



**UNIVERSIDAD TÉCNICA DE AMBATO**

**FACULTAD DE CIENCIA E INGENIERÍA EN  
ALIMENTOS Y BIOTECNOLOGÍA**



**CARRERA DE BIOTECNOLOGIA**

---

**Tema:** Evaluación de posibles dianas terapéuticas para el cáncer de colon aplicando Pipeline Bioinformático como base para el diseño de fármacos asistido por computador

---

Trabajo de titulación, Modalidad Proyecto de Investigación, previo a la obtención de título de Ingeniero Biotecnólogo, otorgado por la Universidad Técnica de Ambato, a través de la Facultad de Ciencia e Ingeniería en Alimentos y Biotecnología.

**Autor:** Mario Fabricio Jácome Campos

**Tutor:** MSc. Cristian Fernando Galarza Galarza

**Ambato – Ecuador**

**Septiembre -2023**

## **APROBACIÓN DEL TUTOR**

MSc. Cristian Fernando Galarza Galarza

### **CERTIFICA:**

Que el presente Trabajo de Titulación ha sido prolijamente revisado. Por lo tanto, autorizo la presentación de este Trabajo de Titulación bajo la modalidad de Proyecto de Investigación, el mismo que responde a las normas establecidas en el Reglamento de Titulación y Grados de la Facultad de Ciencia e Ingeniería en Alimentos y Biotecnología.

Ambato, 25 de Julio del 2023

---

MSc. Cristian Fernando Galarza Galarza

**C.I. 1803160272**

**TUTOR**

## **AUTORÍA DEL TRABAJO DE TITULACION**

Yo, Mario Fabricio Jácome, manifiesto que los resultados obtenidos en el presente Trabajo de Titulación, modalidad Proyecto de Investigación, previo a la obtención del título de Ingeniero Biotecnólogo, son absolutamente originales, auténticos y personales, a excepción de las citas bibliográficas.



Mario Fabricio Jácome Campos

**C.I. 1004127377**

**Autor**

## **DERECHOS DE AUTOR**

Autorizo a la Universidad Técnica de Ambato, para que haga de este Trabajo de Trabajo de Titulación parte de él un documento disponible para su lectura, consulta y procesos de investigación, según las normas de la institución.

Cedo los derechos en línea patrimoniales de mi Trabajo de Titulación, con fines de difusión pública, además apruebo la reproducción de este, dentro de las regulaciones de la Universidad, siempre y cuando esta reproducción no suponga una ganancia económica y se realice respetando mis derechos de autor

A handwritten signature in blue ink, consisting of several loops and strokes, positioned above a horizontal dotted line.

Mario Fabricio Jácome Campos

**C.I. 1004127377**

**Autor**

## **APROBACIÓN DEL TRIBUNAL DE GRADO**

Los suscritos profesores calificadores, aprueban el presente Trabajo de Titulación, modalidad Proyecto de Investigación, el mismo que ha sido elaborado de conformidad con las disposiciones emitidas por la Facultad de Ciencia e Ingeniería en Alimentos y Biotecnología de la Universidad Técnica de Ambato.

Para constancia firman:

---

**Presidente del Tribunal**

---

**Dr. Mario Daniel García Solís**

C.I. 1103605471

---

**Dra. Lorena De Los Ángeles Núñez Villacis**

C.I. 1804256905

Ambato, 24 de Agosto del 2023

## DEDICATORIA

*Con el corazón lleno de gratitud y emoción, quiero expresarles mi más sincero agradecimiento por haber sido parte fundamental de este apasionante viaje académico.*

*A mis abuelitos mamá Esperancita, papá Fernando que están en el cielo agradecerles por el amor que me brindaron y su apoyo inquebrantable dedico esta tesis con todo mi corazón.*

*Dedicarle el presente trabajo a toda mi familia, de mi madre y de mi padre que estuvieron apoyándome en esta etapa de mi vida. A mi madre Fanny Campos que su ejemplo de perseverancia, sabiduría ha sido un faro en mi camino hacia el crecimiento personal. Mi padre Mario Jácome que me ha enseñado a confiar en mí mismo de todo se puede a pesar de las adversidades.*

*En memoria de mi querido amigo Jhon Altamirano, quien ahora brilla como una estrella en el cielo. Te agradezco por brindarme tu apoyo incondicional. Tu amistad sincera fue un regalo inestimable en mi vida.*

*A mis entrañables amigos Susana, Steven, Joseph, Mateo, Anthony, Elvis, Andrés, y a mis primos Blady, Karen: Esta tesis es un testimonio del poder de la amistad y la colaboración. Ustedes han sido mi columna vertebral durante los difíciles momentos que este trabajo ha implicado, y sin su apoyo constante, no habría sido posible alcanzar este logro. Agradezco su incondicional amistad, sus palabras de aliento y sus innumerables horas de compañía en esta travesía académica.*

## AGRADECIMIENTO

*Quiero agradecer primeramente a Dios por brindarme la oportunidad de seguir con la lucha de la vida y de darme retos valiosos para forjar mi carácter.*

*A mi querido tío Jorge Campos que le considero como mi padre le agradezco la enseñanza de no darme por vencido a pesar de las adversidades.*

*Agradezco a mi ñaños Campos Encalada: Xavi, Rene, Fernando, Hiralda, Carmen, Katy, Mónica que me brindaron su apoyo incondicional y me ayudaron siempre en los momentos difíciles de la vida y me dieron la oportunidad de seguir en este camino de la vida.*

*A mi profesor Jaime Benalcázar le doy las gracias por desafiarme constantemente a superar mis límites y por brindarme el apoyo necesario para crecer académicamente.*

*A mi estimado Ángel Lemache agradecer su apoyo me ha dado la fuerza para seguir adelante y enfrentar cualquier obstáculo que se presente en mi camino. De la misma manera agradecer a mi tío Jaime Jácome que siempre ha estado ahí para mí, brindando consejos y ánimos cuando más los necesitaba.*

*Quiero agradecer a mi tutor Msc. Cristian Galarza por compartir su conocimiento y tenerme mucha paciencia de enseñarme y darle las gracias que con su conocimiento se alcanzó el proyecto de investigación.*

*Por último, quiero agradecer a toda mi familia, amigos del colegio, de la universidad como es Joao, Santiago, Luis, Daniel, Pablo, Andrés Sánchez, Andrés Gutiérrez y a mis profesores que me enseñaron que las cosas que uno se propone por más difíciles que sean siempre habrá una luz para terminar lo que se propone.*

## ÍNDICE GENERAL DE CONTENIDOS

APROBACIÓN DEL TUTOR.....	ii
AUTORÍA DEL TRABAJO DE TITULACION.....	iii
DERECHOS DE AUTOR.....	iv
APROBACIÓN DEL TRIBUNAL DE GRADO.....	v
DEDICATORIA.....	vi
AGRADECIMIENTO.....	vii
ÍNDICE DE TABLAS.....	x
ÍNDICE DE FIGURAS.....	xi
ÍNDICE DE ANEXOS.....	xii
RESUMEN EJECUTIVO.....	xiii
ABSTRACT.....	xiv
CAPÍTULO I.....	1
MARCO TEÓRICO.....	1
1.1. Antecedentes Investigativos.....	1
1.1.1. Cáncer de Colon (CC).....	1
1.1.2. Mecanismos Moleculares del Cáncer de colon (CC).....	6
1.1.3. Enfoques terapéuticos.....	9
1.1.4. Búsqueda de tratamiento para el cáncer de colon a través de herramientas Bioinformáticas.....	10
1.3. Objetivos.....	16
1.3.1. Objetivo General.....	16
1.3.2. Objetivos Específicos.....	16
CAPÍTULO II.....	17
METODOLOGÍA.....	17
2.1. Materiales y equipos.....	17

2.1.1. Materiales.....	17
2.2.2 Equipos .....	17
2.3 Métodos.....	18
2.3.1. Identificación de las Bases Genéticas y Moleculares del Cáncer de Colon.....	18
2.3.2. Expresión diferencial de los genes que se expresan o inhiben en el cáncer de colon. .....	18
2.3.3. Pipeline Bioinformático para la identificación de dianas terapéuticas .....	19
2.3.4. Prueba de viabilidad de las dianas terapéuticas .....	22
2.3.5. Búsqueda de moléculas con posible interacción con las dianas.....	23
2.3.6. Cribado Virtual .....	24
2.3.7. Análisis de toxicidad de las moléculas.....	25
CAPITULO III.....	26
RESULTADOS Y DISCUSIÓN.....	26
3.1 Análisis y discusión de los resultados.....	26
3.1.1. Análisis de expresión diferencial .....	26
3.1.2. Análisis de las bases de datos. ....	30
3.1.3. Análisis de la integración de datos.....	36
3.1.4. Análisis de la validación de los genes seleccionados.....	42
3.1.5. Evaluación de la viabilidad de las dianas terapéuticas.....	45
3.1.6 Análisis del cribado virtual .....	49
3.1.7. Análisis de interacción molecular .....	58
3.1.8. Análisis de la toxicidad de la molécula.....	63
CAPITULO IV.....	72
CONCLUSIONES Y RECOMENDACIONES.....	72
4.1. Conclusiones.....	72
4.2. Recomendaciones .....	73
REFERENCIAS BIBLIOGRÁFICAS.....	74
ANEXOS .....	90

## ÍNDICE DE TABLAS

<b>Tabla 1.</b> Principales genes diferenciales. ....	28
<b>Tabla 2.</b> Genes relacionados a la aparición de cáncer de colon obtenidos de las bases de datos con relación gen-enfermedad. ....	32
<b>Tabla 3.</b> Genes elite destacados para la investigación y establecimiento de redes biológicas de interacción por Cytoscape. ....	33
<b>Tabla 4.</b> Interacción de la red biológica de la regulación de transcripción. ....	38
<b>Tabla 5.</b> Características generales de la red biológica de la regulación de la transcripción. ....	39
<b>Tabla 6.</b> Interacción de la red biológica de la hiperactivación de la vía de señalización de MAPK. ....	¡Error! Marcador no definido.
<b>Tabla 7.</b> Características generales de la red biológica de la hiperactivación de la vía de señalización de MAPK. ....	42
<b>Tabla 8.</b> Fármacos relacionados con los genes elite del CC. ....	43
<b>Tabla 9.</b> Puntuación de los sitios de unión de genes elite. ....	46
<b>Tabla 10.</b> Fármacos con mejor energía de afinidad como posibles dianas terapéuticas para el CC relacionadas con la proteína MutL Homolog 1. ....	55
<b>Tabla 11.</b> Fármacos obtenidos con mejor energía de afinidad como posibles dianas terapéuticas para el CC relacionadas con la proteína Mothers Against Decapentaplegic Homolog 2. ....	56
<b>Tabla 12.</b> Fármacos con mejor energía de afinidad como posibles dianas terapéuticas para el CC relacionadas con la proteína Caspasa 9. ....	57
<b>Tabla 13.</b> Afinidad de unión con los principales genes seleccionados. ....	59
<b>Tabla 14.</b> Interpretación para los datos farmacológicos. ....	64
<b>Tabla 15.</b> Prueba de toxicidad de los fármacos seleccionado. ....	64
<b>Tabla 16.</b> Propiedades de absorción del Etopósido. ....	66
<b>Tabla 17.</b> Propiedades de distribución del etopósido. ....	67
<b>Tabla 18.</b> Propiedades del metabolismo del etopósido. ....	68
<b>Tabla 19.</b> Propiedades de excreción del etopósido. ....	68
<b>Tabla 20.</b> Propiedades de toxicidad del etopósido. ....	70

