683	119	41	223	PASS:225,0,370
			124	62	99	PASS:1573,0,1503
			99	46	99	PASS:1174,0,816
1	98058775	97593219	42	17	99	PASS:419,0,748

GQ: Calidad condicional de genotipo | PL Phred-scaled genotype likelihoods rounded to the closest integer

La variante identificada fue evaluada por HOPE (106) y por missense 3D (116), para ambos casos se usó como proteína de referencia la disponible en UniProt de DPD humana (ID Q12882), y para el caso de missense 3D para el modelo estructural PDB ID: 1H7X_B. En el análisis estructural se encontró que el nuevo residuo difiere en la carga, así como la hidrofobicidad frente al residuo *wildtype*, lo que conlleva a impacto nocivo a nivel estructural

por una alteración con la interacción con otros residuos, así como con la conformación derivada de la presencia de este residuo al interior de la proteína, siendo clasificada por missense 3D como nociva. En el análisis poblacional se encontró que esta variante se halló con una frecuencia alélica de 0.0005 en la muestra de datos analizados, estando en equilibrio de Hardy-Weinberg. El resumen de los hallazgos se encuentra en la tabla 12.

Tabla 13. Descriptores de nueva variante identificada

Descriptor		**Variante**
Posición GRCh37		1:98058775
posición GRCh38		1:97593219
Ref.		T
Alt.		G
MAF		0.0005
p-value HWE		0.3171
Proteína		E376A
BayesDel addAF		Nociva
Eigen		Nociva
PROVEAN		Nociva
MetaSNP		Nociva
HOPE	**Tamaño residuo**	Menor tamaño
	Propiedad fisicoquímica	Cambio de carga (carga negativa por neutra) y el nuevo residuo es más hidrofóbico que el Wildtype
	Interacciones	Alteración en la formación de enlaces
	Estructura	Sin cambio
Missense 3D	**Cambio**	Nociva
	Descripción	Esta sustitución reemplaza un residuo cargado al interior de la proteína (GLU, RSA 7.2%) con un residuo no cargado (ALA) y adicionalmente interrumpe un puente de sal formado por el átomo OE1 de GLU 376 y el átomo NE de ARG 235 (distancia: 3.902 Å).

Discusión de resultados

Con el uso de la secuenciación de forma masiva para el diagnóstico médico, ha aumentado la información disponible para el análisis clínico y con esto una de las dificultades que se ha generado con relación a este análisis, es la interpretación del efecto de las variantes identificadas, lo cual no solo sucede en el escenario del diagnóstico de enfermedades monogénicas, sino que también se extrapola a la farmacogenética, donde generar evidencia suficiente al momento de establecer el efecto de las variantes identificadas es un gran reto. De forma clásica para la caracterización de variantes se requieren estudios de evaluación funcional donde no solamente se recolecte información a nivel celular sino también a nivel clínico, que finalmente permita establecer la causalidad entre la presencia de una variante y un efecto clínico. Sin embargo, realizar este tipo de estudios es costoso y requiere mucho tiempo para su ejecución, por lo cual se ha planteado el análisis bioinformático como una medida alternativa para el análisis de la variación en el genoma humano (5).

Actualmente se ha ido avanzando en la generación de evidencia y ampliación del conocimiento frente al uso de la genotipificación para la toma de decisiones a nivel clínico, donde el enfoque en cáncer es uno de los principales frentes de análisis por los costos que implica el tratamiento de esta enfermedad y las implicaciones a nivel de desenlaces clínicos basados en el tratamiento individualizado. Para el caso del gen *DPYD* existe un particular interés por su relación con la toxicidad con el uso de las fluoropirimidinas, sin embargo, una dificultad para su aplicación de forma generalizada en el contexto colombiano son los pocos datos disponibles frente a las frecuencias de las variantes accionables y los pocos estudios realizados con población colombiana, por lo cual en este estudio planteó el uso del análisis computacional como alternativa para la identificación de variantes potencialmente nocivas.

Como parte inicial del análisis se realizó una búsqueda en la literatura que permitiera seleccionar los algoritmos de anotación con mayor rendimiento diagnóstico, por lo cual se diseñó una revisión rápida de la literatura. Inicialmente se efectuó una búsqueda de artículos donde se evaluará el rendimiento diagnóstico de los algoritmos de anotación en genes de interés en farmacología, encontrando un número limitado de artículos para

algunas bases de datos o en otras siendo nulo el número de artículo encontrados, por lo cual se amplió la búsqueda incluyendo artículos que evaluaran estas herramientas en genes relacionados con condiciones monogénicas. Lo cual establece una limitante y sesgo dentro del estudio, ya que a pesar de buscar un enfoque basado en generar evidencia frente al rendimiento de estas herramientas computacionales, la características biológicas entre las variantes de cambio de sentido que se relacionan con la aparición de enfermedad son diferentes a las que se relacionan con la respuesta o la aparición de toxicidad relacionada con el uso de fármacos, donde para el caso de variantes que podrían estar relacionadas con una enfermedad, se espera que el efecto sea mucho más nocivo afectando completamente la función del producto del gen y así la aparición de una condición mórbida, mientas que para las segundas los cambios suelen ser más sutiles frente a la conservación del efecto biológico y solamente ser relevantes al momento de la exposición del individuo a un medicamento (80).

Las diferencias mencionadas pueden implicar un rendimiento diagnóstico diferencial entre ambos tipos de variantes, donde por ejemplo se plantea que los algoritmos basados en evaluación evolutiva de regiones conservadas podrían ser de gran utilidad para establecer relación causal entre la presencia de una variante y la aparición de una enfermedad, pero no ser útiles para la evaluación de variantes en farmacogenes, donde muchas veces la función biológica se conserva. Sin embargo, a pesar de la limitación secundaria a la información disponible, se evidencia la necesidad de generar información de la utilidad de las herramientas disponibles en la identificación de variantes de interés o el desarrollo de herramientas de evaluación específica para este tipo de genes.

Una dificultad al momento de realizar la revisión sistemática fue la heterogeneidad frente a los genes involucrados en cada ensayo, así como el *gold standard* tomado como referente al momento de evaluar el rendimiento diagnóstico de los algoritmos de predicción; a pesar del desarrollo de diversas herramientas experimentales para evaluar la función proteica, el desarrollo individualizado de acuerdo a la función biológica del producto del gen o la característica particular en investigación, dificulta la homogeneidad al momento de establecer un patrón de oro para la identificación de variantes nocivas y su comparación con otras herramientas. A nivel clínico se han desarrollado guías para la consideración y valoración de estudios funcionales como las recomendaciones del Clinical Genome Resource (96), pudiendo ser útiles en investigación, pero aun así se presenta la dificultad

relacionada con la heterogeneidad en los ensayos disponibles incluso para una misma variante.

Lo evidenciado en la revisión sistemática frente a la selección de cohortes y pruebas de referencia para análisis de datos genéticos es un desafío que previamente ha sido planteado, donde la dificultad por la baja prevalencia de muchas condiciones, frecuencias alélicas bajas de variantes de interés, problemas para la selección de cohortes de pacientes, diferencias en la distribución de variantes de acuerdo con la ancestría de los grupos poblacionales y dificultades frente a la validez sea experimental o diagnóstica de la prueba de referencia, limitan el desarrollo de metodologías que puedan brindar suficiente fuerza de evidencia para la validación de test genéticos y herramientas de análisis computacionales aplicables en un entorno clínico (117), por lo cual el cálculo de métricas de rendimiento con validez clínica puede suponer un reto mayor que para otros tipos de herramientas diagnósticas.

En general para los artículos revisados, las herramientas con mayor rendimiento fueron aquellas que se basan en análisis más complejos que no solamente predicen un efecto a partir de la comparación y evaluación de regiones conservadas, así como los meta scores basados en la ponderación de diversas herramientas. Sin embargo, para los hallazgos de los artículos revisados, exceptuando el estudio realizado por Pshennikova y colaboradores (100) estas herramientas presentaron mejor sensibilidad que especificidad, lo cual puede estar relacionado con el bajo número de variantes usadas en este estudio, por lo cual la decisión de los algoritmos que se incluyeron para la evaluación comparativa en el gen *DPYD* se basa principalmente en herramientas altamente sensibles y con un AUC cercana a 1.

A partir de los hallazgos de la revisión de la literatura se seleccionaron las herramientas PROVEAN, SIFT, PhD-SNP, SNP&Go, MutationAssessor, *DPYD*-Varifier, MutPred, Eigen, REVEL, BayesDel y META-SNP. Sin embargo, *DPYD*-Varifier y MutPred fueron excluidos del análisis, la primera por no disponibilidad de la herramienta para su uso en el estudio y la segunda por una limitación frente a la disponibilidad de la herramienta para el conjunto de variantes analizadas. Idealmente la selección de los algoritmos se debió basar en el análisis estadístico de las métricas reportadas para cada caso, pero para este análisis no se consideró apropiado por las diferencias frente al tamaño de la muestra analizada, las

diferencias frente a la función biológica de los genes analizados y la falta de homogeneidad en el tipo de prueba de referencia seleccionado para cada estudio, por lo cual se optó solamente por realizar el análisis descriptivo y a partir de este generar la selección de las herramientas.

Las herramientas de anotación fueron desarrolladas principalmente como alternativa para la evaluación de variantes asociadas con condiciones monogénicas (76), lo cual implica una dificultad al momento de evaluar su utilidad en genes de interés en farmacogenética, siendo el gen *DPYD* un ejemplo de esto. En la literatura se ha reportado un estudio similar al desarrollado, en el que al igual que en el presente estudio, se planteó la necesidad de alternativas de evaluación de las variantes en el gen, ya que del conjunto de variantes reportadas, son pocas las que cuentan con suficiente información sobre su utilidad a nivel clínico y existe infra representación de otras poblaciones diferentes a las caucásicas (91), por lo cual un clasificador *in silico* podría ser una alternativa para el análisis de estas variantes.

A partir del análisis de las variantes, como se mencionó con anterioridad, fue mayor la sensibilidad que la especificidad, lo cual también fue reportado en el estudio realizado por Shrestha y colaboradores. Una mayor sensibilidad no solamente ha sido reportada durante el análisis del gen *DPYD* sino que también ha sido un hallazgo común durante el análisis de genes relacionados con condiciones mórbidas, donde se prioriza el análisis de las regiones relevantes dentro del gen, así como los daños que puedan impactar de forma negativa en la conformación estructural de la proteína, por lo cual en el análisis de grandes grupos de genes, estas herramientas tienden a disminuir la probabilidad de tener falsos positivos (5,73,118). Se plantea la posibilidad de que esto se deba principalmente a la forma en que se han desarrollado estas herramientas, donde se tiende a generar una hipótesis en la cual las regiones altamente conservadas, así como los residuos que se encuentran dentro de dominios activos o se relacionan con la conformación estructural de la proteína son los que pueden afectar considerablemente la función biológica. Lo cual puede ser cierto de forma general al momento de evaluar una variante en el contexto de enfermedad, pero no necesariamente en los genes de interés en farmacología.

Para el caso de este estudio los predictores con mayor sensibilidad fueron PROVEAN, Eigen y Eigen-PC, para los 3 casos de 0,933, lo cual es llamativo por el tipo de análisis

realizado por cada herramienta, donde para el caso de PROVEAN se enfoca en generar una puntuación basada en el análisis de la conservación a partir de evaluación de secuencias homologas y del contexto del residuo frente a la secuencia que lo rodea (78) y para el caso de Eigen su predicción se basa en un algoritmo no supervisado que integra diversas anotaciones funcionales (análisis de conservación evolutiva, análisis funcional y frecuencias alélicas) (79). Lo cual sugiere que el enfoque necesario para identificar las variantes nocivas durante el análisis de genes de interés en farmacología debe hacerse a partir de diferentes enfoques, donde no solamente se realice en función de la conservación, ya que como se ha descrito en la literatura, a pesar de que las variantes relacionadas con la respuesta o aparición de toxicidad con el uso fármacos altera la función del producto proteico, en general la función biológica se conserva lo cual conlleva a que estas variantes no supongan una pérdida de la capacidad adaptativa desde el punto de vista evolutivo (80).

Por lo cual, herramientas en las cuales el valor predictivo se basa en identificar variantes potencialmente nocivas que afectan completamente la funcionalidad del producto del gen y así relacionar el defecto con un fenotipo mórbido, podrían no ser útiles en farmacogenética. Por lo tanto al momento de plantear herramientas candidatas o desarrollar nuevas alternativas para el análisis e interpretación de variantes a nivel de farmacogenética, es clave que se tenga en cuenta que la herramienta debe tener una gran capacidad de correlacionar el efecto de cambios sutiles a nivel estructural con los cambios en la interacción de la proteína de interés con otras moléculas al momento de generar una predicción, ya que al plantear que *per se* estas variantes no generan una alteración fisiológica importante, el enfoque usado a nivel clínico para las variantes relacionadas con condiciones monogénicas no es útil.

Para el caso de la especificidad los algoritmos con mayor rendimiento fueron MetaSNP y Revel, aunque muy inferiores frente a lo reportado para la sensibilidad. Para el caso de los valores del VPP y VPN fueron en general cercanos para todos los predictores e inferiores a 0,6, siendo hallazgos similares a lo reportado en otros estudios donde se anota una mayor sensibilidad frente a las otras métricas de rendimiento (5,91,99), lo cual podría plantear el enfoque de este tipo de predictores basados principalmente en la identificación de variantes potencialmente nocivas sacrificando el rendimiento frente a la exclusión de variantes neutras, con lo cual se podría plantear la hipótesis del uso de estos predictores

como herramienta de análisis para el tamizaje durante la identificación de variantes que podrían ser nocivas, que puedan ser llevadas a análisis adicionales para confirmar su efecto a nivel del producto del gen; pudiendo ser esto una alternativa para la identificación de variantes accionables en genes de interés de farmacología, sugiriendo una alternativa de identificación inicial de variantes de riesgo que puedan ser llevadas a un análisis fenotípico al usar predictores altamente sensibles.