## ÍNDICE DE FIGURAS

<b>Figura 1.</b> Número estimado de casos nuevos en 2020 en el Ecuador expresados en ambos sexos y todas las edades.....	3
<b>Figura 2.</b> Número estimado de muertes en el 2020 en el Ecuador expresados en ambos sexos y todas las edades. ....	4
<b>Figura 3.</b> Fases del desarrollo del cáncer de colon. ....	4
<b>Figura 4.</b> Modelo de acoplamiento molecular. ....	14
<b>Figura 5.</b> Diagrama de Venn.....	27
<b>Figura 6.</b> Gráfica de volcán que muestra una descripción general de la expresión diferencial de todos los genes. ....	27
<b>Figura 7.</b> Análisis de mapa de calor de los primeros 20 genes diferenciales.....	29
<b>Figura 8.</b> Ruta disfuncional metabólica del cáncer de colon .....	31
<b>Figura 9.</b> Red biológica de la regulación de transcripción.....	38
<b>Figura 10.</b> Red biológica de la sobre activación de la vía de señalización del gen MAPK .....	41
<b>Figura 11.</b> Pocket_0 de la zona drogable de la proteína MutL Homolog 1 .....	48
<b>Figura 12.</b> Pocket_0 de la zona drogable de la proteína Mothers Against Decapentaplegic Homolog 2. ....	49
<b>Figura 13.</b> Pocket_0 de la zona drogable de la proteína Caspase 9 .....	49
<b>Figura 14.</b> Red de los genes involucrados la vía de señalización de MAPK con interacción de medicamentos de inhibición. ....	53
<b>Figura 15.</b> Red de los genes involucrados en la regulación de transcripción con interacción de medicamentos de inhibición. ....	54
<b>Figura 16.</b> Pose acoplada de la estructura de la proteína Caspasa 9 (Alphafold) .....	61
<b>Figura 17.</b> Pose acoplada de la estructura de la proteína SMAD2 (PDB:5XOD). ....	63

## ÍNDICE DE ANEXOS

<b>Anexo A.</b> <i>Dataset en RStudio para el reconocimiento de genes de expresión</i> .....	90
<b>Anexo B.</b> <i>Búsqueda en MalaCards para obtener genes elite con la entrada Adenocarcinoma colon</i> .....	90
<b>Anexo C.</b> <i>Análisis de los genes asociados con la enfermedad del CC</i> .....	91
<b>Anexo D.</b> <i>Genes de élite obtenidos en la base de datos Harmonizome</i> .....	91
<b>Anexo E.</b> <i>Genes relacionados a la aparición de cáncer de colon obtenidos de las bases de datos con relación gen-enfermedad</i> .....	92
<b>Anexo F.</b> <i>Búsqueda y análisis de la cristalografía tridimensional del gen MLH1</i> ..	99
<b>Anexo G.</b> <i>Construcción de la red biológica de la alteración bioquímica de la regulación de transcripción</i> .....	99
<b>Anexo H.</b> <i>Genes representativos de la interacción de la red biológica de la regulación de transcripción</i> .....	100
<b>Anexo I.</b> <i>Genes representativos de la interacción de la red biológica de la regulación de transcripción</i> .....	102
<b>Anexo J.</b> <i>Validación de posibles fármacos disponibles para el gen MLH1</i> .....	103
<b>Anexo K.</b> <i>Selección de los aminoácidos residuales para obtener las coordenadas de la GRID-BOX</i> .....	103
<b>Anexo I.</b> <i>Web de cribado virtual basado en la estructura tridimensional y las coordenadas de la GRI-BOX</i> .....	104
<b>Anexo M.</b> <i>QSAR-VEGA in silico evaluación de mutagenicidad, carcinogénesis y hepatotoxicidad para las dianas obtenidas</i> .....	104
<b>Anexo N.</b> <i>pkCSM para la evaluación de pruebas ADMET para la diana seleccionada</i> .....	105

## RESUMEN EJECUTIVO

El cáncer de colon es un tipo común de tumor maligno en el tracto gastrointestinal, representa alrededor del 13 por ciento de todos los tumores. Su desarrollo está influenciado por cambios epigenéticos, genéticos y factores ambientales. No obstante, los métodos actuales para diagnosticar y tratar esta enfermedad suelen ser costosos, dificultando el acceso a tratamientos en muchos pacientes.

Para la identificación de los genes que se expresan o inhiben en el cáncer de colon se efectuó el análisis de expresión diferencial utilizando R. Estos genes se contrastaron con bases de datos tales como: OMIM, Malacards, Harmonizome y KEGG. Los genes más relevantes se agrupan según su alteración bioquímica, con esta información se crean redes de interacción biológica por medio de Cytoscape para identificar proteínas que se alteran dentro del proceso de la enfermedad. Para la identificación de las dianas, se utilizó la base de datos DrugBank, mediante DoGSiteScorer se identificó los sitios drugables, se realiza el cribado virtual a través de MTiopenScreen para encontrar posibles ligandos que presenten alta afinidad de unión. Las redes de interacción obtenidas desde STITCH permiten identificar el medicamento que este ligado a las rutas metabólicas para analizar las interacciones proteína-ligando.

Se identificó el fármaco etopósido, del que se evalúan las propiedades farmacocinéticas y de toxicidad mediante predicciones bioinformáticas utilizando modelos QSAR VEGA y pkCSM. Se encontró que la toxicidad y hepatotoxicidad son negativas, lo que sugiere que este podría ser un candidato viable para el desarrollo de medicamentos dirigidos a los genes SMAD2 y CASP9.

**Palabras clave:** Cáncer, DIFAC, DEG, Docking Molecular

## ABSTRACT

Colon cancer is a common type of malignant tumor in the gastrointestinal tract, accounting for about 13 percent of all tumors. Its development is influenced by epigenetic changes, genetic and environmental factors. However, current methods for diagnosing and treating this disease are often expensive, making access to treatment difficult for many patients.

To identify the genes that are expressed or inhibited in colon cancer, differential expression analysis was performed using R. These genes were contrasted with databases such as: OMIM, Malacards, Harmonizome, and KEGG. The most relevant genes are grouped according to their biochemical alteration, with this information biological interaction networks are created through Cytoscape to identify proteins that are altered within the disease process. For the identification of the targets, the DrugBank database was used, through DoGSiteScorer the drugable sites were identified, virtual screening is performed through MTiopenScreen to find possible ligands that present high binding affinity. The interaction networks obtained from STITCH allow the identification of the drug that is linked to the metabolic pathways to analyze the protein-ligand interactions.

The drug etoposide was identified and its pharmacokinetic and toxicity properties are assessed by bioinformatics predictions using QSAR VEGA and pkCSM models. The toxicity and hepatotoxicity were found to be negative, suggesting that this could be a viable candidate for the development of drugs targeting the SMAD2 and CASP9 genes.

**Keywords:** Cancer, DIFAC, DEG, Molecular Docking

# CAPÍTULO I

## MARCO TEÓRICO

### 1.1. Antecedentes Investigativos

#### 1.1.1. Cáncer de Colon (CC)

El cáncer de colon es una de las enfermedades más comunes del tracto gastrointestinal, representando alrededor del 13% de todos los tumores malignos. Dicha enfermedad se desarrolla a partir de cambios focales dentro de pólipos precancerosos benignos. Estos pólipos son crecimientos o agregaciones localizadas de células anormales dentro de la mucosa intestinal que sobresalen hacia la luz intestinal (**Granados *et al.*, 2017**). A lo largo del tiempo, las células de los pólipos acumulan cambios genéticos que les otorgan la capacidad de invadir la pared intestinal, propagándose a los ganglios linfáticos locales y a sitios metastásicos distantes (**Simon, 2016**).

Se ha encontrado que alrededor del 65% de los casos de cáncer de colon se presentan de manera esporádica, lo que significa que múltiples alteraciones genómicas y epigenéticas somáticas son adquiridas. Por otro lado, el 25% de los casos se asocia con antecedentes familiares hereditarios, el 5% es debido a variaciones genéticas conocidas de baja penetrancia y el 5% restante se debe a otras alteraciones genómicas hereditarias desconocidas (**Keum & Giovannucci, 2019**).

El cáncer de colon es una de las neoplasias más prevalentes en el mundo, afectando tanto a hombres como mujeres. Estudios recientes, es el segundo tipo de cáncer más común en mujeres y el tercero en hombres. La mayoría de los casos de cáncer de colon ocurren en personas mayores de 50 años, lo que sugiere que la edad avanzada es un factor importante en el riesgo de desarrollar esta enfermedad (**Benson *et al.*, 2021**). Se ha demostrado que los niveles de estrógenos tienen un efecto protector contra el desarrollo del cáncer de colon en mujeres. Por lo tanto, las mujeres que experimentan niveles bajos de estrógenos después de la menopausia pueden tener un mayor riesgo de padecer cáncer de colon (**Aghabozorgi *et al.*, 2019**).

Los factores de riesgo que contribuyen al desarrollo del cáncer de colon pueden ser clasificados en factores genéticos, epigenéticos y ambientales. Entre los factores ambientales se encuentran: las enfermedades preexistentes (enfermedades inflamatorias

del intestino, diabetes tipo 2, obesidad), inactividad física, tabaquismo, consumo de alcohol y los patrones dietéticos inadecuados (bajo consumo de: fibra, frutas, verduras, calcio, productos dietéticos y alta en carnes rojas y procesadas **(Duraiyarasan et al., 2022)**). Otros factores como es las alteraciones en el microbioma intestinal, la edad, el sexo, la raza y el nivel socioeconómico son factores importantes que influyen en el desarrollo del cáncer de colon **(Sawicki et al., 2021)**.

#### **1.1.1.1.Epidemiología**

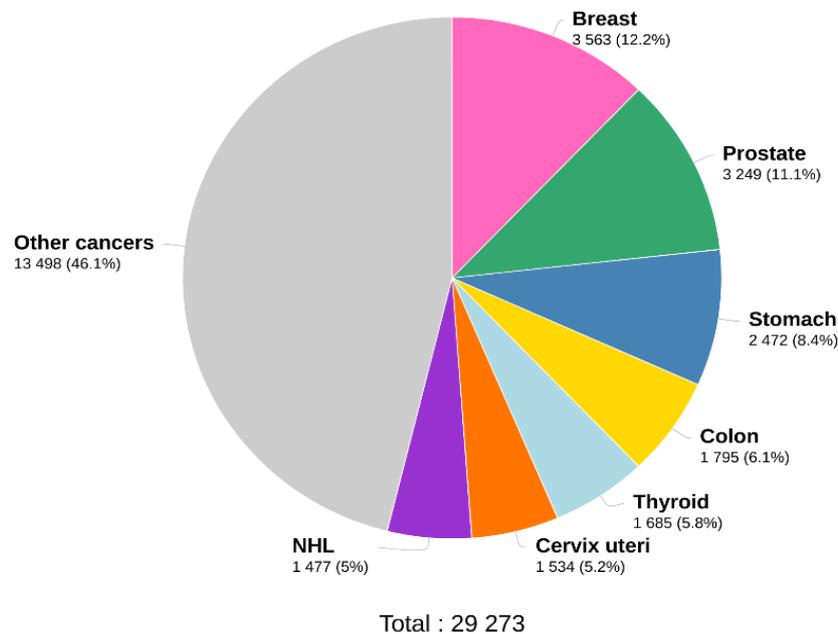
En los últimos años, el cáncer de colon se ha ido incrementando tanto en su incidencia como la mortalidad a nivel global. De acuerdo con estimaciones recientes, el cáncer de colon es la tercera neoplasia más comúnmente diagnosticada y la segunda causa de muerte por cáncer a nivel global **(Bray et al., 2018)**. Dicho aumento se debe en parte a la falta de cumplimiento de los programas de detección temprana o la estrategia terapéutica.

En el 2020, se diagnosticaron 1.93 millones de casos de cáncer de colon en todo el mundo y alrededor de 935000 muertes. Las estimaciones indican que se produjeron alrededor de 515.637 muertes en hombres y 419.536 muertes en mujeres en todo el mundo **(Xi & Xu, 2021)**. Además, se prevé que para el 2040 aumente el número de casos a nivel mundial, alcanzando los 3.2 millones, esto debido al envejecimiento de la población, el crecimiento de la población y el desarrollo humano **(ECIS, 2022)**.