Para todos los algoritmos se encontró que en general la capacidad predictiva para la identificación de variantes nocivas como neutras es baja, en algunos casos como el de Eigen-PC y Mutation Assesor con valores de MCC cercanos a 0. Para el caso de la exactitud las mayores valores fueron para MetaSNP y Revel con valores de MMC igualmente altos, sin embargo, en los valores de AUC los mejores resultados fueron para PROVEAN, Revel y MetaSNP con valores superiores a 0.8 lo cual sugiere un muy buen desempeño de acuerdo a los valores de corte reportados en la literatura para la interpretación de las curvas ROC (119). De acuerdo con los hallazgos en las diferentes métricas se considera que la herramienta con el mejor rendimiento general fue PROVEAN, donde presenta una alta sensibilidad y una AUC de 0.82, pudiendo sugerir su uso como tamizaje para variantes en el gen al momento de la secuenciación completa de *DPYD*, teniendo en cuenta que para pruebas altamente sensibles, un resultado negativo descartaría el efecto nocivo de una variante (119). Por lo cual se considera que esta herramienta por si sola pudiera usarse durante el flujo de análisis de datos de secuenciación para la identificación de variantes de riesgo en *DPYD*.

Los hallazgos reportados en el estudio realizado por Shrestha y colaboradores fueron similares para el análisis de predictores *in silico* en el gen *DPYD*, donde reportaron que las herramientas con el mejor rendimiento fueron PolyPhen-2 (AUC:0.79, ACC: 0.72, MCC:0.42), SIFT (AUC:0.76, ACC:0.73, MCC:0.43) y PROVEAN (AUC:0.77, ACC:0.72, MCC:0.41) (91); para el caso de los datos presentados en esta investigación los valores fueron cercanos SIFT (AUC:0.75, ACC:0.644, MCC:0.346) y PROVEAN (AUC:0.82, ACC:0.632, MCC:0.395), para esta investigación no se incluyó PolyPhen-2 dado el bajo rendimiento reportado en otras investigaciones. Donde en conjunto ambos datos podrían apoyar el uso de PROVEAN como herramienta de tamizaje, a pesar de que para Shrestha y colaboradores la herramienta con mejor rendimiento fue la desarrollada por ellos, sin embargo, no se dispuso de esta para su evaluación en esta investigación.

Una de las dificultades frente al análisis de las variantes en el gen *DPYD* relacionadas con toxicidad con el uso de las fluoropirimidinas, es separar las variantes asociadas con la timina-uracilluria de las que se relacionan con la toxicidad con el uso de fármacos, donde para el segundo grupo existe la presencia de una actividad residual que solo genera un defecto en la vía metabólica ante la exposición del 5-FU y sus derivados. Para la deficiencia completa de DPD, se ha reportado que esta se deriva principalmente de deleciones que afectan los residuos 581-635 por variantes frameshift o las variantes missense relacionados con C29R-R886H que conlleva la perdida completa de la actividad residual enzimática (120). Por lo cual el análisis para cada caso difiere en el sentido de que para el tamizaje para el ajuste de dosis se buscan variantes en las cuales de forma general se conserva la función biológica.

Por lo anterior el análisis estructural puede ser clave en la identificación de variantes de potencial interés en farmacología ya que a diferencia de variantes nocivas que se relacionan con fenotipos mórbidos, para el caso de las variantes accionables en farmacología, estas no están en dominios relevantes o regiones críticas para la función del producto del gen. Durante el análisis estructural se encontró que las variantes nocivas principalmente tenían cambios a nivel de las interacciones del residuo afectado, lo cual se relacionaba con cambios a nivel de la conformación de la cadena de aminoácidos cambiando los ángulos de interacción a nivel de cada aminoácido, generando alertas de colisión y cambios en la estructura y volumen de la proteína asociado a los cambios a nivel de las regiones de unión a ligandos, lo cual es similar a lo reportado en la literatura donde se ha encontrado que variantes conocidas como nocivas cambian los sitios de unión de la enzima afectando la función biológica de forma parcial al limitar la unión con ligandos y alterando el volumen de las cavidades de unión (54).

En estudios como el realizado por Shrestha y colaboradores se encontró que las variantes nocivas analizadas por ellos, estas estaban cercanas a sitios de unión a substratos principalmente en el sitio de unión FAD, afectando la función de la proteína; que podría estar asociado con lo documentado en los datos presentados el estudio, en donde los hallazgos frente a la alteración de la formación de las cavidades y las interacciones de los residuos necesarias para la conformación de la enzima limita la interacción entre el substrato y el sitio de unión, sugiriendo una posible defecto en el intercambio de electrones

necesarios para llevar a cabo la transformación de la primidina, lo cual ha sido planteado previamente (54,91); pudiendo esto ser clave al momento de desarrollar herramientas específicas para el análisis de variantes en el gen.

Asociado a la evaluación de herramientas de anotación y el análisis descriptivo estructural para variantes en el gen *DPYD*, se realizó un rastreo de variantes reportados en PharmVar y PharmGKB en un conjunto de VCF disponibles en la unidad de analítica de datos de Biotegen S.A.S. Una de las variantes de mayor interés dada su alta frecuencia en población caucásica y clara relación con la aparición de toxicidad con el uso de las fluoropirimidinas es DPYD*2A (c.1905+1G>A). En los datos disponibles en gnomAD en su v2.1.1 esta variante ha sido reportada con una frecuencia alélica de 0.0056 en individuos europeos no finlandeses y de 0.0011 en individuos de origen latino (121). En un estudio previo en el que se analizó en una muestra de exomas realizados en población colombiana se encontró una frecuencia alélica de la variante DPYD*2A (rs3918290) es de 0.001 (21) siendo inferior a la que se encontró para este estudio que fue de 0.0020. Sin embargo, evidencia en conjunto la baja frecuencia de esta variante en población colombiana frente a la población caucásica norteamericana y europea, lo cual supone una limitante para su uso en la identificación de pacientes de riesgo, planteando al igual que en otras poblaciones latinoamericanas que el tamizaje de esta variante seria poco informativo, lo cual lo sugiere Farinango y colaboradores quienes no documentaron la presencia de esta variante en una muestra de individuos ecuatorianos (19).

En los datos analizados en este estudio, de las variantes recomendadas para el tamizaje previo inicio de quimioterapia con fluoropirimidinas (12,13), la que tuvo mayor frecuencia fue Hap3 (rs56038477), la cual para la muestra analizada fue de 0.0090, superior a la reportada por Silgado y colaboradores quienes reportaron una frecuencia alélica para esta variante de 0.003 (21). Para el caso de las variantes rs67376798 y rs55886062 (DPYD *13) también con recomendación de tamizaje, la primera tuvo una frecuencia de 0.0035 y la segunda no fue encontrada en la muestra. Es llamativo que en los resultados presentados fuera documentada la variante rs67376798, ya que en estudios previos ninguna de las dos variantes fue encontrada en población colombiana (21) y para otras poblaciones latinoamericanas la variante rs55886062 tampoco fue reportada y las variantes rs67376798 y rs56038477 se reportaron con una frecuencia baja (0.001) para el caso de población Ecuatoriana (19).

En estudios realizados en población Mexicana y de Brasil las variantes rs3918290 y rs55886062 no fueron reportadas (122). Estos hallazgos podrían indicar que las variantes accionables para el ajuste de dosis recomendadas para tamizar en población caucásica presentan una distribución diferente en los grupos poblacionales latinoamericanos incluyendo la población colombiana; a pesar que para el caso del conjunto de datos analizados en este estudio, para algunas variantes la frecuencia alélica fue mayor a lo encontrado por Silgado y colaboradores, quienes realizaron un estudio similar con muestras colombianas, en general las variantes tienen baja frecuencia y una explicación para la diferencia entre estudios puede ser el análisis de poblaciones diferentes. Lo cual es una hipótesis que no se pudo validar por la limitación del uso de los datos durante la investigación, ya que al no poder disponer de información geográfica frente al origen de las muestras, se redujo la posibilidad de realizar inferencias adicionales a nivel poblacional.

La variante con mayor frecuencia dentro de los datos analizados fue DPYD*5, con una frecuencia alélica de 0.2555, siendo similar a lo reportado en otros estudios en poblaciones latinoamericanos (19,21,123), sin embargo, esta variante hasta el momento la evidencia sugiere que no genera un cambio en la actividad enzimática, por lo cual no se encuentra recomendada para el tamizaje previo inicio de tratamiento, considerándola una variante poblacional (124). Los datos evidenciados sugieren un comportamiento diferente de las frecuencias poblacionales de variantes que han sido ampliamente estudiadas en Europa y Norteamérica. Se podría plantear que a diferencia de estas poblaciones, la carga de toxicidad relacionada con fluoropirimidinas para Latinoamérica y Colombia podría ser explicado por otras variantes, por lo cual se hace imperativa la necesidad de realizar estudios con grandes grupos poblacionales y una caracterización demográfica completa, en la cual se puedan identificar las variantes presentes en el gen para cada población y así plantear la relación de estas con la toxicidad con el uso de las fluoropirimidinas.

En un acercamiento inicial en el que se buscaba establecer la presencia de nuevas variantes no reportadas previamente, se realizó una búsqueda en los datos analizados, donde se identificó una variante que no se encuentra reportada en PharmVar y PharmGKB. La variante documentada fue c.1127A>C (p.Glu376Ala) la cual en el conjunto de datos reportó una frecuencia de 0.0005 y para el caso de la base de datos poblacional de gnomAD esta variante no ha sido reportada (121). Durante el análisis *in silico* se encontró

que PROVEAN así como el resto de los algoritmos de anotación sugieren que la variante tendrá un efecto nocivo y en el análisis estructural se sugiere que esta variante conlleva a un cambio de aminoácido con un menor tamaño y una carga diferente que altera la formación de enlaces con estructuras cercanas, que podría alterar la estructura tridimensional de la proteína, sin embargo, no afectaría los dominios funcionales de la proteína. No se encontraron reportes en la literatura para esta variante durante su búsqueda en diferentes bases de datos (Pubmed y Google Scholar).

A pesar del uso de PROVEAN como el principal algoritmo de análisis de variantes de cambio de sentido en el gen *DPYD*, y la sugerencia de su uso dentro del tamizaje en la identificación de variantes de riesgo, el reporte de un efecto nocivo por esta herramienta solamente podría indicar la necesidad de evaluaciones adicionales, se considera que el análisis por medio de herramientas unidimensionales permite un tamizaje inicial, sin embargo es necesario el desarrollo de pruebas adicionales para la confirmación del efecto de la variante.

Para el análisis estructural, las variantes conocidas como dañinas, en la literatura se ha reportado que estas suelen estar cercanas a dominios relevantes para la actividad enzimática de la proteína, donde se ha encontrado que gran parte de estas variantes se encuentran localizadas cercanas a la unión con los substratos lo que finalmente conlleva a una alteración en la función durante eventos nocivos sin que la proteína pierda la función de forma completa (91), para el caso de la variante encontrada en los datos analizados no se encontró que esta afectara directamente los dominios funcionales de la proteína, pero se espera que altere la conformación tridimensional de la enzima lo cual podría alterar la unión con otras moléculas. Por lo cual se ve la necesidad de realizar estudios adicionales, donde desde lo computacional el desarrollo de modelos se evalué el efecto de la variante frente a la interacción con sus sustratos y los ligandos podrían ser una alternativa complementaria para el análisis de estas variantes.

Es importante avanzar en el uso de herramientas computacionales para el análisis en genes de interés en farmacología, ya que a pesar del progreso técnico derivado de la secuenciación en la práctica de la genética médica, que ha llevado a generalizar el uso de estas herramientas en la práctica del diagnóstico clínico, para el caso de la implementación clínica de la farmacogenómica basada en NGS el uso de estas alternativas está muy

rezagada, por lo cual pesar de la ventaja del genotipado basado en NGS que podría llevar al descubrimiento del panorama completo del genoma individual y así mismo de la variabilidad en consecuencias funcionales y recomendaciones clínicas útiles en farmacología, es poco lo que se ha desarrollado frente al uso de alternativas computacionales (125).

Para el caso de las enfermedades congénitas, la correlación entre las alteraciones fenotípicas y el análisis genómico permite determinar el efecto de las variantes de interés, sin embargo, los fenotipos farmacogenómicos son generalmente más difíciles de detectar ya que sólo se presentan ante la exposición a medicamentos específicos, por lo cual el análisis clínico y de laboratorio debe tener en cuenta estas consideraciones, lo cual limita su uso de forma generalizada por el costo económico, las necesidades técnicas, el tiempo requerido para la validación individual de cada análisis y las implicaciones éticas asociadas. Por lo cual, en ausencia de asociaciones de respuesta a fármacos o caracterizaciones experimentales que respalden la interpretación funcional de nuevas variantes en genes de interés en farmacología, las herramientas de predicción computacional podrían ser de utilidad para llenar este vacío en conocimiento, al disminuir el tiempo de análisis, el costo experimental y enfocar las validaciones de laboratorio a variantes potencialmente deletéreas (125)

Adicionalmente, para el caso de las variantes en farmacogenes, la variabilidad Inter experimental, los diferentes grados de severidad con fenotipos leves o moderados y las limitaciones derivadas de los análisis dicotómicos, ha llevado a que los ensayos experimentales deban ser rigurosos y se requieran múltiples replicas para poder generar suficiente evidencia funcional de utilidad clínica, lo cual en general puede no ser posible e implican desviaciones en los resultados. Para lo cual las herramientas computacionales han brindado una solución, donde con la generación de modelos robustos en los cuales se incluyen datos poblacionales, análisis funcionales y herramientas de predicción, se ha logrado generar predicciones con mayor exactitud de las consecuencias fenotípicas en farmacología asociadas a la variabilidad genética (85).