A nivel en Latinoamérica y el Caribe el cáncer de colon se estima que en el 2020 presento 128.006 (6,9 %) de casos mientras la mortalidad de 64.666 (7.3%), esto quiere decir que en la incidencia de los hombres 63.036 de casos mientras que las mujeres mostro 64.970 de casos y en la mortalidad de 64.666 (7,3%) de casos esto se divide en 32.371 en hombres y en mujeres 32.295 de casos que se presentado **(Ferlay et al., 2019)**.

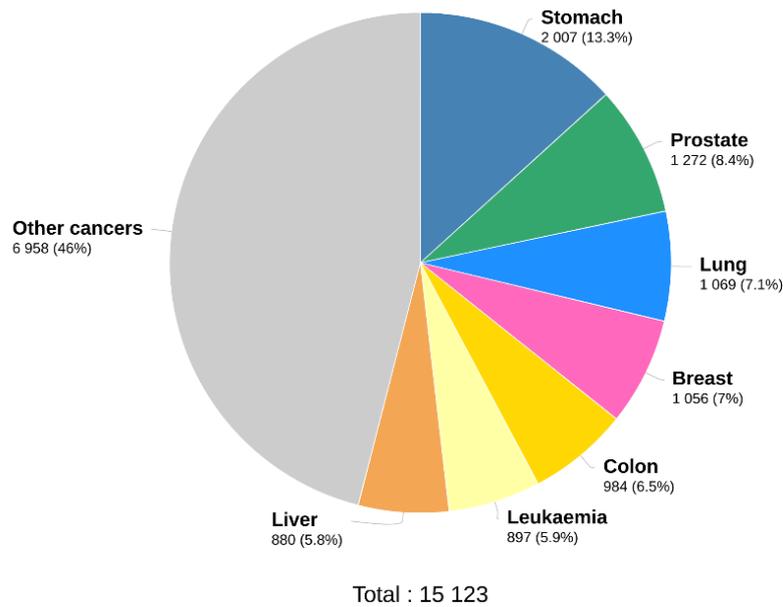
En el Ecuador, la incidencia de cáncer de colon ocupa el quinto lugar en relación con todos los cánceres. Según las estadísticas del 2020 de la Organización Mundial de la Salud se estima que ubo 29273 casos de cánceres, en donde solo el 6,1% (1795 casos) (Figura 1) represento la incidencia de cáncer de colon. Además, los hombres presentaron una mayor incidencia, ocupando el cuarto lugar otorgándole 837 casos (6.3%) del total de canceres en hombres de 13190. En cuanto a la tasa de mortalidad se estimó que hay un

total de 15123 de muertes por cáncer (Figura 2), en donde solo el 6.5% representa 984 muertes. No obstante, se observó que hay una ligera incidencia en mortalidad con un total de 521 fallecidos (6.7%) en comparación con el total de los casos de mortalidad de 7827 (World Health Organization, 2020). Esto puede ser debido, a los niveles bajos de estrógenos en mujeres después de la menopausia aumentan el riesgo de desarrollar cáncer de colon (Aghabozorgi *et al.*, 2019).



**Figura 1.** Número estimado de casos nuevos en 2020 en el Ecuador expresados en ambos sexos y todas las edades.

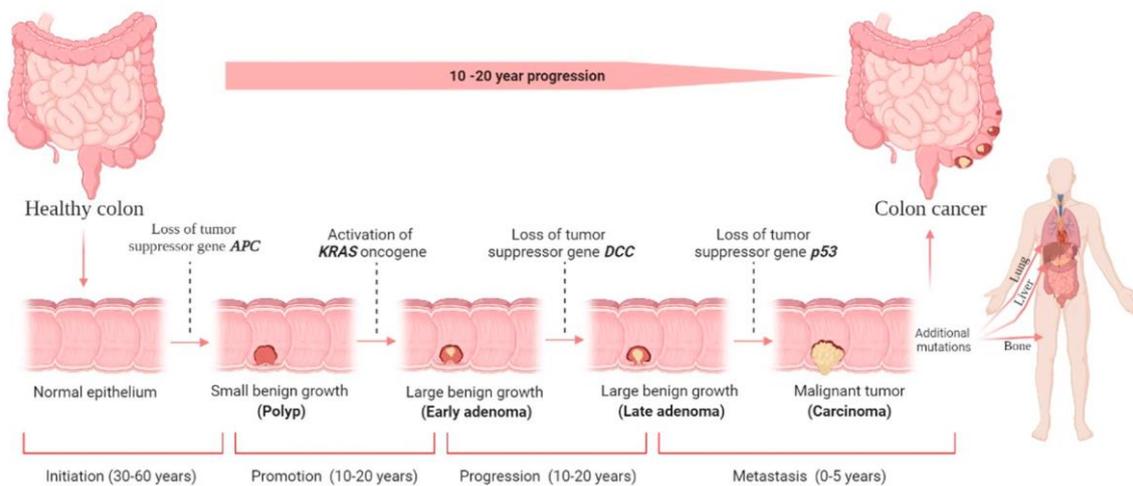
*Nota.* Adaptado de Global Cancer Observatory [Fotografía], por World Health Organization, 2020, Flickr (<https://gco.iarc.fr/>).



**Figura 2.** Número estimado de muertes en el 2020 en el Ecuador expresados en ambos sexos y todas las edades.

*Nota.* Adaptado de Global Cancer Observatory [Fotografía], por World Health Organization, 2020, Flickr (<https://gco.iarc.fr/>).

### 1.1.1.2. Etapas del Cáncer de Colon



**Figura 3.** Fases del desarrollo del cáncer de colon.

*Nota.* La figura muestra una comparación entre un colon normal y un cáncer de colon, Además permite conocer las cuatro etapas del crecimiento de la carcinogénesis del cáncer de colon: iniciación, promoción, progresión y la metástasis. En la etapa de la metástasis el hígado es el sitio que más común es afectado, seguido por el pulmón y el hueso. Adaptado de “Colorectal Cancer: A Review of Carcinogenesis, Global Epidemiology, Current Challenges, Risk Factors, Preventive and Treatment Strategies” (p. 4), por Hossain et al., 2022, *Cancers*, 14(7),

El adenocarcinoma representa el 96% de todos los cánceres del cáncer de colon. A medida que el adenocarcinoma se vuelve invasivo, tiene la capacidad de esparcirse a otras áreas del cuerpo a través de la circulación sanguínea y linfática. La figura 3 representa la formación de la metástasis ya que puede tardar en promedio alrededor de nueve años en desarrollarse, y para que esto suceda, se requieren alrededor de 18 años de cambios progresivos desde un pólipo hasta un cáncer invasivo (**Hossain et al., 2022**).

La progresión del cáncer colon, se han identificado como una serie de alteraciones genéticas cruciales. Estas alteraciones incluyen mutaciones en oncogenes, como ras, y en genes supresores de tumores, especialmente aquellos ubicados en los cromosomas 5q, 17p y 18q. Durante las etapas iniciales del cáncer colon, se observan mutaciones en el gen APC, que resulta en la pérdida de la función de este gen supresor de tumores. Se estima que estas alteraciones del gen APC están presentes en más del 80% de los tumores esporádicos de cáncer de colon (**Testa et al., 2018**).

Después de la presencia de mutaciones en el gen APC, se producen alteraciones en el ADN y la mutación del gen Kras, que actúa como un protooncogén que regula las vías de señalización celular. Se estima que dicha mutación está presente en el 30% al 50% de los casos de cáncer de colon (**Hong, 2018**). A medida que avanza la enfermedad, se produce una pérdida del cromosoma 18q o una supresión del gen DCC, seguida de la pérdida del cromosoma 17p, que afecta al gen supresor de tumores p53. La inactivación de TP53 permite una proliferación celular excesiva que impulsa la progresión del tumor y se asocia principalmente con la evolución de adenoma a cáncer invasivo (**Paper, 2017**).

Adicionalmente, el gen BRAF (Protooncogén B-Raf, serina/treonina quinasa), está presente alrededor del 10 % del cáncer de colon. Este gen impulsa la proliferación, diferenciación, migración, supervivencia y angiogénesis celular. Cuando el gen esta mutado puede asociarse con la metilación del promotor del gen MLH1, que es un gen de reparación de desajuste (MMR), lo que lleva a una disminución de la reparación del ADN y a la acumulación de mutaciones genéticas (**Nakayama et al., 2020**).

El cáncer de colon se caracteriza por ser una enfermedad progresiva que se divide en cuatro etapas (Figura 3). La etapa 0, conocida como carcinoma *in situ* o carcinoma intramucoso, se presenta cuando se identifican células anormales en la mucosa de la pared del colon y se pueden diseminarse al tejido normal cercano. Para la etapa 1, la enfermedad se extiende mediante la mucosa hacia la capa más interna y empieza a extenderse hacia la submucosa y forma parte de la pared del colon. Durante la etapa 2, la enfermedad invade al tejido de la serosa de la pared del colon y recubre los órganos del abdomen, como el peritoneo visceral (Chakrabarti *et al.*, 2020).

La etapa 3 del cáncer de colon se determina por la propagación de las células cancerosas a través de la mucosa propia hacia la subserosa. Esta etapa tiene una particularidad de no haber invadido ningún órgano o tejido, no obstante, se encuentran células cancerosas en uno a tres ganglios regionales (Karim *et al.*, 2017). Finalmente, la etapa 4 se desarrolla la metástasis por lo que el cáncer se propaga a un órgano o un sitio que está lejos del colon como es el claro ejemplo del hígado, pulmones, ovarios o un ganglio linfático distante (Sawicki *et al.*, 2021).

### **1.1.2. Mecanismos Moleculares del Cáncer de colon (CC)**

El cáncer de colon (CC) es una entidad patológica heterogénea causada por la interacción de factores genéticos y ambientales que contribuyen al progreso de la enfermedad. Hay tres vías principales de la patogénesis del CC las cuales son: la convencional de inestabilidad cromosómica (CIN), la inestabilidad de microsatélites (MSI) y las vías del fenotipo metilador de islas CpG (CIMP) (Huang & Yang, 2022).

#### **1.1.2.1. Convencional inestabilidad cromosómica (CIN)**

El 85% de la incidencia de cáncer de colon se debe por el mecanismo molecular de la inestabilidad cromosómica. Esta vía se caracteriza por la alteración de cromosomas en una región específica, lo que contribuye a la mayoría de las transformaciones observadas en adenocarcinoma. Los cambios cromosómicos resultantes pueden inducir modificaciones genéticas significativas en las células tumorales, favoreciendo así su crecimiento y diseminación (Olkinuora *et al.*, 2021).

Este mecanismo se basa en la activación de los genes KRAS (Sarcoma de rata de Kirsten) y BRAF (Serina/treonina-proteína quinasa B-Raf), que son genes conocidos por su potencial cancerígeno. Además, la desactivación de los genes supresores tumorales APC y TP53 aumenta la probabilidad de transformar una célula sana en una célula cancerígena. Adicionalmente, dicho mecanismo conduce la pérdida de heterocigosidad en el brazo del cromosoma 18q, lo cual fomenta la tumorigénesis (**Malki *et al.*, 2021**).

Con respecto al modelo genético existe múltiples pasos, el proceso comienza con el silenciamiento del gen adenomatous polyposis coli (APC). A continuación, se producen mutaciones en el gen KRAS durante la etapa adenomatosa. Posteriormente, se detecta la pérdida del cromosoma 18q y se produce la inactivación del gen TP53. La hibridación genómica basada en matrices y técnicas de polimorfismo de nucleótido único (variantes genómicas en una base única del ADN) lo que ha incurrido en pérdidas en ciertos alelos de los brazos cromosómicos en CC. Al mismo tiempo, se identifican ganancias en el cromosoma 7 y en los brazos cromosómicos 1q, 8q, 12q, 13q y 20q, lo que indica la presencia de oncogenes que estimulan el crecimiento de las células mutadas. Estos cambios genéticos resultan en la disminución de la actividad de los genes supresores tumorales (TSG), lo que da lugar a la formación de células cancerígenas (**Pino & Chung, 2018**).

#### **1.1.2.2. Inestabilidad de microsatélites (MSI).**

En el cáncer de colon, se encuentra la alteración genómica denominada inestabilidad de microsatélites (MSI), que es una característica distintiva de las células cancerosas. Los microsatélites son secuencias de ADN repetitivas que se encuentran dispersas a lo largo de toda la secuencia del ADN. Debido a su naturaleza repetitiva, son propensos a sufrir cambios durante la replicación. Específicamente, son especialmente susceptibles a la acumulación de mutaciones, principalmente debido a la incapacidad de las polimerasas de ADN para unirse eficientemente al ADN durante la síntesis del mismo (**Nojadeh *et al.*, 2018**).

Los microsatélites suelen presentar dos tipos de errores comunes. El primero es el desajuste de bases, en el cual ocurren discrepancias entre las bases que escapan a la capacidad intrínseca de corrección de las polimerasas de ADN. El segundo tipo son los

bucles de inserción-delección, los cuales son secuencias de nucleótidos que se encuentran fuera de la estructura helicoidal normal del ADN, formando horquillas (De'angelis *et al.*, 2018). Estos nucleótidos desapareados ocurren cuando el primer nucleótido y la hebra molde se separan y se vuelven a unir de forma incorrecta en un microsatélite. Cuando se producen inserciones o deleciones en microsatélites ubicados en regiones de codificación del ADN, pueden generarse mutaciones que cambian el marco de lectura, lo que a su vez puede resultar en la truncación de proteínas (Furlan *et al.*, 2022)

Alrededor del 15% de los casos de cáncer de colon presentan inestabilidad de microsatélites (MSI) debido a la supresión epigenética del gen MLH1 o a una mutación hereditaria en uno de los genes MSH2, MSH3, MSH6 o PMS2, que son fundamentales en el sistema de reparación de errores de emparejamiento. Estas proteínas interactúan entre sí como heterodímeros. Cuando se detecta un desajuste, se llevan a cabo varios pasos: MSH2 se une a MSH6 o MSH3 (formando complejos Muts $\alpha$  y Muts $\beta$ , respectivamente), mientras que MLH1 se asocia con PMS2, PMS1 o MLH3 (formando complejos MutL $\alpha$ , MutL $\beta$  o MutL $\gamma$ , respectivamente) (Jones *et al.*, 2017).

El reconocimiento de errores de inserción y eliminación es llevado a cabo por un complejo formado por una proteína Muts y una proteína MutL, que interactúa con el factor de replicación C. La eliminación de los errores se realiza mediante la acción de proteínas como la exonucleasa 1 y el antígeno nuclear de células en proliferación. Finalmente, la reconstrucción y unión de la hebra de ADN es realizada por la ADN polimerasa  $\delta$  y la ADN ligasa. Mutaciones en los genes que participan en el proceso de reconocimiento llevan a una acumulación de errores en el ADN, lo que resulta la inestabilidad de microsatélites (MSI) (Yu *et al.*, 2019).

### **1.1.2.3. Vía del fenotipo metilador de la isla CpG (CIMP).**

El fenotipo de metilación de las islas CpG (CIMP) en el cáncer de colon hace referencia a un cambio epigenético específico que se caracteriza por un aumento en la metilación de las islas CpG presentes en las áreas promotoras de ciertos genes (Rhee *et al.*, 2017). La metilación del ADN es un proceso enzimático mediante el cual se añade un grupo metilo a la posición 5 de la citosina mediante las enzimas ADN metiltransferasas (DNMT), lo que resulta en la formación de 5-metilcitosina. Las islas CpG son segmentos cortos de

regiones que contienen una alta concentración de dinucleótidos CpG y que a menudo se ubican en la región promotora de los genes (**Shen *et al.*, 2017**).

En las células normales, la mayoría de los sitios CpG están fuertemente metilados, mientras que las islas CpG, normalmente ubicadas en las regiones promotoras de los genes, no están metiladas. Sin embargo, después del inicio del cáncer, la hipermetilación dentro de la región promotora puede conducir a la inactivación de los genes supresores de tumores, mientras que la hipometilación global se asocia con inestabilidad genómica y aberraciones cromosómicas (**Kim *et al.*, 2018**).

La expresión regulada al alza de las metiltransferasas de ADN está asociada con CIMP, por lo cual un posible mecanismo subyacente se basa en el silenciamiento de las barreras que inhiben la metilación de islas CpG normalmente no metiladas (**Cohen *et al.*, 2018**).

### **1.1.3. Enfoques terapéuticos.**

#### **1.1.3.1. Tratamientos farmacológicos para los mecanismos moleculares.**

Los mecanismos moleculares han permitido desarrollar estrategias terapéuticas para mejorar los resultados clínicos. Es por ello que se ha basado emplear estrategias para dirigir medicamentos en los mecanismos moleculares del CC.

Para el mecanismo molecular convencional inestabilidad cromosómica (CIN), se ha diseñado moléculas pequeñas, péptidos y anticuerpos para inhibir la vía Wnt/ $\beta$ -catenina, esto con el objetivo de regular la expresión de genes que están involucrados en la regulación del ciclo celular. Como tratamiento en la alteración del mecanismo se ha empleado el uso del cloruro de litio suprimir la supervivencia y proliferación de células. Además, los fármacos antiinflamatorios no esteroideos (AINE) y Celecoxib bloquean la transcripción dependiente de CTNNB1 en el CC y reducen la formación de pólipos en pacientes (**Li *et al.*, 2014**).

Para el tratamiento farmacológico en la inestabilidad del mecanismo de microsatélites (MSI), se basa en la proteína PDL-1 (bloqueo de la muerte programada 1) que está ligada con dicho mecanismo. Según **André *et al.*, (2020)** menciona que es eficaz utilizar la terapia a base de la proteína PDL-1 por lo que se relaciona con el MSI-H que es un tipo de inestabilidad de microsatélite ya que cuando se expresa mayores niveles de PD-L1

conlleva a evadir la respuesta inmunológica. Es por ello que se ha utilizado el medicamento Pembrolizumab, que actúa bloqueando las células tumorales y esto permite que los linfocitos T reanuden su actividad antitumoral.

En el mecanismo del fenotipo metilador de isla CpG (CIMP), se relaciona con la vía del EGFR (receptor del factor de crecimiento epidérmico) por lo cual para el tratamiento se ha utilizado Cetuximab que es un anti-EGFR que inhibe la tirosina quinasa (TKI) dirigidos a las quinasas intracelulares (**Malki *et al.*, 2021**). No obstante, en los tumores que portan mutaciones en los genes KRAS, BRAF son resistentes a la terapia anti-EGFR debido a la activación continua de las vías de señalización aguas abajo de EGFR. Para estos casos existe el uso de Panitumumab, que no provoca citotoxicidad mediada por células dependiente de anticuerpos (**Zhang *et al.*, 2021**).

#### **1.1.4. Búsqueda de tratamiento para el cáncer de colon a través de herramientas Bioinformáticas.**

La bioinformática aplica la combinación de la biología molecular, la computación y la estadística para analizar grandes cantidades de datos biológicos y genéticos. Además, ha permitido utilizar softwares complejos para recuperar, clasificar, analizar, predecir y almacenar datos de secuencias de ADN y proteína (**Islam & Kalimuthu, 2019**). El pipeline bioinformático se basa en las tecnologías ómicas, que incluyen la genómica, transcriptómica y proteómica. Estas técnicas permiten analizar los marcadores de ADN, las transcripciones de ARN, las proteínas y a la vez ayuda a descubrir nuevos biomarcadores para aplicaciones en el pronóstico, diagnóstico y tratamiento del cáncer de colon (**Dalal *et al.*, 2020**).

El pipeline bioinformático usa la genómica para analizar material genético de las células del cáncer, esto permite determinar las mutaciones que impulsan el crecimiento del tumor. Asimismo utiliza el análisis transcriptómica que es un enfoque que se utiliza para explorar biomarcadores genómicos y analizan la expresión de genes en las células del cáncer para determinar los genes que están regulados en el tumor (**Orlov *et al.*, 2021**). La proteómica permite detallar la expresión de proteínas en las células de cáncer de colon, esto conlleva identificar las proteínas que están sobre expresadas o sub expresadas en el tumor (**Yang *et al.*, 2019**).

#### **1.1.4.1. Diseño de fármacos por medio de la Bioinformática**

La Bioinformática utiliza algoritmos basados en simuladores de eventos químicos que permite calcular las propiedades químicas de los átomos y de las moléculas. El objetivo de diseñar los fármacos se basa en las propiedades químicas que presentan tanto el fármaco como su objetivo y a la vez permite calcular la predicción de unión del fármaco a su objetivo (**Sliwoski *et al.*, 2014**).

El método computacional, para el diseño de fármaco se basa en las estructuras y se fundamenta sobre el conocimiento de la estructura tridimensional de la proteína objetivo que normalmente se proporciona por cristalografía de rayos X de alta gama o datos RMN. Estos métodos incluyen el acoplamiento molecular, que tiene como objetivo de identificar los compuestos que se acoplen dentro de un bolsillo objetivo, ya que se dice que si un compuesto encaja en un bolsillo de destino representa un andamio químico prometedor para el diseño de nuevo fármaco (**Palermo & Diego, 2016**).

Este método sirve para localizar sitios específicos de unión entre proteína-ligando y permite analizar las interacciones moleculares. Además, facilita al investigador modificar las características fisicoquímicas que presentan dichas interacciones. Adicionalmente permite conocer los procesos de inhibición de algunas proteínas que se sobre expresan en el cáncer de colon esto mediante el acoplamiento molecular y esto conlleva a reconocer posibles soluciones de tratamiento (**De Vivo *et al.*, 2005**).

El resultado final del método computacional permite reconocer compuestos bioactivos que normalmente incluyen residuos de aminoácidos y expone varios efectos físicos como actividades antioxidantes, antitrombóticas (**Esmeeta *et al.*, 2022**). Una de las ventajas que provee el método computacional es predecir las propiedades farmacocinéticas que permite determinar el movimiento del fármaco a través del cuerpo y a la vez se puede estimar las concentraciones del fármaco en diferentes partes del cuerpo (**Grogan *et al.*, 2022**).

### **1.1.4.2. Cribado Virtual (VS)**

El cribado virtual es un enfoque computacional utilizado para la identificación de fármacos. Su aplicación se lleva a cabo mediante simulaciones por computadora, lo que permite buscar y seleccionar moléculas que posean la capacidad de interactuar con un objetivo molecular específico, generalmente una proteína o un receptor enzimático. En esencia, esta técnica actúa como un filtro al descartar aquellas moléculas que no muestran afinidad hacia la proteína en cuestión, y solo selecciona las moléculas candidatas que tienen el potencial de convertirse en fármacos efectivos (**Zhang et al., 2022**).

Durante el proceso de cribado virtual, se realiza una búsqueda exhaustiva en bibliotecas de ligandos con el objetivo de identificar aquellas estructuras que presentan una mayor probabilidad de interactuar con el objetivo proteico. Estas moléculas se clasifican en función de su capacidad para unirse al receptor proteico. Mediante este proceso, se puede predecir con mayor precisión cuáles ligandos tienen más posibilidades de exhibir una actividad farmacológica relevante para el tratamiento deseado (**Maia et al., 2020**).

### **1.1.4.3. Acoplamiento Molecular**

El acoplamiento molecular (Docking molecular) es una técnica molecular bioinformática que facilita la predicción de interacción entre proteína-ligando y calcula computacionalmente la posición favorable, a partir de representaciones tridimensionales. Es importante mencionar que mientras más estable, y específica sea la unión entre el ligando (fármaco) y su proteico (diana terapéutica) mayor será su actividad biológica. Por lo tanto, el método bioinformático juega un rol importante en el descubrimiento y desarrollo de nuevos principios activos o fármacos (**Ballón Paucara & Grados Torrez, 2019**).