Por todo lo anterior, el enfoque diseñado en este estudio, en el cual se plantea el uso de herramientas computacionales, podría contribuir a fomentar el desarrollo de alternativas para la identificación de variantes que son de interés en farmacogenética en población

colombiana, donde a partir de la combinación del análisis computacional junto al análisis poblacional, suponga un fortalecimiento frente a la disponibilidad de información y la identificación de forma exhaustiva de variantes que podrían ser relevantes. En la literatura de forma generalizada se ha planteado el uso del tamizaje sea por genotipado o fenotipo, solo indicando el uso de una sola alternativa para la identificación de variantes de riesgo, por lo cual desde este estudio se considera necesario el cambio de paradigma donde se combinen diferentes herramientas de análisis, como también lo plantearon Pallet y colaboradores, donde se identifiquen de manera total los pacientes de riesgo, pudiendo realizar un genotipado y fenotipado en conjunto buscando impactar a nivel del salud disminuyendo el número de pacientes en riesgo con el uso de fluoropirimidinas, como lo sugiere Pallet y colaboradores en su investigación (24).

Por la inclusión de la secuenciación en el sistema de salud de Colombia, el análisis de genotipo podría ser de gran utilidad en la identificación de variantes de riesgo en el gen *DPYD*, por lo cual estos algoritmos computacionales son fundamentales en la identificación y tamizaje de los pacientes que van a ser expuestos a las fluoropirimidinas. A pesar de que la información disponible frente al uso de estas herramientas es amplia, para el caso de la farmacogenética la evidencia disponible es limitada, encontrando solamente algunas investigaciones para variantes puntuales.

Se plantea que al igual que como sucede en el análisis de variantes en condiciones monogénicas, las herramientas de análisis *In Silico* permitirían a partir de datos derivados de la secuenciación identificar variantes reconocidas como de riesgo y tamizar variantes potencialmente riesgosas con el uso de algoritmos de anotación unidimensional y el análisis estructural, lo cual podría disminuir la limitante que tiene el genotipado, el cual solo identifica las variantes que son incluidas dentro del experimento o conocidas. Con un análisis computacional exhaustivo, se podría plantear el tamizaje del fenotipo únicamente en pacientes en los que se encontraron variantes de riesgo, lo cual también permitirá el análisis de nuevas variantes y la clasificación final de los pacientes que, tamizando solo los casos en los cuales se plantee un mayor riesgo desde el genotipado.

Lo cual ya se ha descrito previamente, donde, se ha planteado que el análisis y la interpretación de la variabilidad farmacogenética y farmacogenómica usando herramientas computacionales puede ser un pilar para la implementación de terapias guiadas a partir de

datos de secuenciación, lo cual se apoya por la necesidad de evaluar de forma exhaustiva la asociación entre factores genéticos y la complejidad en los fenotipos toxicológicos y los efectos adversos relacionados con el uso de fármacos, siendo esto facilitado por la posibilidad de análisis multivariados y con la inclusión de variables de difícil reproducibilidad a nivel de laboratorio en los modelos computacionales (85), lo cual impactaría de forma positiva en salud pública, al desarrollar recomendaciones de acuerdo con las necesidades propias de cada población teniendo en cuenta la diversidad genética, los factores medio ambientales y las necesidad propias de salud, siendo ya ampliamente conocido, que con estas intervenciones se logra aumentar la efectividad y tolerabilidad con ajuste de dosis de acuerdo al genotipo, además, de lograr una disminución de los costos en salud asociados al manejo de reacciones adversas y cambios en la terapia (126)

Lo anteriormente mencionado se plasma en el protocolo de análisis de datos de secuenciación del gen *DPYD* propuesto, el cual se desarrolló a partir de la premisa de la utilidad del análisis computacional como alternativa de identificación de pacientes de riesgo y la facilidad de aplicación del análisis de genotipo dada su cobertura por el sistema de salud colombiano. A pesar de la limitante que plantea un análisis por algoritmos de anotación unidimensionales, se considera que el protocolo propuesto abre la puerta para continuar con la generación de información y desarrollo de herramientas que fortalezcan la posibilidad de un análisis mixto, lo cual se justifica en lo propuesto por Pallet y colaboradores (24), quienes indican que realizar solamente el análisis de fenotipo limita el esfuerzo médico al ajuste de dosis, mientras que correlacionar los hallazgos fenotípicos con el genotipo del paciente, no solamente implica un impacto en el tratamiento, sino que adicionalmente supone un análisis familiar del riesgo relacionado con condiciones mórbidas asociadas al gen *DPYD*. Por lo cual continuar con la investigación de otras herramientas computacionales junto con la validación del protocolo propuesto, podría ser una alternativa en el contexto colombiano para la aplicación de la farmacogenética.

Actualmente se ha ampliado el interés en la aplicación de las herramientas computacionales en la práctica clínica, donde no solamente se ha planteado su uso para el desarrollo de nuevos fármacos, sino que adicionalmente ha generado interés en el campo del desarrollo de nuevos biomarcadores, donde diversos estamentos como la FDA, ha planteado que con la integración de información clínica y multiomica, se podrían tener hallazgos con el suficiente nivel de evidencia para la aplicación directa en la practica

clínica, incluso sin una validación previa a nivel experimental (127,128), por lo cual plantear el análisis de variantes potencialmente nocivas en genes de interés en farmacología por medio de estas herramientas, podría verse beneficiado por el aumento de herramientas de uso clínico, adicional al impacto económico en salud, donde por ejemplo se ha planteado que la implementación generalizada de herramientas basadas en inteligencia artificial para el diagnóstico médico y la investigación en salud podría reducir hasta un 10% de los costos totales en salud (129).

En Colombia, a pesar de la poca disponibilidad de información para el inicio de la aplicación de las recomendaciones frente al tamizaje de genotipo previo inicio del tratamiento con fluoropirimidinas, se dispone de las herramientas para la secuenciación y análisis del gen *DPYD*, por lo cual, dado el contexto colombiano frente a la posibilidad del análisis molecular incluido en el plan de beneficios en salud, se plantea que al realizar un uso generalizado de las herramientas computacionales para la identificación de pacientes en riesgo, podría permitir la aplicación de la farmacogenética de forma generalizada, siendo esto relevante por las consecuencias no solamente para cada individuo al reducir para el caso particular de las fluoropirimidinas el impacto social y carga de salud mental asociados a la aparición de los efectos adversos asociados al uso del medicamento y a nivel económico la reducción en costos en salud al reducir los gastos directos e indirectos relacionados con el tratamiento de eventos adversos (56), adicionalmente también el aumento de evidencia clínica y la generación de datos poblacionales permitiría tomar decisiones en salud publica frente al desarrollo de protocolos acordes a las necesidades poblacionales, como ya ha sido descrito dentro de los beneficios de la aplicación de farmacogenética en salud pública (130).

Por lo cual es imperativo el desarrollo de alternativas para la aplicación de la medicina de precisión y la farmacogenética en Colombia, donde se permita el desarrollo de bases de datos completas incluyendo no solamente información clínica sino también información con fenotipos, genotipos y datos poblacionales, lo cual permitirá la reducción del riesgo derivado de las intervenciones en salud, este trabajo planteó un primer análisis poblacional y computacional, en el cual se evidencia la necesidad de aumentar la información poblacional frente a la presencia de variantes ya conocidas y las alternativas frente a la identificación de nueva información, por lo cual se deberá continuar con la investigación

en farmacogenética y la generación de evidencia para el desarrollo de guías de práctica clínica acorde a las características de la diversidad poblacional colombiana.

Conclusiones

Actualmente gran parte de la información disponible acerca de las frecuencias alélicas de diferentes variantes en el gen *DPYD* es derivada principalmente de población caucásica, por lo cual se ha planteado la duda frente al uso de esta información y guías de manejo en la población latinoamericana, dada la diversidad existente por la confluencia de diversos grupos poblacionales con diferentes ancestrías que supondría una distribución diferente en las frecuencias alélicas de las variantes encontradas en caucásicos, por lo cual se implementó un análisis poblacional y computacional en una muestra de datos derivados de secuenciación de exoma completo de pacientes colombianos, con el fin de realizar un acercamiento a la realidad colombiana.

Como primer paso de este acercamiento, se realizó una búsqueda en la literatura bajo el modelo de una revisión sistemática rápida para la selección de los algoritmos con mejor rendimiento para la anotación de variantes de cambio de sentido. En general, tanto las revisiones sistemáticas como los meta-análisis ayudan a condensar la evidencia disponible en la literatura frente a un tema y extraer información de calidad para la toma de decisiones, sin embargo, para el análisis realizado se encontró que la presencia de diversos tipos de análisis comparativos y la gran diversidad frente a la decisión de las pruebas Gold Standard usadas durante la evaluación de rendimiento de los algoritmos de anotación son un limitante importante para la generación de recomendaciones, principalmente asociado a la aparición de los posibles sesgos derivados por la falta de homogeneidad en los estudios. Pero, a pesar de las dificultades que se documentaron durante el desarrollo de la revisión sistemática, realizar este tipo de análisis no solamente plantea un diferencial por las implicaciones en la selección de herramientas con evidencia disponible frente a su rendimiento diagnóstico para el análisis comparativo, sino que adicionalmente permitió evidenciar las falencias existentes, así como la necesidad de generar conocimiento frente alternativas para el desarrollo de herramientas Gold Standard que permitan la comparación de diversos algoritmos de anotación al momento de establecer el impacto de la variación en el genoma humano.

Y para el caso de genes de interés en farmacogenética es un reto aun mayor, ya que al igual que para otro tipo de genes la información entre estudios no es homogénea y

adicionalmente se dispone de un número menor de estudios en los que se evalúe de forma comparativa las herramientas de anotación y el posible uso de estas herramientas en estos genes, por lo cual el estudio desarrollado es de gran importancia, ya que permitió generar información acerca de la utilidad real de estos algoritmos. A partir de la búsqueda en la literatura se seleccionaron las herramientas PROVEAN, SIFT, PhD-SNP, SNP&Go, MutationAssessor, Eigen, REVEL, BayesDel y META-SNP, las cuales fueron reportadas en diferentes estudios con mayor rendimiento, sin embargo, este hallazgo fue principalmente en genes relacionados con enfermedades monogénicas.

Dentro de los objetivos del estudio se planteó identificar la herramienta de anotación con mayor rendimiento para la clasificación de las variantes en el gen *DPYD*, a partir de los datos analizados se encontró que PROVEAN fue el algoritmo con mejor desempeño (alta sensibilidad, exactitud y valor AUC), sin embargo, al igual que las otras herramientas analizadas fue mayor la sensibilidad reportada frente a las otras métricas de rendimiento, por lo cual su uso a nivel clínico se puede ver limitado por la dificultad para excluir principalmente variantes verdaderamente neutras, lo cual no solamente se evidenció para PROVEAN sino también para todas las herramientas analizadas. Pero en forma general esta herramienta podría ser de utilidad en el flujo de identificación inicial de variantes nocivas, sin embargo, para la toma de decisiones clínicas es necesario profundizar en estudios adicionales con los cuales se puedan confirmar los hallazgos anotados in silico.

Asociado al análisis *in silico* unidimensional, se realizó un análisis descriptivo estructural del impacto de las variantes de cambio de sentido en la proteína DPD, encontrando que en pocos casos con el análisis estructural básico se logra separar claramente entre una variante nociva vs una variante neutra, lo cual plantea la posibilidad que el impacto de estas variantes pueda relacionarse principalmente con cambios en la interacción de la proteína con otras moléculas, lo cual limite la función ante eventos nocivos como el uso de fármacos y no de forma global. Mostrando de esta forma la necesidad de ampliar la información acerca de los focos de investigación para el desarrollo de herramientas computacionales de predicción en genes de interés en farmacogenética. Ya que como se evidenció durante el análisis del conjunto de VCF tomados del banco de datos de Biotecgen S.A.S., donde a pesar de identificar la variante c.1127A>C (p.Glu376Ala) en la cual los algoritmos de anotación y de análisis estructural sugirieron un efecto nocivo, se

vio limitada la asignación de la clasificación por la ausencia de información frente a la interacción de la proteína con otras.

Como parte del acercamiento poblacional de la frecuencia de variantes de interés en población colombiana, se realizó una búsqueda de las SNV de interés en una muestra de datos derivados de secuenciación de exoma completa, donde se encontró que las variantes que son recomendadas para el tamizaje en las guías europeas (c.1905+1G>A, c.1679T>G, c.2846A>T y c.1129–5923C>G) tiene una baja frecuencia en la población de estudio frente a la población caucásica, por lo cual asignar toda la carga de riesgo de toxicidad a estas variantes para la población colombiana podría ser equivocado. Sin embargo, en los resultados no se encontró ninguna otra variante conocida como de riesgo con una alta frecuencia lo cual podría sugerir que pueden haber variantes nocivas no identificadas en otras poblaciones presentes en individuos colombianos o un mayor número de variantes nocivas que suponga la necesidad de tamizaje de un número mayor de variantes, por lo cual para el desarrollo de recomendaciones frente a la aplicación del tamizaje y generación de guías se requiere un estudio de mayor envergadura donde se analicen individuos de diferentes regiones y se plantee un análisis fenotípico adicional, lo cual se constituye como una gran limitante de este estudio donde no se tuvo en cuenta el fenotipo y no se accedieron a los datos demográficos del origen de las muestras.