Este método bioinformático se rige en dos algoritmos, el primero es el que se encarga predecir las conformaciones estructurales del ligando y asume el sitio activo que conforman entre proteína- ligando y el otro algoritmo actúa en función de predecir las energías de unión entre el ligando y el receptor para cada una de las conformaciones establecidas por el primer algoritmo. Luego de aplicar estos dos algoritmos se clasificará

las conformaciones de unión del ligando y su receptor de acuerdo a su afinidad de unión **(Agarwal & Mehrotra, 2016)**.

El método consta dos técnicas que son de cuerpo rígido y flexible. El primero se basa en modelo propuesto por Fischer, se fundamenta en el ligando-receptor y se consideran como cuerpos independientes que interactúan entre sí, esto por su forma y volumen. El segundo, se fundamenta en el efecto mutuo del reconocimiento de proteínas y ligandos en la conformación de cada uno de ellos **(Maden *et al.*, 2022)**.

#### **1.1.4.4. Desarrollo del ligando y proteínas.**

Para el uso de las proteínas se requiere una estructura proteica inicial en formato de archivo PDB que incluye un ligando cristalizado. Esta estructura se obtiene en Protein Data Bank y Protein Data Bank in Europe Knowledge Base (PDBE-KB) o Uniprot. Dichas proteínas contienen sus propias estructuras tridimensionales y se diferencian por su difracción de rayos X, resonancia magnética, nuclear (RMN) o crio-microscopía electrónica (CryoEM) **(Guide, 2015)**.

Este proceso se debe primeramente eliminar todas las interferencias una de ellas es las aguas (excepción de las con los que están ancladas los metales) que se asemejen por el átomo de oxígeno ya que estas no tienen hidrógenos unidos. No obstante, si se quiere mantener las aguas se debe agregar hidrógenos en el desarrollo de la preparación. Adicionalmente, se debería simplificar si existe los multímeros con sitios de unión duplicados donde se debe eliminar los sitios de unión redundantes **(Madhavi Sastry *et al.*, 2013)**.

Se debe tener en cuenta de corregir los residuos que por lo general son relativamente son inofensivos si están alejados del sitio activo. De igual forma se debe comprobar la estructura de la proteína en busca de iones metálicos y cofactores. Donde sí existe se debe eliminar enlaces para luego ajustar las cargas de los átomos que están unidas al metal. Estos procedimientos analizados se pueden realizar en los programas Pymol y Chimera que permite evaluar las moléculas, mapas de densidad, ensamblajes supramoleculares, alineaciones de secuencias, resultados de docking, trayectorias y conjuntos conformacionales **(Bender *et al.*, 2021)**.

Para la selección del ligando, se debe igual realizar un bosquejo en Drugbank, las cuales propone ligantes que se acopla a la proteína que se va a analizar lo cual se debe tener igual presente determinar mediante difracción de rayos X, resonancia magnética nuclear. Se puede decir que para trabajar con los ligandos se puede usar las estructuras tridimensionales como son las base de datos CCDC o Pubchem (Huggins *et al.*, 2012).

#### 1.1.5.5. Acoplamiento rígido.



**Figura 4.** *Modelo de acoplamiento molecular rígido*

Nota. La figura muestra modelo llave (receptor) y candado (receptor), en donde se refiere a un acoplamiento rígido de receptores y ligandos para localizar la orientación correcta de la llave para así poder abrir la cerradura. Cabe mencionar que es un modelo de complementariedad geométrica. Adaptado de “Progress in molecular docking” (p.2.) por Fan *et al.*, 2019, Quantitative Biology, 7(2).

El acoplamiento rígido (Figura 4) es una técnica que no requiere cálculos, pero puede a veces obtener errores al no considerar que las dos especies pueden producir una gran cantidad de conformaciones acopladas. Estas conformaciones se reorganizan aprovechando la energía libre y es fundamental resaltar que, al considerar el ligando y el receptor como cuerpos rígidos, el espacio de búsqueda es específico ya que solo posee tres grados de libertad de traslación y tres de rotación. Asimismo, se asume que la estructura del acoplamiento que es receptor y el ligando permaneces invariable (Fan *et al.*, 2019;Raval & Ganatra, 2022).

#### 1.1.4.6. Interacción de complejo proteína con el ligando

Es impredecible la puntuación de la interacción proteína con el ligando, ya que este definirá la drugabilidad del ligando que se ha unido a la proteína. Estas puntuaciones estiman la energía del enlace del complejo evaluando las interacciones intermoleculares

de solvatación efectos electrostáticos y entrópicos, para llegar lo más cerca posible de la energía de enlace real en un tiempo mínimo (**Dar & Mir, 2017**). Una puntuación ideal en las interacciones proteína-ligando sea apto para reconocer la posición de unión de un ligando en la superficie de la proteína entre las posiciones de señuelo y predecir con precisión la afinidad de unión (o la energía libre de unión) para que las moléculas activas puedan ser discriminados de los no activos (**Wang & Lin, 2017**).

#### **1.1.4.7. R como lenguaje para el análisis de expresión diferencial.**

La expresión diferencial de genes permite detectar qué genes muestran variaciones significativas en su expresión durante una enfermedad. De esta manera, el análisis diferencial facilita la identificación de las disparidades entre la expresión génica en condiciones normales y en estados de enfermedad, lo que contribuye a una mejor comprensión de la patología de las enfermedades (**Rodriguez-Esteban & Jiang, 2017**).

El lenguaje R utiliza paquetes Bioconductor para analizar datos RNA-seq (secuenciación de ARN) para generar los resultados de expresión diferencial. Es importante que al analizar los DEG en R se debe normalizar los datos ya que con ello se pueda que los niveles de expresión de genes ortólogos sean comparables entre diferentes especies. Los paquetes que normalmente se utiliza es limma que determina la alta correlación entre genes que puede ocurrir en los datos de micromatrices (**Zhou *et al.*, 2019**).

Además, R cuenta con Volcano-plot, que es un diagrama de tipo de dispersión que muestra la significación estadística frente a la magnitud de cambio. Tiene la particularidad de identificar los genes con grandes cambios de pliegue que también son estadísticamente significativos (**Guan & Tian, 2017**). El pheatmap se utiliza para representar mapas de calor donde se expresa los genes que muestra cambios significativos entre dos o más grupos experimentales, como muestras sanas frente a muestras enfermas o muestras de control frente a muestras de tratamiento. Adicionalmente, pheatmap utiliza el algoritmo de agrupación jerárquica que reestablece filas y columnas de calor con la finalidad de aglomerar genes con patrones de expresiones similares (**Evans *et al.*, 2018**).

### **1.3. Objetivos**

#### **1.3.1. Objetivo General**

- Evaluar posibles dianas terapéuticas para el cáncer de Colon aplicando pipeline bioinformático como base para el diseño de fármacos asistido por computador.

#### **1.3.2. Objetivos Específicos**

- Comprender las bases genéticas y moleculares del Cáncer de colon enfocados en los genes relacionados.
- Estudiar la expresión diferencial de los genes que se expresan o inhiben en el cáncer de colon.
- Aplicar Pipeline Bioinformático para la identificación de dianas terapéuticas en el cáncer de colon.
- Valorar las propiedades toxicológicas y farmacodinámicas de los compuestos bioactivos y su potencial uso para el diseño de nuevos fármacos.

## CAPÍTULO II

### METODOLOGÍA

#### 2.1. Materiales y equipos

##### 2.1.1. Materiales

##### 2.1.1.1. Software de bioinformática

- MalaCards
- HARMONIZOME
- OMIM
- KEGG
- Cytoscape (GeneMania)
- GEO (Gene Expression Omnibus)
- Lenguaje de programación R
- Drugbank
- Protein Data Bank
- ProteinPlus
- Chimera 1.16
- AutoDock Tools
- AutoDock Vina
- MTIOpenScreen
- STITCH
- Discovery Studio
- Avogadro
- pkCSM
- AlphaFold
- GeneCards
- Protein Data Bank in Europe Knowledge Base
- Vega

##### 2.2.2 Equipos

Laptop personal

## **2.3 Métodos**

### **2.3.1. Identificación de las Bases Genéticas y Moleculares del Cáncer de Colon.**

En la presente investigación, se empleó una búsqueda sobre las bases moleculares y las alteraciones bioquímicas relacionadas con el cáncer de colon en las siguientes bases de datos como es Genecards, PUBMed, OMIM, MalaCards, Harmonizome y NCBI. Esto se realizó con finalidad de recopilar información sobre los genes implicados en la enfermedad.

### **2.3.2. Expresión diferencial de los genes que se expresan o inhiben en el cáncer de colon.**

#### **2.3.2.1. Búsqueda de datos para Cáncer de Colon.**

Se busco los datasets mediante la metodología de **Dong et al., (2020)**, donde otorga los siguientes criterios para la búsqueda las cuales son:

- Conjuntos con muestras enfocadas en el cáncer de colon.
- Conjuntos de datos con un respaldo en publicaciones indexadas en bases de datos confiables.
- Estudios que tengan información sobre la tecnología y plataforma usada para su creación.

Siguiendo los criterios mencionados, se accedió al sitio web de la base de datos GEO (Gene Expression Omnibus) y se ingresó el término "colon cáncer". A continuación, se seleccionó las siguientes opciones: el uso de GEO DataSets, el organismo Homo sapiens, el tipo de experimento que implicó la generación de perfiles de expresión mediante arrays, y se consideró el diseño experimental que incluyó tanto pacientes normales como pacientes afectados por la enfermedad. Posteriormente, se descargó el archivo en formato CEL para su análisis preliminar.

### **2.3.2.2. Análisis de expresión diferencial**

Con el dataset previamente ya obtenido, se llevó a cabo el análisis de expresión de los genes, se utilizó el lenguaje de programación R. No obstante, se tomó en cuenta de instalar previamente los paquetes de Bioconductor como es: Rbowtie2, ggplot2, MA Plots, pheatmap, sva y lima.

Como primer es normalizar los genes de expresión por medio del paquete “Rbowtie2”. Este paquete permitió realizar el análisis de expresión diferencial para cada uno de los conjuntos de datos como es los datos de expresión de colon humano no afectado y canceroso, en lo cual lo comparo entre ellos.

Para la determinación de genes de expresión diferencial (DEGs) se utilizó el paquete “ggplot2” donde se generó un diagrama de tipo volcán (Volcano plots). Este diagrama permitió identificar la expresión de genes que están en alza y baja en las células normales de colon y células cancerosas de colon.

Asimismo, se realizó un mapa de calor (heatmap) mediante el paquete “heatmap”. en R. Posteriormente con ayuda del paquete “sva” y “limma” se efectuó una corrección por lotes y a la par una normalización de los arrays, esto permitió eliminar la heterogeneidad de los conjuntos de datos. Después, se realizó un análisis de los componentes de expresiones de genes que se obtuvo para así comparar los diferentes DEGs consensuados entre los grupos control y los de análisis.

### **2.3.3. Pipeline Bioinformático para la identificación de dianas terapéuticas**

#### **2.3.3.1. Base de datos MalaCards**

La base de datos Malacards permitió llevar a cabo una búsqueda acerca de los genes que afectan al cáncer de colon. Como resultado de este análisis, se identificó la entrada más relevante para su inclusión en la investigación, la cual corresponde al "Adenocarcinoma Colon". Esta entrada específica se caracteriza por contener un conjunto de 119 genes relacionados con la enfermedad y un MIFTS de 55. Dicha puntuación de MIFTS (siglas en inglés para "Minimum Information for the functional Annotation of Transcripts and

their Protein Products") refleja el nivel de información acumulada en la base de datos en relación con el cáncer de colon.