A pesar de las limitaciones del estudio, se considera que este permitió evidenciar las limitaciones de aplicar la información disponible para el gen DPYD y las guías de ajuste de dosis por genotipo, al evidenciar las diferencias entre las frecuencias alélicas de las variantes disponibles en bases de datos sugiere que un tamizaje basado en las recomendaciones actuales puede ser insuficiente para la detección de pacientes de riesgo. Como aspecto positivo se abre la puerta para continuar la investigación en farmacogenética, evidenciando las necesidades para el desarrollo de trabajos de mayor envergadura donde se analice una cohorte mayor de individuos, que permita generar una base de datos poblacional colombiana. Adicionalmente, plantea a las herramientas computacionales como una alternativa en la identificación de variantes de riesgo, lo cual puede ir de la mano con el crecimiento en la generación de datos de secuenciación del ADN que se ha ido dando en Colombia, por lo cual aprovechar la capacidad computacional podría abrir una puerta para el desarrollo de la farmacogenética a nivel nacional.

Finalmente, como se mencionó durante la discusión se debe plantear la posibilidad de un análisis escalonado para identificación de pacientes con mayor riesgo de toxicidad, donde tal vez el primer paso sea un análisis genético exhaustivo del gen con el desarrollo de herramientas computacionales con alto rendimiento para posteriormente realizar un análisis fenotípico bioquímico, para lo cual se deberá continuar con investigaciones no solamente desde lo clínico donde se integren estas herramientas, sino que también realicen una evaluación del impacto en salud pública y del impacto económico a largo plazo que puedan tener este tipo de intervenciones.

Recomendaciones y limitaciones

Este estudio se planteó la necesidad de desarrollar una alternativa para la identificación y análisis de variantes en el gen de *DPYD*, lo cual se propone por la poca información disponible acerca de la frecuencia alélica de variantes de interés en el gen, la heterogeneidad poblacional colombiana y la necesidad de iniciar la aplicación de medicina personalizada y de precisión. En los hallazgos reportados en este trabajo se encontró la dificultad que plantea el análisis de SNV de interés en farmacogenética a partir de herramientas computacionales, donde a pesar del desarrollo de diferentes alternativas para el análisis *In Silico* de variantes de tipo cambio de sentido, estas se han desarrollado principalmente para la evaluación de condiciones monogénicas y son pocos los estudios que se han realizado para genes de interés en farmacogenética. Por lo cual, su uso en SNV que principalmente tienen un rol como biomarcador se ve limitado por la falta de evidencia y las limitaciones técnicas de estas herramientas por sus particularidades. Por lo cual se debe plantear la necesidad de estudios en los cuales se desarrollen nuevas alternativas en los cuales se tenga en cuenta las particularidades biológicas de los farmacogenes y sus variantes de interés

A pesar de que la investigación realizada plantea el uso de herramientas computacionales como una alternativa para la evaluación de variantes en el gen *DPYD*, se ve limitado por la inclusión en el análisis comparativo solamente de herramientas unidimensionales, excluyendo dentro del análisis de rendimiento, los hallazgos documentados durante la evaluación del impacto estructural de las variantes analizadas, pudiendo esto ser relevante para mejorar la capacidad predictiva de los algoritmos evaluados. Adicionalmente, se evaluó de forma individual cada una de las herramientas planteadas, sin incluir la evaluación conjunta de las herramientas, lo cual también podría ser relevante al momento de plantear algoritmos con mayor exactitud y precisión en la identificación de variantes nocivas.

Por lo cual, se considera que se deberá continuar la investigación en dirección de plantear la posibilidad de estudios computacionales en los que el análisis desarrollado incluya el desarrollo de estrategias basadas en modelos en biología de sistemas usando información física molecular y bioquímica de la interacción entre la enzima, el sustrato y sus ligandos,

con lo cual se dé un análisis exhaustivo del efecto de una variante, donde se detalle la diferencia entre variantes nocivas y benignas con lo cual no solamente se podría ampliar la información acerca de variantes en genes de interés en farmacología, sino que adicionalmente se podría impactar en el análisis de cambios relacionados con enfermedades monogénicas.

El análisis computacional bajo el contexto colombiano supone una ventaja frente a otros países de la región, dada la inclusión de los estudios moleculares tipo exoma y genoma en el sistema de salud y la existencia de bancos de datos en los laboratorios de biología molecular con información del genoma de individuos colombianos, por lo que se plantea que se deberá direccionar la investigación no solamente al desarrollo de capacidad técnica y tecnológica, sino a fortalecer la integración de estos datos para aumentar la información disponible frente a datos poblacionales de acuerdo a las particularidades de las regiones colombianas.

A pesar de incluir datos poblacionales dentro del estudio, existe una limitación derivada de la ausencia de datos demográficos relacionados con el origen de los individuos de las muestras incluidas, lo cual se derivo de las limitaciones frente al uso de los datos y las consideraciones éticas asociadas a las limitaciones del consentimiento informado usado en Biotecgen S.A.S. Por lo cual, a pesar de que se aporta información relevante frente a las frecuencias de variantes de interés en el gen *DPYD*, existen limitaciones claras para un análisis poblacional estructurado donde se pueda generar más información frente a la distribución de las variantes a nivel geográfico, análisis de estructura poblacional y posibles recomendaciones dentro del análisis genético poblacional.

Adicionalmente, dadas las características de los datos usados, se debe anotar que puede existir un sesgo frente a la identificación de variantes en regiones no codificantes del gen, ya que, al usar datos de secuenciación de exoma completo, se limita la cobertura y profundidad a nivel de regiones intrónicas. Siendo esto importante para futuras investigaciones, en las cuales se hace relevante realizar la secuenciación completa del gen, para así poder identificar otro tipo de variantes, no incluidas en el presente estudio.

Sin embargo, se debe mencionar a pesar de las limitaciones del estudio, el abordaje planteado es apoyado por un número mayor de revisiones e investigaciones originales, en

los cuales los abordajes a partir del análisis de datos multiomicos, el modelado computacional con datos experimentales de función, estructural y propiedades físico-químicas ha llevado a la reducción de los tiempos de análisis, así, como de los costos de investigación, al permitir disminuir los potenciales biomarcadores y así enfocar el análisis de laboratorio a un grupo pequeño de moléculas, lo cual también puede ser llevado al campo del análisis de la variación en el genoma humano al desarrollar predictores específicos y con un alto rendimiento diagnóstico. Por lo cual las investigaciones encaminadas a la aplicación de este tipo de herramientas a nivel de farmacogenética, no solamente permitiría la generación de evidencia para la toma de decisiones en salud, sino que también, permitirá innovar con el desarrollo de herramientas de uso clínico y de metodologías de análisis de datos de secuenciación del genoma humano.

Por lo cual, el desarrollo de investigación en el campo de la tecnología y de la computación en salud no solamente impactaría a nivel de salud pública al permitir avanzar en la obtención de evidencia frente a la aplicación de medidas en farmacogenética, sino que adicionalmente podría llevar a ampliar la capacidad técnica y tecnológica del país, al generar no solamente conocimiento, sino ampliar la capacidad para resolver problemas y desarrollar herramientas para solucionar problemas en salud, generando una oportunidad para el desarrollo de productos con miras comerciales.

Finalmente, a pesar del progreso en la investigación en el país en farmacogenética y el mayor interés por las implicaciones en salud pública que tienen este tipo de estudios, se sugiere la necesidad de desarrollar alternativas de análisis y generación de recomendaciones a partir de las investigaciones desarrolladas en el campo de la biología computacional aplicada a la medicina, ya que como se planteó en esta investigación, herramientas como revisiones sistemáticas y meta-análisis como están actualmente planteados, pueden verse limitados al momento de extraer recomendaciones para herramientas de uso computacional, ya que por la heterogeneidad en la función biológica de los productos de los genes, la disparidad entre las pruebas de referencia y métodos experimentales, el análisis estadístico puede encontrarse con muchas limitaciones y sesgos que deberán ser corregidos al momento del análisis experimental y desarrollo de guías.

Anexo A: Instrumento de evaluación de la calidad de estudios de precisión diagnóstica QUADAS-2

Dominio 1: Selección de pacientes	
Pregunta original	**Pregunta ajustada**
1.1 ¿Se enroló una muestra consecutiva o aleatoria de pacientes?	1.1 ¿Se seleccionaron variantes con suficiente evidencia clínica o biológica de patogenicidad o benignidad?
1.2 ¿Se evitó un diseño de casos y controles	1.2 ¿Se evitó un diseño de casos y controles?
1.3 ¿Se evitaron exclusiones inapropiadas?	1.3 ¿Se eliminaron variantes de forma inadecuada o por no disponibilidad de algún algoritmo de anotación para la variante?
Dominio 2: Prueba índice	
2.1 ¿Fueron interpretados los resultados de la prueba índice sin conocimiento de los resultados de la de referencia?	2.1 ¿La interpretación del resultado del algoritmo de anotación se interpretó sin conocimiento de los resultados de la de referencia?[1]
2.2 Si se utilizó un umbral para definir la positividad o la negatividad de la prueba índice, ¿fue especificado previamente?	2.2 Se estableció en la metodología el uso de los puntos de corte para la interpretación del efecto de cada variante en el producto del gen (Se evaluará la aceptación de los puntos de cortes establecidos por los desarrolladores de cada herramienta o la indicación de la adopción de otro punto de corte sugerido por algún otro autor con su respectiva referencia)
Dominio 3: Prueba de referencia	
3.1 ¿Es probable que la prueba de referencia valore correctamente la condición diana?	3.1 Para cada variante incluida en el ensayo se evaluó la disponibilidad de estudios in vitro o clínicos donde se evidenciará la suficiente evidencia de causalidad entre la variante y la condición de interes[2].
3.2 ¿Fueron interpretados los resultados de la prueba de referencia sin conocimiento de los resultados de la prueba índice?	3.2 ¿La interpretación del resultado de la prueba de referencia se interpretó sin conocimiento de los resultados de la de referencia?[3]
Dominio 4: Flujo y tiempos	
4.1 ¿Hubo un intervalo apropiado entre la prueba índice y la prueba de referencia?	4.1 No aplica[4].
4.2 ¿Fue aplicada en todos los individuos la misma prueba de referencia?	4.2 ¿Se aplicaron todos los algoritmos de anotación a todas las variantes seleccionadas?
4.3 ¿Fueron incluidos todos los pacientes en el análisis?	4.3 ¿Fueron incluidas todas las variantes seleccionadas en el análisis? ¿Se realizo alguna exclusión inapropiada?[5]
Aplicabilidad	Para cada dominio se deberá evaluar la aplicabilidad de la investigación con la pregunta de la revisión sistemática, donde se informe si la aplicabilidad es baja, alta o incierta frente al tipo de prueba de referencia, prueba índice y población de interés.

[1]Se **DEBERA** responder en esta pregunta **UNCLEAR** para los casos en los cuales no se informe sobre el desconocimiento de los investigadores del resultado de la prueba de referencia al momento de interpretar el resultado de la prueba índice. En los casos en los que se usen herramientas *in silico* validadas previamente o desarrollados por autores externos a la investigación y en el artículo se establezca que se usaran puntos de corte recomendados por otros autores o por los desarrolladores de la herramienta, no se considerara como posible riesgo de sesgo el conocimiento por parte de los investigadores del resultado de la prueba de referencia al momento de interpretar los resultados de las pruebas índice. En los casos en los que se incluyan herramientas de evaluación in silico desarrollada por los investigadores del articulo original las cuales fueron evaluadas durante la investigación se considerara un **ALTO RIESGO** de sesgo el conocimiento del resultado de la prueba de referencia al momento de interpretar la prueba índice.

[2]Se considerara **BAJO RIESGO** de sesgo que las variantes incluidas en el ensayo no cuenten con el mismo estudio de referencia (estudio clínico, análisis *in vitro*, *in vivo* o ambos), con la salvedad que la asignación del efecto de cada variante en la proteína (deletéreo o neutro) para cada caso cuente con suficiente evidencia que justifique el efecto de la misma. Esta decisión basada en la dificultad del análisis experimental para la anotación del efecto de las variantes genéticas.

[3]Esta pregunta solo **DEBERA** responder en los casos en los que en el mismo ensayo se realicen estudios de análisis experimental o clínico de las variantes, en investigaciones en las cuales se incluyan variantes que cuenten con estudios funcionales *in vitro*, *in vivo* o clínicos realizados previamente se deberá asignar un **NO APLICA/ UNCLEAR**.

[4]Se estableció que esta pregunta no aplicaba; esta decisión se basa en que los análisis experimentales del efecto de una variante de cambio de sentido son las pruebas de referencia al momento de establecer la causalidad entre la presencia de una variante y la aparición de una enfermedad, sin embargo, la dificultad y costo de realizar estos estudios plantea una dificultad al momento de realizar análisis de precisión de algoritmos de anotación, por lo cual en pruebas de precisión se acepta la inclusión de variantes con datos experimentales reportados previamente y se compara el resultado curado previamente vs el resultado del algoritmo de anotación. Adicionalmente el efecto de una variante desde el punto de vista funcional no progresa ni se agudiza en el tiempo, por lo cual la evaluación entre la prueba de referencia y un algoritmo de anotación con un intervalo de tiempo prolongado *a priori* no implicaría un sesgo derivado del cambio de la enfermedad que se evidencian en otros test diagnósticos como pruebas serológicas o imagenológicas.

[5]Se acepta como exclusiones apropiadas aquellas donde se informa la exclusión de una variante por evidencia no disponible frente a su efecto en estudios funcionales. En los casos en los que un algoritmo de anotación no se encuentre disponible para una variante se podrá aceptar la exclusión de una variante, sin embargo, se deberá realizar la anotación en el riesgo de sesgo, se considerara solamente sin riesgo cuando se excluya la herramienta de análisis si no está disponible para una variante o se realice un análisis ajustado para esa herramienta teniendo en cuenta el cambio en el número de variante analizadas.

Anexo B: Variantes usadas para la evaluación comparativa

Descripción de variante						Efecto PharmVar	Evidencia	Efecto estudios	% Perdida de función	Nivel de evidencia
ID	dbSNP	GRCh38	GRCh37	SNV	proteína					
	rs376273539	1:97920887	1:98386443	G/C	I12M	No asignado	Shrestha. et al 2019	Función conservada	-	PS3/BS3_Moderate
	rs367619008	1:97828160	1:98293716	T/C	K63E	No asignado	Shrestha. et al 2019	Función disminuida	<15%	PS3/BS3_Moderate
	rs527580106	1:97828118	1:98293674	T/C	M77V	No asignado	Shrestha. et al 2019	Función disminuida	15-35%	PS3/BS3_Moderate
	rs528152707	1:97740453	1:98206009	C/A	C87F	No asignado	Shrestha. et al 2019	Función disminuida	15-35%	PS3/BS3_Moderate
	rs375990187	1:97740423	1:98205979	A/G	I97T	No asignado	Shrestha. et al 2019	Función conservada	-	PS3/BS3_Moderate
	rs370615432	1:97721651	1:98187207	C/A	K114N	No asignado	Shrestha. et al 2019	Función conservada	-	PS3/BS3_Moderate
	rs377169736	1:97721648	1:98187204	C/G	M115I	No asignado	Shrestha. et al 2019	Función conservada	-	PS3/BS3_Moderate
	rs538336580	1:97721599	1:98187155	T/A	T132S	No asignado	Shrestha. et al 2019	Función disminuida	35-70%	PS3/BS3_Moderate
	rs536577604	1:97721526	1:98187082	T/C	Q156R	No asignado	Shrestha. et al 2019	Función disminuida	15-35%	PS3/BS3_Moderate
	rs368970772	1:97699542	1:98165098	G/T	F163L	No asignado	Shrestha. et al 2019	Función conservada	-	PS3/BS3_Moderate
	rs374827081	1:97699510	1:98165066	G/C	P174R	No asignado	Shrestha. et al 2019	Función conservada	-	PS3/BS3_Moderate
	rs371792178	1:97699507	1:98165063	G/A	S175L	No asignado	Shrestha. et al 2019	Función conservada	-	PS3/BS3_Moderate
	rs779728902	1:97699486	1:98165042	A/T	M182K	No asignado	Shrestha. et al 2019	Función conservada	-	PS3/BS3_Moderate
	rs368152149	1:97699463	1:98165019	T/C	I190V	No asignado	Shrestha. et al 2019	Función conservada	-	PS3/BS3_Moderate
	rs376073289	1:97699408	1:98164964	C/T	R208Q	No asignado	Shrestha. et al 2019	Función disminuida	<15%	PS3/BS3_Moderate
	rs781184141	1:97699375	1:98164931	T/C	K219R	No asignado	Shrestha. et al 2019	Función conservada	-	PS3/BS3_Moderate
	rs376128878	1:97691740	1:98157296	G/T	L247I	No asignado	Shrestha. et al 2019	Función conservada	-	PS3/BS3_Moderate

	rs141726921	1:97679173	1:98144729	C/T	G258S	No asignado	Shrestha. et al 2019	Función conservada	-	PS3/BS3_Moderate
	rs369103276	1:97679103	1:98144659	A/G	I281T	No asignado	Shrestha. et al 2020	Función disminuida	<15%	PS3/BS3_Moderate
	rs143879757	1:97595124	1:98060680	G/T	T298K	No asignado	Shrestha. et al 2019	Función conservada	-	PS3/BS3_Moderate
	rs575763449	1:97593369	1:98058925	G/A	S326F	No asignado	Shrestha. et al 2019	Función conservada	-	PS3/BS3_Moderate
	rs377143350	1:97593276	1:98058832	C/T	R357H	No asignado	Shrestha. et al 2019	Función conservada	-	PS3/BS3_Moderate
	rs573299212	1:97593273	1:98058829	C/T	R358H	No asignado	Shrestha. et al 2019	Función conservada	-	PS3/BS3_Moderate
	rs201785202	1:97593226	1:98058782	G/A	P374S	No asignado	Shrestha. et al 2019	Función conservada	-	PS3/BS3_Moderate
	rs144935781	1:97573970	1:98039526	T/C	M377V	No asignado	Shrestha. et al 2019	Función conservada	-	PS3/BS3_Moderate
	rs201648613	1:97573967	1:98039523	C/G	E378Q	No asignado	Shrestha. et al 2019	Función conservada	-	PS3/BS3_Moderate
	rs528430685	1:97573871	1:98039427	G/A	R410W	No asignado	Shrestha. et al 2019	Función conservada	-	PS3/BS3_Moderate
	rs199646142	1:97573870	1:98039426	C/T	R410Q	No asignado	Shrestha. et al 2019	Función conservada	-	PS3/BS3_Moderate
	rs532341730	1:97573827	1:98039383	A/T	D424E	No asignado	Shrestha. et al 2019	Función conservada	-	PS3/BS3_Moderate
	rs200693895	1:97573819	1:98039375	A/G	V427A	No asignado	Shrestha. et al 2019	Función conservada	-	PS3/BS3_Moderate
	rs370569731	1:97573792	1:98039348	C/G	S436T	No asignado	Shrestha. et al 2019	Función conservada	-	PS3/BS3_Moderate
	rs200709381	1:97573762	1:98039318	T/G	K446T	No asignado	Shrestha. et al 2019	Función conservada	-	PS3/BS3_Moderate
	rs370707404	1:97549693	1:98015249	A/G	V464A	No asignado	Shrestha. et al 2019	Función conservada	-	PS3/BS3_Moderate
	rs141439344	1:97549637	1:98015193	C/T	V483I	No asignado	Shrestha. et al 2019	Función conservada	-	PS3/BS3_Moderate
	rs140039091	1:97515821	1:97981377	C/G	A549P	No asignado	Shrestha. et al 2019	Función disminuida	<15%	PS3/BS3_Moderate
	rs55886062	1:97515787	1:97981343	A/T	I560N	No asignado	Shrestha. et al 2019	Función disminuida	15-35%	PS3/BS3_Moderate
	rs558354142	1:97515748	1:97981304	G/A	T573I	No asignado	Shrestha. et al 2019	Función disminuida	15-35%	PS3/BS3_Moderate
	rs374527058	1:97450207	1:97915763	A/G	V586A	No asignado	Shrestha. et al 2019	Función disminuida	35-70%	PS3/BS3_Moderate
	rs759249769	1:97450183	1:97915739	G/T	T594N	No asignado	Shrestha. et al 2019	Función conservada	-	PS3/BS3_Moderate

	rs371258350	1:97450156	1:97915712	C/T	G603E	No asignado	Shrestha. et al 2019	Función disminuida	35-70%	PS3/BS3_Moderate
	rs368146607	1:97450118	1:97915674	T/G	K616Q	No asignado	Shrestha. et al 2019	Función disminuida	35-70%	PS3/BS3_Moderate
	rs201433243	1:97450099	1:97915655	C/T	C622Y	No asignado	Shrestha. et al 2019	Función conservada	-	PS3/BS3_Moderate
	rs548783838	1:97382436	1:97847992	C/T	S644N	No asignado	Shrestha. et al 2019	Función disminuida	35-70%	PS3/BS3_Moderate
	rs529019871	1:97382404	1:97847960	T/C	K655E	No asignado	Shrestha. et al 2019	Función conservada	-	PS3/BS3_Moderate
	rs575853463	1:97373592	1:97839148	C/T	G676E	No asignado	Shrestha. et al 2019	Función disminuida	35-70%	PS3/BS3_Moderate
	rs187713395	1:97373580	1:97839136	A/G	M680T	No asignado	Shrestha. et al 2019	Función disminuida	<15%	PS3/BS3_Moderate
	rs202212118	1:97306285	1:97771841	C/A	V691L	No asignado	Shrestha. et al 2019	Función conservada	-	PS3/BS3_Moderate
	rs538703919	1:97306282	1:97771838	G/A	R692W	No asignado	Shrestha. et al 2019	Función disminuida	15-35%	PS3/BS3_Moderate
	rs375436137	1:97306281	1:97771837	C/T	R692Q	No asignado	Shrestha. et al 2019	Función conservada	-	PS3/BS3_Moderate
	rs569661196	1:97306215	1:97771771	A/G	V714A	No asignado	Shrestha. et al 2019	Función disminuida	35-70%	PS3/BS3_Moderate
	rs570122671	1:97305348	1:97770904	G/A	T737I	No asignado	Shrestha. et al 2019	Función disminuida	<15%	PS3/BS3_Moderate
	rs368327291	1:97305346	1:97770902	C/G	V738L	No asignado	Shrestha. et al 2019	Función conservada	-	PS3/BS3_Moderate
	rs568367673	1:97234994	1:97700550	C/A	G767V	No asignado	Shrestha. et al 2019	Función disminuida	funcioón	PS3/BS3_Moderate
	rs374825099	1:97234964	1:97700520	G/T	A777D	No asignado	Shrestha. et al 2019	Función disminuida	35-70%	PS3/BS3_Moderate
	rs758649719	1:97234935	1:97700491	C/T	G795R	No asignado	Shrestha. et al 2019	Función disminuida	<15%	PS3/BS3_Moderate
	rs371313778	1:97234860	1:97700416	C/T	V812I	No asignado	Shrestha. et al 2019	Función conservada	-	PS3/BS3_Moderate
	rs200687447	1:97193209	1:97658765	C/G	E828Q	No asignado	Shrestha. et al 2019	Función conservada	-	PS3/BS3_Moderate
	rs199777072	1:97193206	1:97658762	C/T	D829N	No asignado	Shrestha. et al 2019	Función conservada	-	PS3/BS3_Moderate
	rs372058915	1:97193164	1:97658720	T/C	I843V	No asignado	Shrestha. et al 2019	Función conservada	-	PS3/BS3_Moderate
	rs372909322	1:97193087	1:97658643	T/C	I868M	No asignado	Shrestha. et al 2019	Función conservada	-	PS3/BS3_Moderate
	rs368519011	1:97193071	1:97658627	T/C	K874E	No asignado	Shrestha. et al 2019	Función conservada	-	PS3/BS3_Moderate

	rs137878450	1:97082431	1:97547987	C/A	G936C	No asignado	Shrestha. et al 2019	Función conservada	-	PS3/BS3_Moderate
	rs372307932	1:97082394	1:97547950	A/T	I948N	No asignado	Shrestha. et al 2019	Función disminuida	<15%	PS3/BS3_Moderate
	rs201268750	1:97079120	1:97544676	G/T	H978Q	No asignado	Shrestha. et al 2019	Función disminuida	<15%	PS3/BS3_Moderate
	rs140989814	1:97079073	1:97544629	C/G	S994T	No asignado	Shrestha. et al 2019	Función disminuida	15-35%	PS3/BS3_Moderate
	rs151074666	1:97079056	1:97544612	C/T	D1000N	No asignado	Shrestha. et al 2019	Función conservada	-	PS3/BS3_Moderate
DPYD*13	rs55886062	1:97515787	1:97981343	A/C	I560S	No function	Offer. et al 2013	Función disminuida	25%	PS3/BS3_Moderate
DPYD*8	rs1801266	1:97691776	1:98157332	G/A	R235W	No function	Offer. et al 2014	Función disminuida	12.5-25%	PS3/BS3_Moderate
DPYD*10	rs1801268	1:97079071	1:97544627	C/A	V995F	No function	Offer. et al 2014	Función disminuida	<12.5%	PS3/BS3_Moderate
	rs55674432	1:97098616	1:97564172	C/A	G880V	No function	Offer. et al 2014	Función disminuida	<12.5%	PS3/BS3_Moderate
	rs59086055	1:97450190	1:97915746	G/A	R592W	No function	Offer. et al 2014	Función disminuida	<12.5%	PS3/BS3_Moderate
	rs72547601	1:97079121	1:97544677	T/C	H978R	No function	Offer. et al 2014	Función disminuida	<12.5%	PS3/BS3_Moderate
	rs72549304	1:97549609	1:98015165	G/A	S492L	No function	Offer. et al 2014	Función disminuida	12.5-25%	PS3/BS3_Moderate
	rs72549307	1:97699399	1:98164955	T/C	Y211C	No function	Offer. et al 2014	Función disminuida	12.5-25%	PS3/BS3_Moderate
	rs72549308	1:97699430	1:98164986	T/G	S201R	No function	Offer. et al 2014	Función disminuida	<12.5%	PS3/BS3_Moderate
	rs111858276	1:97549600	1:98015156	T/C	D495G	No function	Offer. et al 2014	Función disminuida	12.5-25%	PS3/BS3_Moderate
	rs138616379	1:97450189	1:97915745	C/T	R592Q	No function	Offer. et al 2014	Función disminuida	12.5-25%	PS3/BS3_Moderate
	rs141044036	1:97082365	1:97547921	T/C	K958E	No function	Offer. et al 2014	Función disminuida	<12.5%	PS3/BS3_Moderate
	rs143154602	1:97593289	1:98058845	G/A	R353C	No function	Offer. et al 2014	Función disminuida	0%	PS3/BS3_Moderate
	rs145773863	1:97450187	1:97915743	C/T	G593R	No function	Offer. et al 2014	Función disminuida	<12.5%	PS3/BS3_Moderate
	rs183385770	1:97593322	1:98058878	C/T	D342N	No function	Offer. et al 2014	Función disminuida	12.5-25%	PS3/BS3_Moderate
	rs67376798	1:97082391	1:97547947	T/A	D949V	Decreased function	Offer. et al 2014	Función disminuida	>25%	PS3/BS3_Moderate
	rs112766203	1:97305279	1:97770835	G/A	T760I	Decreased function	Offer. et al 2014	Función disminuida	>25%	PS3/BS3_Moderate