#### **2.3.3.2. Base de datos Harmonizome**

Dentro de esta base de datos Harmonizome existe un gran conjunto de genes relacionadas con la enfermedad es por ello por lo que se eligió la opción "Gene Sets" con la entrada de la enfermedad de "Adenocarcinoma Colon", esto con el objetivo de evaluar a la par con la base de datos Malacards las similitudes y congruencias genéticas. Dentro de este procedimiento se escogió los genes que comparten interacciones y los que obtuvieron una mayor puntuación esto con la finalidad de desarrollar un análisis posterior.

#### **2.3.3.3. Base de datos OMIN**

La base de datos de OMIM, se ingresó la entrada de "colon cáncer", se encontró la opción de COLORECTAL CANCER; CRC y presentó veinte y ocho genes que están establecidos en dicha enfermedad. Con estos genes encontrados, se correlacionó con las bases de datos de Harmonizome y Malacards para establecer los genes que están más relacionados con la enfermedad.

#### **2.3.3.4. Base de datos KEGG**

Se han obtenido las vías de activación de las señales del cáncer de colon, así como información sobre las rutas metabólicas relacionadas con esta enfermedad. Además, se ha recopilado información sobre los genes que participan en las cascadas que provocan alteraciones bioquímicas y desregulaciones en el proceso de desarrollo de la carcinogénesis en el colon.

#### **2.3.3.5. Selección de los genes con mayor relevancia.**

En la selección de los genes se basó primeramente en la obtención de las puntuaciones estándares de cada uno de los genes que están establecidas en las bases de datos como es Malacards y Harmonizome, además se tomó en cuenta la relación del gen con la

enfermedad. De igual modo se consideró la información acerca de las alteraciones bioquímicas.

Luego de la revisión, se procedió a seleccionar los genes que se consideran más significativos en el cáncer de colon, en donde se clasifico según las alteraciones bioquímicas y sus puntuaciones. Se encontró diferentes alteraciones bioquímicas en los genes, sin embargo, se basó en el score que obtuvieron los genes para así descartar las diferentes alteraciones por lo cual se analizó aquellos que presentaron mayor score por lo cual se clasifico en: (regulación) actividad transcripcional, activación descontrolada de señalización MAPK.

De los 119 genes que se encontró en la base de datos de Malacards se correlacionó con el score que presentaba cada gene en la base de datos de Harmonizome, para así genes más representativos que están asociados en la enfermedad por lo cual se pone los datos en hoja de cálculo con sus respectivas puntuaciones e integración dentro de las bases de datos y fueron organizadas en función del nombre del gen, su puntuación y alteración bioquímica.

#### **2.3.3.6. Integración de los datos**

Para la integración de datos, se utilizó el software Cytoscape para el diseño de redes biológicas, en donde se utilizó el plugin GeneMania que permitió la red de interacción de los genes más representativos para el cáncer de colon. En este software, se realizó los siguientes pasos para la integración de datos:

- Se ingresó los nombres de los genes de acuerdo a las alteraciones bioquímicas que presentan y escoger la opción de Homo sapiens.
- Luego se generó la red biológica mediante del plugin GeneMania.

Posteriormente, se escogió los genes que obtuvieron el mayor número de interacciones en los nodos y a la vez las puntuaciones más altas y se examinó los bordes para identificar los genes que tengan mayor iteración.

### **2.3.3.7. Validación de las dianas seleccionados**

La base de datos de DrugBank, contiene información acerca de dianas y medicamentos que actúan en el cáncer de colon. Esta base posee dos partes, la primera es información química de los fármacos, mientras que la segunda corresponde a la información de proteínas o dianas terapéuticas.

Esta base de datos, se efectuó la búsqueda de genes que obtuvimos previamente que presentaron mayor relevancia en la enfermedad, para así encontrar posibles opciones farmacológicas y ligandos. Donde se consiguió varios fármacos en fase de desarrollo y en fase experimental.

Posteriormente, en Protein Data Bank se encontró los cristales de las proteínas de los genes, además, se tomó en cuenta los parámetros del R-value Work y del R-Free que no debe ser mayor de 0.3 (**Gore *et al.*, 2017**). Sin embargo, cabe mencionar que en algunos casos no se encontraron todos los cristales es por ello que se utilizó el modelamiento *in silico* utilizando AlphaFold apoyado de alineamiento estructural y de secuencia con proteínas homologas (**Evans *et al.*, 2022**).

### **2.3.4. Prueba de viabilidad de las dianas terapéuticas**

En la identificación, de la drugabilidad para las proteínas obtenidas se realizó en la web Protein Plus de la Universidad de Hamburgo y ChEMBL del European Bioinformatics Institute (EBI). Esta web permitió mapear la proteína mediante el código que proporciona el PDB. No obstante, si la proteína que se obtuvo en AlphaFold se debe solo subir el archivo en formato PDB en DoGSiteScorer.

Una vez que se ingresó el código de la proteína se eligió la opción de DoGSiteScorer, esta herramienta permitió la identificación de posibles bolsillos y subbolsillos de unión en estas estructuras proteicas. Asimismo, presentó un modelado en 3D por lo cual se obtuvo el grado de drugabilidad y a la vez se obtiene los pockets. Estos pockets contienen los aminoácidos residuales que presentan las proteínas.

Los parámetros evaluados en el DoGSiteScorer fueron la capacidad volumétrica del gen MLH1 que presentó 1580.8 Å<sup>3</sup>, la profundidad de 1921.99 Å<sup>2</sup> y la puntuación de drugabilidad de 0.79. De igual manera, en el gen SMAD2 se obtuvo la capacidad volumétrica de 529.13 Å<sup>3</sup>, la profundidad de 984.5 Å<sup>2</sup> y la puntuación de drugabilidad de 0.77. Igualmente, en el gen CASP9 su capacidad volumétrica es de 1131.99 Å<sup>3</sup>, la profundidad de 1620.98 Å<sup>2</sup> y la puntuación de drugabilidad de 0.78.

### **2.3.5. Búsqueda de moléculas con posible interacción con las dianas.**

La web Protein Data Bank permitió obtener cristales validados en formato pdb. Luego de obtener las proteínas, se utilizó el software Chimera 1.16 con la finalidad de eliminar todo lo que no es parte de la proteína como es el agua y los complejos que normalmente suelen tener los cristales que se encuentran en las bases de datos.

En la web Pubchem, se descargó los posibles ligandos que se va ocupar para luego utilizar el software Avogadro para poder bajar la energía para obtener un mejor acoplamiento. Posteriormente, se empleó Autodocktools para calcular las dimensiones del Grid-Box tanto para las dimensiones como para la localización del sitio de unión con los restos de aminoácidos puntuales especificados con este mismo software, de modo que se situó la Grid-Box dentro del área de dominio más probable para que exista una interacción de proteína-ligando.

Las coordenadas que se obtuvieron en la Grid-Box en la proteína de MutL Protein Homolog 1 fueron:

- Coordenadas del centro de la cuadrícula: x\_20.958, y\_-24.486, z\_9.442
- Tamaño del espacio de búsqueda: x\_78, y\_92, z\_92

Para la proteína de Mothers Against Decapentaplegic Homolog 2 se obtuvo la Grid-Box de:

- Coordenadas del centro de la cuadrícula: x\_25.63, y\_7.226, z\_4.078
- Tamaño del espacio de búsqueda: x\_100, y\_68, z\_64

En las coordenadas que se obtuvieron en la Grid-Box en la proteína Caspasa 9 de fueron:

- Coordenadas del centro de la cuadrícula: x\_0.37, y\_-4.604, z\_-14.941
- Tamaño del espacio de búsqueda: x\_108, y\_86, z\_76

### **2.3.6. Cribado Virtual**

Para el cribado virtual, se tomó en cuenta el archivo que se obtiene en el Autodocktools de la Grid-Box la cual fue de formato pdb. A continuación, se utilizó la web MTIOpenScreen, que permitió utilizar el empleo de Autodock Vina para realizar un cribado virtual. Seguidamente, se empleó los cálculos comenzando desde los datos de las moléculas menores que se incorporan por defecto en el software (**Labbé *et al.*, 2015**). Además, se excluyó, los disolventes que están presentes en la estructura 3D de la proteína en formato pdb y se insertó los datos de las moléculas pequeñas en formato sdf.

Luego se incluyó las coordenadas del sitio de unión y sus dimensiones a través de la opción Custom parameters en función de los cálculos enfocados a esa zona. Se obtuvo estos parámetros mediante la extracción del inhibidor de la estructura 3D en formato pdb. Los resultados mostraron una lista descargada de compuestos de la base de datos según la energía de unión proteína-ligando en formato .pdb, lo cual facilitó la identificación de la unión con las proteínas.

Además, para una mejor comprensión de interacción entre proteína con fármaco se utilizó la base de datos STITCH por lo cual se integró los nombres de los genes de acuerdo con la alteración bioquímica de la regulación de transcripción en la opción de multiple ítems y a la vez se seleccionó el organismo Homo Sapiens, esto con la finalidad de generar las redes de interacción entre los productos químicos y las proteínas.

### **2.3.7. Acoplamiento Molecular**

Se realizó el acoplamiento molecular por medio del software autoDockTools, primero se subió la proteína para agregar los hidrógenos polares, para luego eliminar los hidrógenos no polares y agregar la carga de Kollman, seguidamente se guardó en formato pdbqt. El

ligando se cargó al software y se agregó igualmente todos los hidrógenos y se eliminó los no polares y se agregó la carga de Compute Gasteiger para luego guardar en formato pdbqt.

Para la ejecución del acoplamiento, se utilizó la opción de acoplamiento rígido para agregar la proteína y luego se escogió el ligando, y guardar en formato dpf. Después, se analizó las conformaciones de afinidad de unión que se obtuvieron. Luego se escogió la mejor conformación de afinidad para guardar en formato pdb para su respectivo análisis.

Para una mejor comprensión de los resultados obtenidos de acoplamiento molecular se ocupó el software Discovery Studio, lo cual se ingresó la conformación de la mejor afinidad de unión de formato pdb. Luego se debe definir el receptor proteico con su ligando para agregar la interacción. Posteriormente en la opción display receptor surfaces se escogió la opción H-bond. Esto se realizó con la finalidad de ver los aminoácidos de interacción. Finalmente se guarda las imágenes en formato image files para el respectivo análisis.

### **2.3.7. Análisis de toxicidad de las moléculas.**

Para el análisis de la actividad biológica de los ligandos formados por el complejo ligando-proteína, se empleó el software Vega esto con la ayuda de modelos QSAR en función de análisis fisicoquímicos. Luego se realizó pruebas de mutagenicidad, carcinogénesis y hepatotoxicidad utilizando los códigos SMILES de dichos fármacos. Finalmente, se realizó el análisis de las propiedades de las moléculas con el programa pkCSM, en el cual se realizó pruebas ADMET (Absorción, Distribución, Metabolismo, Excreción y Toxicidad) utilizando también los códigos SMILES de las moléculas identificadas.

## CAPITULO III

### RESULTADOS Y DISCUSIÓN

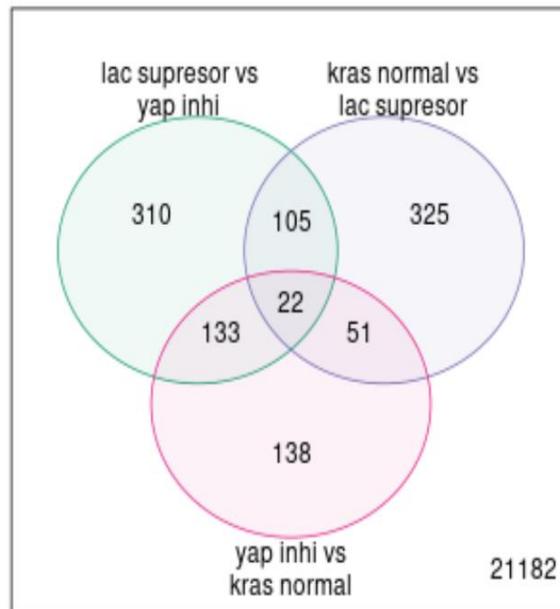
#### 3.1 Análisis y discusión de los resultados.