	rs115232898	1:97699474	1:98165030	T/C	Y186C	Decreased function	Offer. et al 2014	Función disminuida	>25%	PS3/BS3_Moderate
	rs146356975	1:97595149	1:98060705	T/C	K290E	Decreased function	Offer. et al 2014	Función disminuida	>25%	PS3/BS3_Moderate
	rs186169810	1:97573785	1:98039341	A/C	F438L	Decreased function	Offer. et al 2014	Función disminuida	>25%	PS3/BS3_Moderate
DPYD*4	rs1801158		1:97981421	C/T	S534N	Normal	Offer. et al 2013	Función conservada	-	PS3/BS3_Moderate
DPYD*5	rs1801159		1:97981395	T/C	I543V	Normal	Offer. et al 2013	Función conservada	-	PS3/BS3_Moderate
DPYD*6	rs1801160		1:97770920	C/T	V732I	Normal	Offer. et al 2013	Función conservada	-	PS3/BS3_Moderate
DPYD*9B	rs1801267		1:97564154	C/T	R886H	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate
	rs2297595		1:98165091	T/C	M166V	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate
	rs3918289		1:97915615	G/C	N635K	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate
	rs45589337		1:98144726	T/C	K259E	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate
	rs55971861		1:97848017	T/G	I636L	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate
	rs56005131		1:97700547	G/T	T768K	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate
	rs60139309		1:97658665	T/C	K861R	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate
	rs60511679		1:97770919	A/C	V732G	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate
	rs61622928		1:98039437	C/T	M406I	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate
	rs72547602		1:97544689	T/A	D974V	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate
	rs72549305		1:98058794	T/C	I370V	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate
DPYD*11	rs72549306		1:98058899	C/A	V335L	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate
	rs72975710		1:98015291	G/A	A450V	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate
	rs80081766		1:98348908	C/T	R21Q	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate
	rs114096998		1:97544543	G/T	P1023T	Normal	Offer. et al 2013	Función conservada	-	PS3/BS3_Moderate
	rs138391898		1:98015121	C/T	V507I	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate

	rs138545885		1:97839185	C/A	A664S	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate
	rs139459586		1:97544632	A/C	L993R	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate
	rs139834141		1:98165089	C/T	M166I	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate
	rs140114515		1:97544561	C/T	V1017I	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate
	rs140602333		1:98039475	G/A	R394W	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate
	rs141462178		1:98187206	T/C	M115V	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate
	rs142512579		1:98039361	C/T	D432N	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate
	rs142619737		1:97981407	C/T	G539R	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate
	rs143815742		1:98039474	C/A	R394L	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate
	rs144395748		1:98015282	G/C	P453R	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate
	rs145112791		1:98060639	G/A	L312F	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate
	rs145529148		1:97544695	T/C	Q972R	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate
	rs145548112		1:97771751	C/T	A721T	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate
	rs146529561		1:97770928	G/A	A729V	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate
	rs147545709		1:97564155	G/A	R886C	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate
	rs147601618		1:97915724	A/G	M599T	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate
	rs148799944		1:97544549	C/G	V1021L	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate
	rs148994843		1:97981479	C/T	V515I	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate
	rs150036960		1:98348924	G/C	L16V	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate
	rs150385342.1		1:98205956	C/T	A105T	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate
	rs150437414		1:98060644	A/G	L310S	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate
	rs190951787		1:97981445	G/C	T526S	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate

	rs199549923		1:98015237	G/T	T468N	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate
	rs199634007		1:97700514	G/T	T779N	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate
	rs200064537		1:98039395	A/T	N420K	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate
	rs200562975		1:98187098	T/C	N151D	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate
	rs200687447		1:97658765	C/T	E828K	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate
	rs201018345		1:98058935	C/T	A323T	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate
	rs201035051		1:97564188	T/G	K875Q	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate
	rs201615754		1:97981340	C/A	R561L	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate
	rs202144771		1:97544633	G/A	L993F	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate
	rs764666241		1:98039377	C>A	M426I	Normal	Offer. et al 2014	Función conservada	-	PS3/BS3_Moderate

Anexo C: Hallazgos durante el análisis estructural por Missense 3D-DB

ID	SNV	Proteína	Efecto funcional	Cambio estructural reportado en Missense 3D – DB
	C/A	C87F	Nociva	El residuo de tipo wildtype forma un enlace disulfuro con el residuo CYS 140 (Distancia: 2.528 Å) en la cadena de la estructura del tipo salvaje q12882_03-1h7x_b, y la sustitución interrumpe este enlace.
	T/C	M77V	Nociva	La sustitución conduce a la contracción del volumen de la cavidad en 195.48 Å^3.
	T/G	S201R	Nociva	Esta sustitución reemplaza un residuo no ionizado (SER, RSA 0.0%) con un residuo cargado (ARG).
	G/A	R353C	Nociva	Esta sustitución reemplaza un residuo cargado (ARG, RSA 0.4%) por un residuo no cargado (CYS).
DPYD*8	G/A	R235W	Nociva	Esta sustitución interrumpe un puente salino formado por el átomo NH2 del ARG 235 y el átomo OD1 del ASP 346 (distancia: 3.979 Å). El residuo del tipo wildtype tiene un área superficial relativa (RSA) del 8.4%.
	G/A	S492L	Nociva	Esta sustitución interrumpe todos los enlaces de hidrógeno entre cadenas laterales y/o enlaces de hidrógeno entre cadenas laterales y la cadena principal formados por un residuo SER (RSA 6.1%).
	T/C	Q156R	Nociva	Esta sustitución reemplaza un residuo no cargado (GLN, RSA 4.5%) al interior de la proteína, por un residuo cargado (ARG).
	C/G	A549P	Nociva	Esta sustitución introduce una prolina al interior de la proteína.
	A/T	I560N	Nociva	Esta sustitución reemplaza un residuo hidrofóbico al interior de la proteína (ILE, RSA 0.0%) con un residuo hidrofílico (ASN, RSA 0.6%).
	C/T	G795R	Nociva	Esta sustitución desencadena una alerta de colisión. El puntaje local de colisión para el wildtype es 55.20 y el puntaje local de colisión para el mutante es 75.51.
	C/A	G767V	Nociva	Esta sustitución desencadena una alerta de ángulos phi/psi no permitidos. Los ángulos phi/psi se encuentran en la región favorable para el residuo del tipo wildtype, pero en una región atípica para el residuo mutante. Esta sustitución reemplaza un residuo GLY al interior de la proteína (RSA 2.3%) por un residuo VAL también al interior de la proteína (RSA 2.1%)
	C/T	G603E	Nociva	La sustitución conduce a la contracción del volumen de la cavidad en 137.808 Å^3.
	C/T	G676E	Nociva	Esta sustitución desencadena una alerta de ángulos phi/psi no permitidos. Los ángulos phi/psi se encuentran en la región favorable para el residuo del tipo wildtype, pero en una región atípica para el residuo mutante.
	G/T	A777D	Nociva	Esta sustitución reemplaza un residuo hidrofóbico al interior de la proteína (ALA, RSA 0.9%) por un residuo hidrofílico (ASP, RSA 3.6%) y cargado (ASP).
	C/T	G593R	Nociva	Esta sustitución desencadena una alerta de choque. El puntaje de choque local para el tipo wildtype es de 37.61 y el puntaje de choque local para el mutante es de 61.91. Esta sustitución reemplaza un residuo GLY al interior de la proteína (RSA 7.1%) por un residuo ARG expuesto (RSA 24.1%).

	C/A	G880V	Nociva	Esta sustitución desencadena una alerta de phi/psi no permitida. Los ángulos phi/psi están en la región permitida para el residuo de tipo wildtype pero en la región atípica para el residuo mutante. Esta sustitución reemplaza un residuo GLY al interior de la proteína (RSA 5.9%) por un residuo VAL (RSA 3.5%).
	C/A	K114N	Neutra	Esta sustitución reemplaza un residuo cargado al interior de la proteína (LYS, RSA 7.8%) por un residuo no cargado (ASN), adicionalmente, esta sustitución interrumpe un puente salino formado por el átomo NZ de LYS 114 y el átomo OD1 de ASP 529 (distancia: 3.095 Å).
	C/T	D829N	Neutra	Esta sustitución reemplaza un residuo cargado al interior de la proteína (ASP, RSA 2.4%) por un residuo no cargado (ASN).
DPYD*9B	C/T	R886H	Neutra	Esta sustitución interrumpe todos los enlaces de hidrógeno de cadena lateral / cadena lateral y/o enlaces de hidrógeno de cadena lateral / cadena principal formados por un residuo ARG al interior de la proteína (RSA 8.4%).
	G/A	R886C	Neutra	La sustitución conduce a la expansión del volumen de la cavidad en 150.336 Å^3.
	T/C	K219R	Neutra	La sustitución conduce a la contracción del volumen de la cavidad en 125.28 Å^3.
	G/T	T768K	Neutra	La sustitución conduce a la expansión del volumen de la cavidad en 127.44 Å^3.
	G/C	P174R	Neutra	La sustitución conduce a la expansión del volumen de la cavidad en 74.088 Å^3.
	G/C	P453R	Neutra	Esta sustitución reemplaza una prolina cis.
	A/C	V732G	Neutra	Esta sustitución resulta en un cambio entre un estado enterrado y un estado expuesto del residuo de la variante objetivo. VAL está al interior de la proteína (RSA 0.0%) y GLY está expuesto (RSA 9.5%).
	C/T	G258S	Neutra	Esta sustitución reemplaza una glicina que originalmente se encontraba en una curvatura o doblez.

Bibliografía

1. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. diciembre de 1977;74(12):5463-7.

2. Wake DT, Ilbawi N, Dunnenberger HM, Hulick PJ. Pharmacogenomics: Prescribing Precisely. Med Clin North Am. noviembre de 2019;103(6):977-90.

3. Verma M, Kulshrestha S, Puri A. Genome Sequencing. En: Keith JM, editor. Bioinformatics: Volume I: Data, Sequence Analysis, and Evolution [Internet]. New York, NY: Springer; 2017 [citado 25 de marzo de 2023]. p. 3-33. (Methods in Molecular Biology). Disponible en: https://doi.org/10.1007/978-1-4939-6622-6_1

4. Petrosino M, Novak L, Pasquo A, Chiaraluce R, Turina P, Capriotti E, et al. Analysis and Interpretation of the Impact of Missense Variants in Cancer. Int J Mol Sci. 21 de mayo de 2021;22(11):5416.

5. Rodrigues C, Santos-Silva A, Costa E, Bronze-da-Rocha E. Performance of In Silico Tools for the Evaluation of UGT1A1 Missense Variants. Hum Mutat. diciembre de 2015;36(12):1215-25.

6. Tsimberidou AM, Fountzilas E, Nikanjam M, Kurzrock R. Review of precision cancer medicine: Evolution of the treatment paradigm. Cancer Treat Rev [Internet]. 1 de junio de 2020 [citado 25 de marzo de 2023];86. Disponible en: https://www.cancertreatmentreviews.com/article/S0305-7372(20)30057-8/fulltext

7. Morganti S, Tarantino P, Ferraro E, D´Amico P, Achutti B, Curigliano G. Next Generation Sequencing (NGS): A Revolutionary Technology in Pharmacogenomics and Personalized Medicine in Cancer. In: Ruiz-Garcia, E., Astudillo-de la Vega, H. (eds) Translational Research and Onco-Omics Applications in the Era of Cancer Personal Genomics. Advances in Experimental Medicine and Biology [Internet]. Vol. 1168. Switzerland: Springer Nature; [citado 25 de marzo de 2023]. 9–30 p. Disponible en: https://link.springer.com/chapter/10.1007/978-3-030-24100-1_2

8. Micaglio E, Locati ET, Monasky MM, Romani F, Heilbron F, Pappone C. Role of Pharmacogenetics in Adverse Drug Reactions: An Update towards Personalized Medicine. Front Pharmacol. 2021;12(651720):1-17.

9. Lunenburg CATC, Henricks LM, Guchelaar HJ, Swen JJ, Deenen MJ, Schellens JHM, et al. Prospective DPYD genotyping to reduce the risk of fluoropyrimidine-induced severe toxicity: Ready for prime time. Eur J Cancer. 1 de febrero de 2016;54:40-8.

10. Sharma V, Gupta SK, Verma M. Dihydropyrimidine dehydrogenase in the metabolism of the anticancer drugs | SpringerLink. 4 de septiembre de 2019;84(6):1157-66.

11. Campbell JM, Bateman E, Peters MD, Bowen JM, Keefe DM, Stephenson MD. Fluoropyrimidine and platinum toxicity pharmacogenetics: an umbrella review of

systematic reviews and meta-analyses. Pharmacogenomics. marzo de 2016;17(4):435-51.

12. Amstutz U, Henricks LM, Offer SM, Barbarino J, Schellens JHM, Swen JJ, et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for Dihydropyrimidine Dehydrogenase Genotype and Fluoropyrimidine Dosing: 2017 Update - Amstutz - 2018 - Clinical Pharmacology & Therapeutics - Wiley Online Library. 103(2):210-2016.

13. Lunenburg CATC, van der Wouden CH, Nijenhuis M, Crommentuijn-van Rhenen MH, de Boer-Veger NJ, Buunk AM, et al. Dutch Pharmacogenetics Working Group (DPWG) guideline for the gene–drug interaction of DPYD and fluoropyrimidines. Eur J Hum Genet. abril de 2020;28(4):508-17.

14. Cacabelos R, Naidoo V, Corzo L, Cacabelos N, Carril JC. Genophenotypic Factors and Pharmacogenomics in Adverse Drug Reactions. Int J Mol Sci. 10 de diciembre de 2021;22(24):13302.

15. Deenen MJ, Meulendijks D, Cats A, Sechterberger MK, Severens JL, Boot H, et al. Upfront Genotyping of DPYD*2A to Individualize Fluoropyrimidine Therapy: A Safety and Cost Analysis. J Clin Oncol. 20 de enero de 2016;34(3):227-34.

16. Henricks LM, Lunenburg CATC, Man FM de, Meulendijks D, Frederix GWJ, Kienhuis E, et al. DPYD genotype-guided dose individualisation of fluoropyrimidine therapy in patients with cancer: a prospective safety analysis. Lancet Oncol. 1 de noviembre de 2018;19(11):1459-67.

17. Innocenti F, Mills SC, Sanoff H, Ciccolini J, Lenz HJ, Milano G. All You Need to Know About DPYD Genetic Testing for Patients Treated With Fluorouracil and Capecitabine: A Practitioner-Friendly Guide. JCO Oncol Pract. diciembre de 2020;16(12):793-8.

18. Barin-Le Guellec C, Lafay-Chebassier C, Ingrand I, Tournamille JF, Boudet A, Lanoue MC, et al. Toxicities associated with chemotherapy regimens containing a fluoropyrimidine: A real-life evaluation in France. Eur J Cancer. 1 de enero de 2020;124:37-46.

19. Farinango C, Gallardo-Cóndor J, Freire-Paspuel B, Flores-Espinoza R, Jaramillo-Koupermann G, López-Cortés A, et al. Genetic Variations of the DPYD Gene and Its Relationship with Ancestry Proportions in Different Ecuadorian Trihybrid Populations. J Pers Med. 10 de junio de 2022;12(6):950.

20. Rodrigues JCG, Fernandes MR, Ribeiro-dos-Santos AM, de Araújo GS, de Souza SJ, Guerreiro JF, et al. Pharmacogenomic Profile of Amazonian Amerindians. J Pers Med. 10 de junio de 2022;12(6):952.

21. Silgado-Guzmán DF, Angulo-Aguado M, Morel A, Niño-Orrego MJ, Ruiz-Torres DA, Contreras Bravo NC, et al. Characterization of ADME Gene Variation in Colombian Population by Exome Sequencing. Front Pharmacol. 2022;13(931531):1-14.

22. Danchin A. In vivo, in vitro and in silico: an open space for the development of microbe-based applications of synthetic biology. Microb Biotechnol. 2022;15(1):42-64.

23. Carvalho C, Varela SAM, Bastos LF, Orfão I, Beja V, Sapage M, et al. The Relevance of In Silico, In Vitro and Non-human Primate Based Approaches to Clinical Research on Major Depressive Disorder. Altern Lab Anim. 1 de julio de 2019;47(3-4):128-39.

24. Pallet N, Hamdane S, Garinet S, Blons H, Zaanan A, Paillaud E, et al. A comprehensive population-based study comparing the phenotype and genotype in a pretherapeutic screen of dihydropyrimidine dehydrogenase deficiency. Br J Cancer. septiembre de 2020;123(5):811-8.

25. De Metz C, Hennart B, Aymes E, Cren PY, Martignène N, Penel N, et al. Complete DPYD genotyping combined with dihydropyrimidine dehydrogenase phenotyping to prevent fluoropyrimidine toxicity: A retrospective study. Cancer Med. 2024;13(6):e7066.

26. Coleman JJ, Pontefract SK. Adverse drug reactions. Clin Med. octubre de 2016;16(5):481-5.

27. Montané E, Santesmases J. Reacciones adversas a medicamentos. Med Clínica. 13 de marzo de 2020;154(5):178-84.

28. Elzagallaai AA, Carleton BC, Rieder MJ. Pharmacogenomics in Pediatric Oncology: Mitigating Adverse Drug Reactions While Preserving Efficacy. Annu Rev Pharmacol Toxicol. 2021;61(1):679-99.

29. Lavan AH, O'Mahony D, Buckley M, O'Mahony D, Gallagher P. Adverse Drug Reactions in an Oncological Population: Prevalence, Predictability, and Preventability. The Oncologist. septiembre de 2019;24(9):e968-77.

30. Freites-Martinez A, Santana N, Arias-Santiago S, Viera A. CTCAE versión 5.0. Evaluación de la gravedad de los eventos adversos dermatológicos de las terapias antineoplásicas. Actas Dermo-Sifiliográficas. 1 de enero de 2021;112(1):90-2.

31. Schütte M, Ogilvie LA, Rieke DT, Lange BMH, Yaspo ML, Lehrach H. Cancer Precision Medicine: Why More Is More and DNA Is Not Enough. Public Health Genomics. 2017;20(2):70-80.

32. Goetz LH, Schork NJ. Personalized medicine: motivation, challenges, and progress. Fertil Steril. 1 de junio de 2018;109(6):952-63.

33. Grandori C, Kemp CJ. Personalized Cancer Models for Target Discovery and Precision Medicine. Trends Cancer. septiembre de 2018;4(9):634-42.

34. Low S, Zembutsu H, Nakamura Y. Breast cancer: The translation of big genomic data to cancer precision medicine. Cancer Sci. marzo de 2018;109(3):497-506.

35. Malki MA, Pearson ER. Drug–drug–gene interactions and adverse drug reactions. Pharmacogenomics J. 2020;20(3):355-66.

36. Cacabelos R, Cacabelos N, Carril JC. The role of pharmacogenomics in adverse drug reactions. Expert Rev Clin Pharmacol. 4 de mayo de 2019;12(5):407-42.

37. Rodríguez-Vicente AE, Lumbreras E, Hernández JM, Martín M, Calles A, Otín CL, et al. Pharmacogenetics and pharmacogenomics as tools in cancer therapy. Drug Metab Pers Ther. 1 de marzo de 2016;31(1):25-34.

38. Mhandire DZ, Goey AKL. The Value of Pharmacogenetics to Reduce Drug-Related Toxicity in Cancer Patients. Mol Diagn Ther. marzo de 2022;26(2):137-51.

39. Kobuchi S, Ito Y. Application of Pharmacometrics of 5-Fluorouracil to Personalized Medicine: A Tool for Predicting Pharmacokinetic–Pharmacodynamic/Toxicodynamic Responses. Anticancer Res. 1 de diciembre de 2020;40(12):6585-97.

40. Sethy C, Kundu CN. 5-Fluorouracil (5-FU) resistance and the new strategy to enhance the sensitivity against cancer: Implication of DNA repair inhibition. Biomed Pharmacother. 1 de mayo de 2021;137:111285.

41. Lam SW, Guchelaar HJ, Boven E. The role of pharmacogenetics in capecitabine efficacy and toxicity. Cancer Treat Rev. 1 de noviembre de 2016;50:9-22.

42. Vodenkova S, Buchler T, Cervena K, Veskrnova V, Vodicka P, Vymetalkova V. 5-fluorouracil and other fluoropyrimidines in colorectal cancer: Past, present and future. Pharmacol Ther. 1 de febrero de 2020;206:107447.

43. ABC Transporter-Mediated Multidrug-Resistant Cancer | SpringerLink [Internet]. [citado 24 de abril de 2023]. Disponible en: https://link.springer.com/chapter/10.1007/978-981-13-7647-4_12

44. Varma A, Mathaiyan J, Shewade D, Dubashi B, Sunitha K. Influence of ABCB-1, ERCC-1 and ERCC-2 gene polymorphisms on response to capecitabine and oxaliplatin (CAPOX) treatment in colorectal cancer (CRC) patients of South India. J Clin Pharm Ther. 2020;45(4):617-27.

45. Nishibeppu K, Komatsu S, Imamura T, Kiuchi J, Kishimoto T, Arita T, et al. Plasma microRNA profiles: identification of miR-1229-3p as a novel chemoresistant and prognostic biomarker in gastric cancer. Sci Rep. 21 de febrero de 2020;10:3161.

46. Thorn CF, Marsh S, Carrillo MW, McLeod HL, Klein TE, Altman RB. PharmGKB summary: fluoropyrimidine pathways. Pharmacogenet Genomics. abril de 2011;21(4):237-42.

47. Castro-Rojas C, Ortiz-López R, Rojas-Martínez A. Farmacogenómica del tratamiento de primera línea en el cáncer gástrico: avances en la identificación de los biomarcadores genómicos de respuesta clínica. Investig Clínica. junio de 2014;55(2):185-202.

48. Wu XP, Dolnick BJ. 5-Fluorouracil alters dihydrofolate reductase pre-mRNA splicing as determined by quantitative polymerase chain reaction. Mol Pharmacol. 1 de julio de 1993;44(1):22-9.

49. Greenhalgh DA, Parish JH. Effect of 5-fluorouracil combination therapy on RNA processing in human colonic carcinoma cells. Br J Cancer. marzo de 1990;61(3):415-9.

50. Noordhuis P, Holwerda U, Wilt CLV der, Groeningen CJV, Smid K, Meijer S, et al. 5-Fluorouracil incorporation into RNA and DNA in relation to thymidylate synthase inhibition of human colorectal cancers. Ann Oncol. 1 de julio de 2004;15(7):1025-32.

51. Wei X, Elizondo G, Sapone A, McLeod HL, Raunio H, Fernandez-Salguero P, et al. Characterization of the Human Dihydropyrimidine Dehydrogenase Gene. Genomics. 1 de agosto de 1998;51(3):391-400.

52. Johnson MR, Wang K, Tillmanns S, Albin N, Diasio RB. Structural Organization of the Human Dihydropyrimidine Dehydrogenase Gene1. Cancer Res. 1 de mayo de 1997;57(9):1660-3.

53. Dobritzsch D, Ricagno S, Schneider G, Schnackerz KD, Lindqvist Y. Crystal Structure of the Productive Ternary Complex of Dihydropyrimidine Dehydrogenase with NADPH and 5-Iodouracil: IMPLICATIONS FOR MECHANISM OF INHIBITION AND ELECTRON TRANSFER∗. J Biol Chem. 12 de abril de 2002;277(15):13155-66.

54. Dobritzsch D, Schneider G, Schnackerz KD, Lindqvist Y. Crystal structure of dihydropyrimidine dehydrogenase, a major determinant of the pharmacokinetics of the anti-cancer drug 5-fluorouracil. EMBO J. 15 de febrero de 2001;20(4):650-60.

55. Brutcher E. 5-Fluorouracil and Capecitabine: Assessment and Treatment of Uncommon Early-Onset Severe Toxicities Associated With Administration. Number 6 Dec 2018. 1 de diciembre de 2018;22(6):627-34.

56. Brooks GA, Tapp S, Daly AT, Busam JA, Tosteson ANA. Cost-effectiveness of DPYD genotyping prior to fluoropyrimidine-based adjuvant chemotherapy for colon cancer. Clin Colorectal Cancer. septiembre de 2022;21(3):e189-95.

57. García-González X, Kaczmarczyk B, Abarca-Zabalía J, Thomas F, García-Alfonso P, Robles L, et al. New DPYD variants causing DPD deficiency in patients treated with fluoropyrimidine. Cancer Chemother Pharmacol. julio de 2020;86(1):45-54.

58. White C, Scott RJ, Paul C, Ziolkowski A, Mossman D, Ackland S. Ethnic Diversity of DPD Activity and the DPYD Gene: Review of the Literature. Pharmacogenomics Pers Med. 9 de diciembre de 2021;14:1603-17.

59. Hamzic S, Schärer D, Offer SM, Meulendijks D, Nakas C, Diasio RB, et al. Haplotype structure defines effects of common DPYD variants c.85T > C (rs1801265) and c.496A > G (rs2297595) on dihydropyrimidine dehydrogenase activity: Implication for 5-fluorouracil toxicity. Br J Clin Pharmacol. agosto de 2021;87(8):3234-43.

60. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for Dihydropyrimidine Dehydrogenase Genotype and Fluoropyrimidine Dosing: 2017 Update - Amstutz - 2018 - Clinical Pharmacology & Therapeutics - Wiley Online

Library [Internet]. [citado 2 de abril de 2023]. Disponible en: https://ascpt.onlinelibrary.wiley.com/doi/10.1002/cpt.911

61. Implementing DPYD*2A Genotyping in Clinical Practice: The Quebec, Canada, Experience - PMC [Internet]. [citado 24 de abril de 2023]. Disponible en: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8018309/

62. Dihydropyrimidine Dehydrogenase Testing prior to Treatment with 5-Fluorouracil, Capecitabine, and Tegafur: A Consensus Paper - FullText - Oncology Research and Treatment 2020, Vol. 43, No. 11 - Karger Publishers [Internet]. [citado 24 de abril de 2023]. Disponible en: https://www.karger.com/Article/FullText/510258

63. Ng PC, Henikoff S. Predicting Deleterious Amino Acid Substitutions. Genome Res. mayo de 2001;11(5):863-74.

64. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods. abril de 2010;7(4):248-9.

65. Tang B, Li B, Gao LD, He N, Liu XR, Long YS, et al. Optimization of in silico tools for predicting genetic variants: individualizing for genes with molecular sub-regional stratification. Brief Bioinform. 25 de septiembre de 2020;21(5):1776-86.

66. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. 1 de julio de 2003;31(13):3812-4.

67. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. Nat Methods. abril de 2014;11(4):361-2.

68. Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res. 1 de septiembre de 2011;39(17):e118.

69. Reva B, Antipin Y, Sander C. Determinants of protein function revealed by combinatorial entropy optimization. Genome Biol. 1 de noviembre de 2007;8(11):R232.