##### 3.1.1. Análisis de expresión diferencial

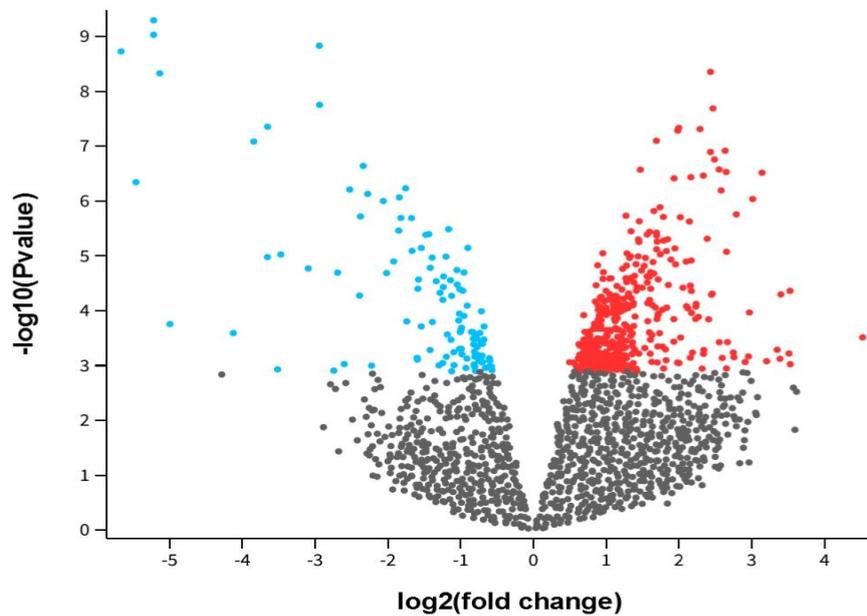
El análisis de expresión diferencial de genes (DEG) se utiliza para examinar los patrones de expresión de los genes y descubrir los mecanismos biológicos involucrados en enfermedades complejas. En términos generales, los perfiles de expresión genética se caracterizan por tener una gran cantidad de genes (decenas de miles) y fuertes correlaciones entre ellos (**Abbas & El-Manzalawy, 2020**). El presente estudio, se empleó para identificar los DEG en muestras de líneas celulares que tiene la expresión normal del gen KRAS en células normales del colon y muestras ya en expresión adenocarcinomas colon en función del perfil de expresión génica de GEO: GSE55942.

El análisis de expresión de la expresión se estudió la expresión de las mutaciones activadoras del protooncogén KRAS. Tiene la función de iniciar la vía de señalización de MAPK (proteína quinasa activada por mitógenos), el fosfatidilinositol 3-quinasa (PI3K). La supresión de la vía ayuda a prevenir la iniciación del tumor. Cualquier alteración en la expresión o función de este gen podría tener implicaciones en la vía MAPK y, en algunos casos, podría estar relacionada con trastornos genéticos o enfermedades relacionadas con la división celular anormal (**Liu *et al.*, 2022**).

Se comparo la expresión de KRAS en 12 muestras de células, en las cuales se hacen 2 réplicas de que no está inhibido y suprimido la función del KRAS. El estudio de (**Guo *et al.*, 2019**) utiliza el gen KRAS por medio de LacZ y YAP1 para identificar los genes que están involucrados tanto en el mecanismo de transcripción cuando no está en inhibido y esta suprimido. En la figura 5 representa el diagrama de Venn en donde se puede observar los genes totales que muestran el DEG que es de 21182. En estos comparten similitud 22 de ellos, lo cual los 105 representa que se puedes analizar para comprender las funciones biológicas, rutas metabólicas.



**Figura 5.** Diagrama de Venn de las muestras de células normales y supresión de la función.



**Figura 6.** Volcano plot que muestra una descripción general de la expresión diferencial de todos los genes.

Nota: El umbral en la gráfica del volcán fue  $-\log_2$  y  $\log_2$  cambio de pliegue. El rojo indica genes regulados al alza y el azul indica genes regulados a la baja.

En las células de cáncer de colon (HT116) se obtuvo la supresión de la función del KRAS, lo cual se reveló un total de 570 genes expresados diferencialmente en el cáncer de colon. De estos se encontró 105 genes regulados al alza y 465 genes regulados a la baja. Los genes que se encuentran en la parte superior de la figura 6, en el eje x representan aquellos con cambios altamente significativos al  $p < 0.05$  relacionados con las diferencias biológicas.

Los genes que se sobre expresaron están en la tabla 1 KRAS, MAP3K5, MAPK14, ERK, CASP9. Estos genes son de gran importancia ya que participan en la vía MAPK y estos genes servirán para el estudio de la investigación de la vía de señalización de apoptosis.

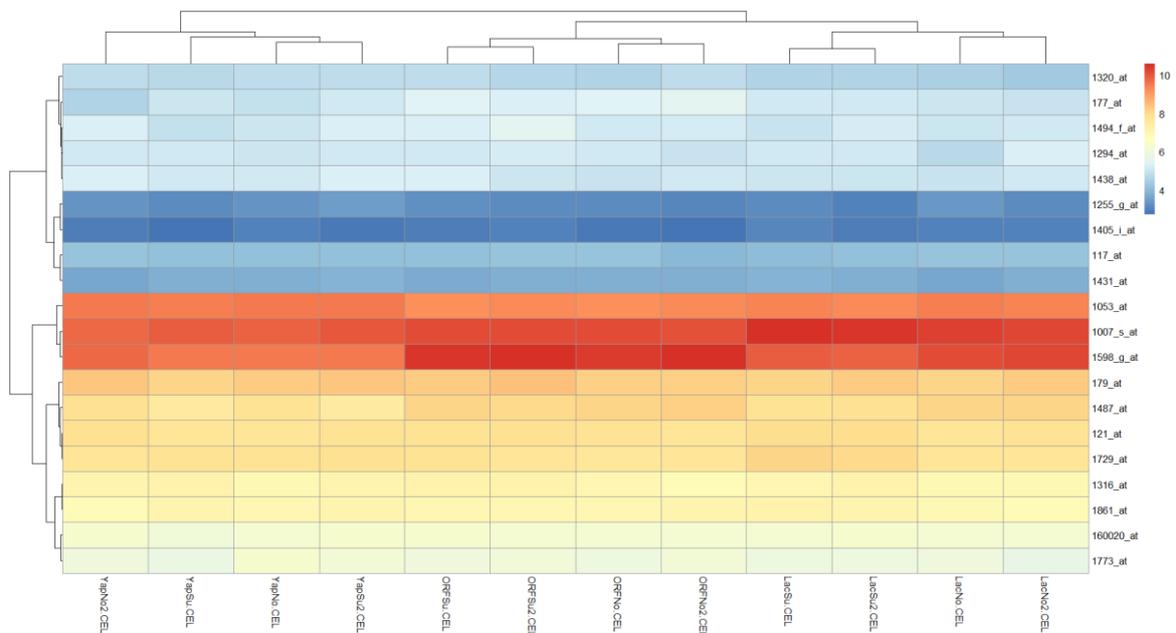
**Tabla 1.** Principales genes diferenciales.

#	Gen	Log FC	Valor-P	Valor P-ajustado
1	KRAS	14.15030	2.966814e-30	7.453210e-28
2	MAOB	14.36048	1.080603e-30	7.161925e-28
3	CDH1	14.31419	1.630872e-30	7.161925e-28
4	JUN	14.28529	1.530220e-30	7.161925e-28
5	CYR61	14.28024	1.361202e-30	7.161925e-28
6	MAP3K5	14.27770	1.415675e-30	7.161925e-28
7	CTNNB1	14.24173	1.950912e-30	7.161925e-28
8	MAPK14	14.20448	1.897595e-30	7.870332e-28
9	CASP9	14.19242	4.390918e-30	7.161925e-28
10	ERK	14.18807	1.588781e-30	7.453210e-28

La figura 7 se analizo los 20 genes diferenciales, de los cuales muestran de los cuales se representan en nivel alto 3 genes en nivel alto, en medio se presenta 8 genes y en cuanto a nivel bajo se encuentra 9. Cada columna indica un conjunto de datos y cada fila indica un gen. Se analiza que cada columna representa lo siguiente de las 12 columnas se subdivide en 4 cada uno; gen KRAS, gen Lacz, y el codificador de Yap. La cual de los 4 que se divide en dos uno no suprime la función y el otro actua en forma de supresor de la actividad del gen KRAS. Para identificar el nombre de los genes se analiza mediante el

codigo de proporcionado de R la cual otorga con `print_genenames` la cual proporciona los nombres de los genes.

En la tabla 1, muestra los genes que estan sobreexpresados ya que cuando se obtiene un valor positivo en el logFC, asimismo, los principales genes estan sobreexpresados y esto conlleva a que estos genes podrían desempeñar un papel importante en la viabilidad, el crecimiento y la propagación de las células cancerosas, lo que contribuye a las alteraciones fenotípicas. El gen KRAS al ser un oncogen cuando se sobreexpresa se relaciona con el desarrollo del cáncer. Además, al evaluar el valor-p de los resultados obtenidos se consideran que los genes está diferencialmente expresado entre los grupos que se compararon ya que es  $< 0.05$  (Y. Chen *et al.*, 2016).



**Figura 7.** Mapa de calor de los primeros 20 genes diferenciales

Los genes que inhiben la función de KRAS actúan como inhibidor PI3K, por lo tanto, el gen YAP1 al ser un coactivador transcripcional actúa de forma parcialmente para activar las vías de PI3K. No obstante, la vía PI3K no toma en cuenta la capacidad de inhibir YAP1 para promover la supervivencia celular del KRAS. El estudio sugiere Shen *et al.*, (2022) que la supresión de la vía MAPK de KRAS inhibe el crecimiento del cáncer de colon ya que tiene un potencial oncogénico en el cáncer de colon. Esta inhibición de la expresión de la vía MAPK activada descontroladamente puede ser una estrategia anticancerígena prometedora.