70. biocompute.org.uk. fathmm. [citado 22 de agosto de 2023]. Functional Analysis through Hidden Markov Models (v2.3). Disponible en: http://fathmm.biocompute.org.uk/about.html

71. Quang D, Chen Y, Xie X. DANN: a deep learning approach for annotating the pathogenicity of genetic variants. Bioinformatics. 1 de marzo de 2015;31(5):761-3.

72. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. marzo de 2014;46(3):310-5.

73. Dong C, Wei P, Jian X, Gibbs R, Boerwinkle E, Wang K, et al. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. Hum Mol Genet. 15 de abril de 2015;24(8):2125-37.

74. Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S. Identifying a High Fraction of the Human Genome to be under Selective Constraint Using GERP++. PLoS Comput Biol. 2 de diciembre de 2010;6(12):e1001025.

75. Feng BJ. PERCH: A Unified Framework for Disease Gene Prioritization. Hum Mutat. 2017;38(3):243-51.

76. Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, et al. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. Am J Hum Genet. 6 de octubre de 2016;99(4):877-85.

77. Wu Y, Li R, Sun S, Weile J, Roth FP. Improved pathogenicity prediction for rare human missense variants. Am J Hum Genet. 7 de octubre de 2021;108(10):1891-906.

78. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the Functional Effect of Amino Acid Substitutions and Indels. PLoS ONE. 8 de octubre de 2012;7(10):e46688.

79. IONITA-LAZA I, MCCALLUM K, XU B, BUXBAUM J. A SPECTRAL APPROACH INTEGRATING FUNCTIONAL GENOMIC ANNOTATIONS FOR CODING AND NONCODING VARIANTS. Nat Genet. febrero de 2016;48(2):214-20.

80. Li B, Seligman C, Thusberg J, Miller JL, Auer J, Whirl-Carrillo M, et al. In silico comparative characterization of pharmacogenomic missense variants. BMC Genomics. 20 de mayo de 2014;15(4):S4.

81. Duzkale H, Shen J, McLaughlin H, Alfares A, Kelly M, Pugh T, et al. A systematic approach to assessing the clinical significance of genetic variants. Clin Genet. noviembre de 2013;84(5):453-63.

82. Zloh M, Kirton SB. The benefits of in silico modeling to identify possible small-molecule drugs and their off-target interactions. Future Med Chem. febrero de 2018;10(4):423-32.

83. Masica DL, Karchin R. Towards Increasing the Clinical Relevance of In Silico Methods to Predict Pathogenic Missense Variants. PLoS Comput Biol. 12 de mayo de 2016;12(5):e1004725.

84. Tavtigian SV, Greenblatt MS, Lesueur F, Byrnes GB. In silico analysis of missense substitutions using sequence-alignment based methods. Hum Mutat. noviembre de 2008;29(11):1327-36.

85. Zhou Y, Mkrtchian S, Kumondai M, Hiratsuka M, Lauschke VM. An optimized prediction framework to assess the functional impact of pharmacogenetic variants. Pharmacogenomics J. 2019;19(2):115-26.

86. Lahti JL, Tang GW, Capriotti E, Liu T, Altman RB. Bioinformatics and variability in drug response: a protein structural perspective. J R Soc Interface. 7 de julio de 2012;9(72):1409.

87. Pandi MT, Koromina M, Tsafaridis I, Patsilinakos S, Christoforou E, van der Spek PJ, et al. A novel machine learning-based approach for the computational functional assessment of pharmacogenomic variants. Hum Genomics. 9 de agosto de 2021;15:51.

88. Zhou Y, Lauschke VM. Computational Tools to Assess the Functional Consequences of Rare and Noncoding Pharmacogenetic Variability. Clin Pharmacol Ther. septiembre de 2021;110(3):626-36.

89. Farajzadeh-Dehkordi M, Mafakher L, Samiee-Rad F, Rahmani B. Computational analysis of missense variant CYP4F2*3 (V433M) in association with human CYP4F2 dysfunction: a functional and structural impact. BMC Mol Cell Biol. 9 de mayo de 2023;24:17.

90. Joshi K, Kaur S, Kumar R. Cytochrome P450 2C19 gene polymorphisms (CYP2C19*2 and CYP2C19*3) in chronic myeloid leukemia patients: in vitro and in silico studies. J Biomol Struct Dyn. 2022;40(19):9389-402.

91. Shrestha S, Zhang C, Jerde CR, Nie Q, Li H, Offer SM, et al. Gene-specific variant classifier (DPYD-Varifier) to identify deleterious alleles of dihydropyrimidine dehydrogenase. Clin Pharmacol Ther. octubre de 2018;104(4):709-18.

92. Capitain O, Seegers V, Metges JP, Faroux R, Stampfli C, Ferec M, et al. Comparison of 4 Screening Methods for Detecting Fluoropyrimidine Toxicity Risk: Identification of the Most Effective, Cost-Efficient Method to Save Lives. Dose-Response. 14 de septiembre de 2020;18(3):1559325820951367.

93. Tricco AC, Langlois EV, Straus SE, Research A for HP and S, Organization WH. Rapid reviews to strengthen health policy and systems: a practical guide [Internet]. World Health Organization; 2017 [citado 22 de agosto de 2023]. xix, 119 p. Disponible en: https://apps.who.int/iris/handle/10665/258698

94. Kelly SE, Moher D, Clifford TJ. Quality of conduct and reporting in rapid reviews: an exploration of compliance with PRISMA and AMSTAR guidelines. Syst Rev. 10 de mayo de 2016;5(1):79.

95. Ciapponi A. QUADAS-2: instrumento para la evaluación de la calidad de estudios de precisión diagnóstica. Evid Actual En Práctica Ambulatoria [Internet]. 1 de abril de 2015 [citado 23 de agosto de 2023];18(1). Disponible en: https://www.evidencia.org/index.php/Evidencia/article/view/6341

96. Brnich SE, Abou Tayoun AN, Couch FJ, Cutting GR, Greenblatt MS, Heinen CD, et al. Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. Genome Med. 31 de diciembre de 2019;12(1):3.

97. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. GigaScience. 2021;10(2):1-4.

98. Liu X, Li C, Mou C, Dong Y, Tu Y. dbNSFP v4: a comprehensive database of transcript-specific functional predictions and annotations for human nonsynonymous and splice-site SNVs. Genome Med. 2 de diciembre de 2020;12(1):103.

99. Tian Y, Pesaran T, Chamberlin A, Fenwick RB, Li S, Gau CL, et al. REVEL and BayesDel outperform other in silico meta-predictors for clinical variant classification. Sci Rep. 4 de septiembre de 2019;9(1):12752.

100. Pshennikova VG, Barashkov NA, Romanov GP, Teryutin FM, Solov'ev AV, Gotovtsev NN, et al. Comparison of Predictive In Silico Tools on Missense Variants in GJB2, GJB6, and GJB3 Genes Associated with Autosomal Recessive Deafness 1A (DFNB1A). Sci World J. 20 de marzo de 2019;2019:5198931.

101. Pejaver V, Byrne AB, Feng BJ, Pagel KA, Mooney SD, Karchin R, et al. Calibration of computational tools for missense variant pathogenicity classification and ClinGen recommendations for PP3/BP4 criteria. Am J Hum Genet. 1 de diciembre de 2022;109(12):2163-77.

102. Hicks SA, Strümke I, Thambawita V, Hammou M, Riegler MA, Halvorsen P, et al. On evaluation metrics for medical applications of artificial intelligence. Sci Rep. 8 de abril de 2022;12:5979.

103. Fawcett T. An introduction to ROC analysis. Pattern Recognit Lett. 1 de junio de 2006;27(8):861-74.

104. Cerda J, Cifuentes L. Uso de curvas ROC en investigación clínica: Aspectos teórico-prácticos. Rev Chil Infectol. abril de 2012;29(2):138-41.

105. scikit-learn [Internet]. [citado 10 de octubre de 2023]. sklearn.metrics.roc_curve. Disponible en: https://scikit-learn/stable/modules/generated/sklearn.metrics.roc_curve.html

106. Venselaar H, te Beek TA, Kuipers RK, Hekkelman ML, Vriend G. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. BMC Bioinformatics. 8 de noviembre de 2010;11(1):548.

107. Khanna T, Hanna G, Sternberg MJE, David A. Missense3D-DB web catalogue: an atom-based analysis and repository of 4M human protein-coding genetic variants. Hum Genet. 1 de mayo de 2021;140(5):805-12.

108. Arthur R, O´Conell J. akt — ancestry and kinship toolkit. [citado 19 de octubre de 2023]. akt — ancestry and kinship toolkit. Disponible en: https://illumina.github.io/akt/

109. Variant calling [Internet]. [citado 27 de abril de 2023]. Disponible en: https://samtools.github.io/bcftools/howtos/variant-calling.html

110. Ramudo-Cela L, López-Martí JM, Colmeiro-Echeberría D, de-Uña-Iglesias D, Santomé-Collazo JL, Monserrat-Iglesias L, et al. Desarrollo y validación de un panel de secuenciación masiva en paralelo para farmacogenética clínica. Farm Hosp. diciembre de 2020;44(6):243-53.

111. NGSEP [Internet]. NGSEP; 2023 [citado 23 de octubre de 2023]. Disponible en: https://github.com/NGSEP/NGSEPcore

112. Ghosh R, Harrison SM, Rehm HL, Plon SE, Biesecker LG. Updated Recommendation for the Benign Stand Alone ACMG/AMP Criterion. Hum Mutat. noviembre de 2018;39(11):1525-30.

113. Leong IU, Stuckey A, Lai D, Skinner JR, Love DR. Assessment of the predictive accuracy of five in silico prediction tools, alone or in combination, and two metaservers to classify long QT syndrome gene mutations. BMC Med Genet. 13 de mayo de 2015;16:34.

114. Offer SM, Fossum CC, Wegner NJ, Stuflesser AJ, Butterfield GL, Diasio RB. Comparative functional analysis of DPYD variants of potential clinical relevance to dihydropyrimidine dehydrogenase activity. Cancer Res. 1 de mayo de 2014;74(9):2545-54.

115. Offer SM, Wegner NJ, Fossum C, Wang K, Diasio RB. Phenotypic profiling of DPYD variations relevant to 5-fluorouracil sensitivity using real-time cellular analysis and in vitro measurement of enzyme activity. Cancer Res. 15 de marzo de 2013;73(6):1958-68.

116. Ittisoponpisan S, Islam SA, Khanna T, Alhuzimi E, David A, Sternberg MJE. Can Predicted Protein 3D Structures Provide Reliable Insights into whether Missense Variants Are Disease Associated? J Mol Biol. 17 de mayo de 2019;431(11):2197-212.

117. Hagenkord J, Funke B, Qian E, Hegde M, Jacobs KB, Ferber M, et al. Design and Reporting Considerations for Genetic Screening Tests. J Mol Diagn. 1 de mayo de 2020;22(5):599-609.

118. Montenegro LR, Lerário AM, Nishi MY, Jorge AAL, Mendonca BB. Performance of mutation pathogenicity prediction tools on missense variants associated with 46,XY differences of sex development. Clinics. 2021;76:e2052.

119. Santini A, Man A, Voidăzan S. Accuracy of Diagnostic Tests. J Crit Care Med. 5 de agosto de 2021;7(3):241-8.

120. Rosenbaum K, Jahnke K, Curti B, Hagen WR, Schnackerz KD, Vanoni MA. Porcine recombinant dihydropyrimidine dehydrogenase: comparison of the spectroscopic and catalytic properties of the wild-type and C671A mutant enzymes. Biochemistry. 15 de diciembre de 1998;37(50):17598-609.

121. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. mayo de 2020;581(7809):434-43.

122. Bonifaz-Peña V, Contreras AV, Struchiner CJ, Roela RA, Furuya-Mazzotti TK, Chammas R, et al. Exploring the Distribution of Genetic Markers of Pharmacogenomics Relevance in Brazilian and Mexican Populations. PLoS ONE. 24 de noviembre de 2014;9(11):e112640.

123. Rodrigues-Soares F, Suarez-Kurtz G. Pharmacogenomics research and clinical implementation in Brazil. Basic Clin Pharmacol Toxicol. 2019;124(5):538-49.

124. Cavalcante GC, Freitas NDSDC, Ribeiro-Dos-Santos AM, Carvalho DCD, Silva EMD, Assumpção PPD, et al. Investigation of Potentially Deleterious Alleles for Response to Cancer Treatment with 5-Fluorouracil. Anticancer Res. 1 de diciembre de 2015;35(12):6971-7.

125. Zhou Y, Fujikura K, Mkrtchian S, Lauschke VM. Computational Methods for the Pharmacogenetic Interpretation of Next Generation Sequencing Data. Front Pharmacol. 4 de diciembre de 2018;9:1437.

126. Principi N, Petropulacos K, Esposito S. Impact of Pharmacogenomics in Clinical Practice. Pharmaceuticals. noviembre de 2023;16(11):1596.

127. Frontiers | A practical guide for the generation of model-based virtual clinical trials [Internet]. [citado 25 de febrero de 2024]. Disponible en: https://www.frontiersin.org/articles/10.3389/fsysb.2023.1174647/full

128. Methods to Develop an in silico Clinical Trial: Computational Head-to-Head Comparison of Lisdexamfetamine and Methylphenidate - PMC [Internet]. [citado 25 de febrero de 2024]. Disponible en: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8595241/

129. Alnasser B. A Review of Literature on the Economic Implications of Implementing Artificial Intelligence in Healthcare. E-Health Telecommun Syst Netw. 15 de septiembre de 2023;12(3):35-48.

130. Moldrup C. Ethical, social and legal implications of pharmacogenomics: a critical review. Community Genet. 2001;4(4):204-14.