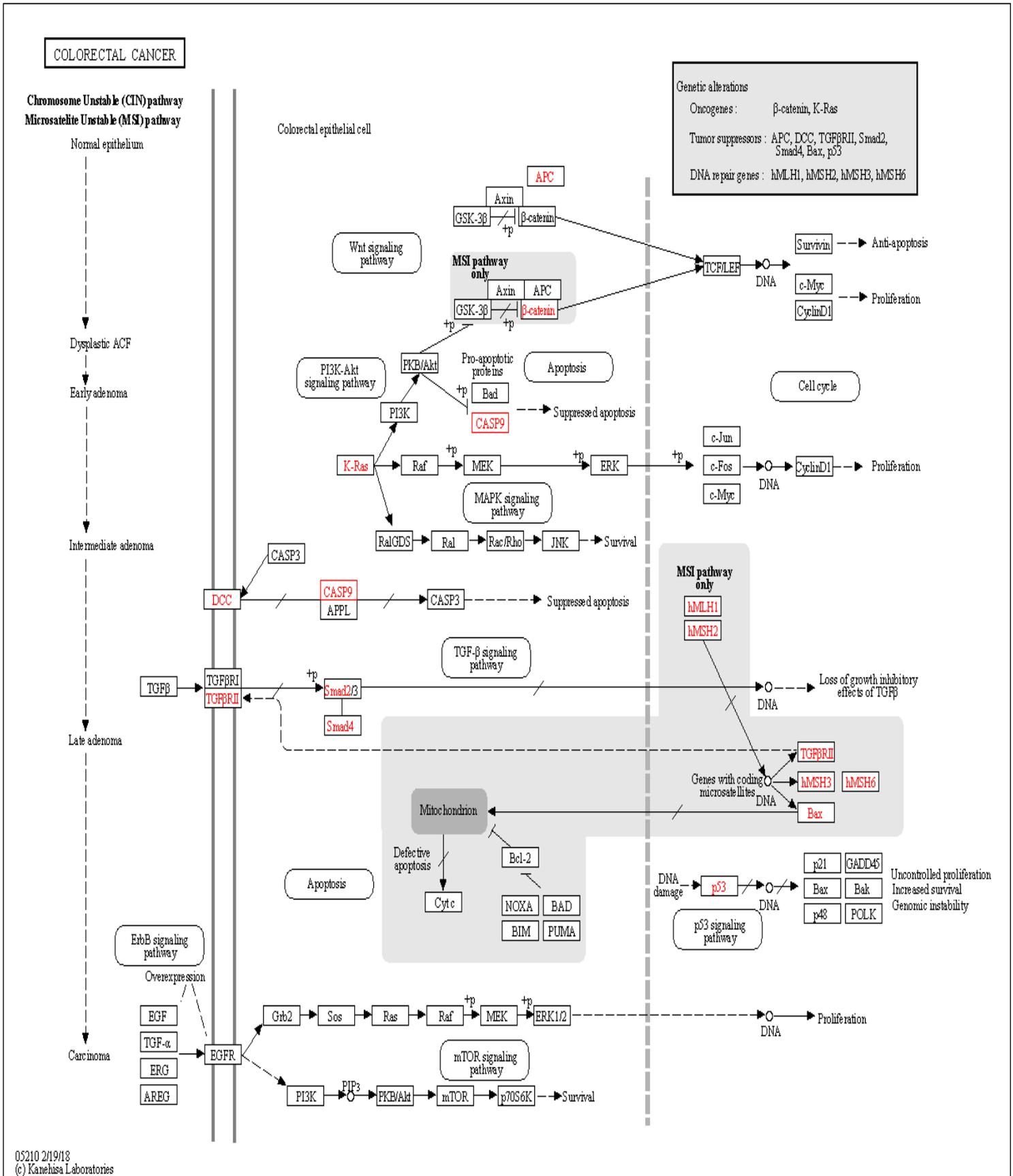
### **3.1.2. Análisis de las bases de datos.**

Se seleccionó los 119 genes encontrados en la base de datos Malacards, esto como punto de partida para la investigación para el tratamiento del cáncer de colon. Estos genes fueron seleccionados a base de las interacciones analizadas que involucran la relación "gen-enfermedad" y que son considerados elementos clave por este sitio web. Además, estos genes constan de un score que indican la incidencia que está ligada al cáncer.

A pesar de los hallazgos en la base de datos mencionada anteriormente, hay investigaciones en la misma área que sugieren la existencia de otros genes que podrían estar involucrados en las complicaciones asociadas con el cáncer de colon (CC). Esto conllevó a una búsqueda exhaustiva en otras bases de datos como es OMIM, Genecards para verificar dicha información. Como resultado, se confirmó que varios de los genes previamente identificados están directa o indirectamente relacionados con el cáncer de colon.

Posteriormente, para la selección de los genes de élite se correlacionó con la base de datos Harmonizome para así agruparlos y relacionarles con la base de datos de OMIM para saber si se correlaciona directamente con la enfermedad. Por otro lado, en el análisis de las rutas metabólicas que presenta la enfermedad se efectuó en la base de datos KEGG, tal y como se presenta en la figura 8, en donde se observa las diferentes vías moleculares que participan los diferentes genes que están marcados de color rojo, estos son los responsables de provocar el CC.

La figura 8, se puede ver las diferentes vías de señalización molecular que participan los diferentes genes que actúan en el CC, sin embargo, para la investigación terapéutica se tomó en cuenta las fases de tempranas de la enfermedad, es por ello que se sobre sale el estudio de la alteración de la vía MAPK con su participación en la vía apoptótica. Además, es imprescindible analizar la función de la regulación de transcripción con sus respectivos genes ligados. Como objetivo terapéutico en la vía MAPK del análisis del gen CASP9 y de la regulación de transcripción como es SMAD2 y MLH1.



**Figura 8. Ruta disfuncional metabólica del cáncer de colo**

**Tabla 2.** Genes representativos en la aparición de cáncer de colon.

#	Genes	Alteración Bioquímica	Proteína	DATABASE			
				MalaCards	OMIM	Harmonizome	KEGG
1	APC	Regulación transcripcional-supresor de tumores	APC Regulator of WNT Signaling Pathway	420.39	si	1.66375	si
2	MSH6	Reparación del ADN	MutS Homolog 6	406.48	si	1.65655	si
3	CTNNB1	Actúa como coactivador de los factores de transcripción.	Catenin Beta 1	26.94	si	1.11982	si
4	KRAS	Activación anormal de la vía MAPK	KRAS Proto-Oncogene, GTPase	26.3	si	1.00963	si
5	TP53	Regulación transcripcional-supresor de tumores	Tumor Protein P53	23.15	si	1.02187	si
6	SMAD4	Regulación transcripcional	SMAD Family Member 4	23.01	si	0.95996	si
7	CDH1	Regulación transcripcional	Cadherin 1	20.29	no	0.785165	si
8	SMAD2	Regulación transcripcional	Mothers Against Decapentaplegic Homolog 2	20.27	no	0.596012	si
9	CASP3	Alteración en la vía MAPK	Caspase 3	20.15	si	0.828639	si
10	BRAF	Activación anormal de la vía MAPK	B-Raf Proto-Oncogene, Serine/Threonine Kinase	19.38	si	0.858543	si

**Tabla 3.** *Genes elite destacados para la investigación y establecimiento de redes biológicas de interacción por Cytoscape.*

#	Genes	Alteración Bioquímica	Proteína
1	APC	Regulación transcripcional-supresor de tumores	Adenomatosis Poliposis Coli Tumor Suppressor
2	MSH6	Regulación transcripcional	MutS (E. Coli) Homólogo 6
3	CTNNB1	Activación anormal de la vía MAPK	Catenin Beta 1
4	KRAS	Activación anormal de la vía MAPK	KRAS Proto-Oncogene, Gene, GTPase
5	BAX	Activación anormal de la vía MAPK	BCL2 Associated X, Apoptosis Regulator
6	TP53	Regulación transcripcional-supresor de tumores	Tumor Protein P53
7	SMAD4	Regulación transcripcional	SMAD Family Member 4
8	PIK3CA	Activación anormal de la vía MAPK	Phosphatidylinositol-4
9	ABCB1	Regulación transcripcional	ATP Binding Cassette Subfamily B Member 1
10	TYMS	Regulación transcripcional	Thymidylate Synthetase
11	CDH1	Regulación transcripcional	Cadherin 1
12	SMAD2	Regulación transcripcional	SMAD, Mothers Against DPP Homolog 2
13	PTGS2	Activación anormal de la vía MAPK	Prostaglandin-Endoperoxide Synthase 2
14	ODC1	Regulación transcripcional	Ornithine Decarboxylase 1
15	BRAF	Activación anormal de la vía MAPK	B-Raf Proto-Oncogene, Serine/Threonine Kinase
16	KRT20	Regulación transcripcional	Keratin 20
17	ST6GAL1	Activación anormal de la vía MAPK	ST6 Beta-Galactoside Alpha-2,6-Sialyltransferase 1
18	CDKN1A	Regulación transcripcional	Cyclin Dependent Kinase Inhibitor 1A

19	SYNE1	Activación anormal de la vía MAPK	Spectrin Repeat Containing Nuclear Envelope Protein 1
20	LGALS3	Regulación transcripcional	Galectin 3
21	TNF	Regulación transcripcional	Tumor Necrosis Factor
22	CCND1	Regulación transcripcional	Cyclin D1
23	CASP9	Activación anormal de la vía MAPK	Caspase 9
24	CCNB1	Regulación transcripcional	Cyclin B1
25	TGFB1	Regulación transcripcional	Transforming Growth Factor Beta 1
26	MLH1	Regulación transcripcional	MutL Homolog 1
27	PPARG	Activación anormal de la vía MAPK	Peroxisome Proliferator Activated Receptor Gamma
28	XIAP	Regulación transcripcional	X-Linked Inhibitor Of Apoptosis
29	HRAS	Regulación transcripcional	HRas Proto-Oncogene, GTPase
30	CASP3	Activación anormal de la vía MAPK	Caspase 3

---

De los 119 genes de élite (anexo E) se escogió los más representativos que están en la tabla 2, lo cual representa con sus respectivas alteraciones bioquímicas, proteínas que contienen los genes y la participación en las bases de datos Malacards, Harmonizome, OMIM y KEGG. Sin embargo, hay diferentes alteraciones bioquímicas que afectan en la aparición de la enfermedad, además aún están en fases tempranas de investigación por que se requiere una investigación a profundo para poder argumentar la posible aparición de la enfermedad, es por ello que algunas alteraciones aún no se ha comprobado experimentalmente con ensayos clínicos, a pesar de ello se tiene predicciones y simulaciones que se han hecho en el ámbito de cómo influye las dichas alteraciones bioquímicas de como fluye en la aparición de la enfermedad.

Las alteraciones bioquímicas que se identifico es la regulación de la actividad transcripcional. Como principal gen que actúa en esta alteración es el gen TP53 ya que tiene la capacidad de activar o inhibir la transcripción de numerosos genes involucrados en una variedad de procesos celulares, incluyendo la respuesta al estrés celular, la reparación del ADN, el control del ciclo celular y la apoptosis (**Williams & Schumacher, 2016**).

Otra alteración bioquímica, es la activación descontrolada de la vía de señalización RAS/RAF/MEK/ERK, también conocida como vía MAPK. Las MAPK tienen la capacidad biológica de diversos procesos regulares mediante la intervención en múltiples mecanismos celulares. En varios de estos procesos, como la apoptosis, las MAPK realizan un papel dual, ya que su efecto puede variar entre activar o inhibir, dependiendo del tipo de célula y el estímulo específico que esté presente (**Bellio et al., 2021**).

De acuerdo a los datos de la tabla 3 se registró 30 genes de élite. Esto implicó descartar los genes que tienen un rango inferior a 0.5 en la base de datos de Harmonizome, así como un rango menor en la base de datos de Malacards. La razón detrás de esta exclusión se debe a la falta de evidencia experimental que vincule estos genes con el CC en ambas bases de datos. Además, su inclusión podría sesgar la búsqueda de dianas en una dirección incorrecta.

### **3.1.3. Análisis de la integración de datos**

De acuerdo con los resultados obtenidos se definió dos alteraciones bioquímicas presentes las cuales son la regulación de actividad transcripcional y la activación anormal de la vía de señalización de MAPK. En este análisis de integración de datos, por medio del software Cytoscape, donde permitió representar la interacción genética que consta de cuatro niveles de interacción. Los genes que se ingresó en el software están pintados de negro, los genes relacionados agregados por GeneMANIA se representan en gris y los enlaces de colores representan las interacciones que conectan los nodos (genes).

#### **3.1.3.1. Regulación de la actividad transcripcional**

La figura 9, se evidencia la interacción de los genes que están implicados en la regulación transcripcional, por lo cual están en forma de nodos (nodes) y los bordes (edges) que son líneas que están unidas dos nodos y representa una interacción entre los genes. El nodo MLH1 es de alta interacción y el nodo de CCNB1 es el que presenta mayor score, no obstante, los nodos de SMAD2, SMAD4, CDKN1A, TP53 son de muy alta interacción (tabla 4). Cabe mencionar que el software Cytoscape tiene la herramienta de GeneMania, que otorgó de resultado 42 nodos (genes) y 294 bordes que esta presentado en la tabla 5, donde se evidencia los datos generales de la red biológica.

De los 42 genes (anexo H) que se obtuvieron en la interacción se escogió los genes más representativos que tiene mayor interacción. La tabla 4 consta de los 10 genes que presentaron mayor interacción y consta del nombre del gen, los números de nodos, los pares de nodos de múltiples aristas y el puntaje (score) de interacción de los genes. Adicionalmente, se aprecia una gran cantidad de bordes en el gen TP53, es por ello que en la figura 9 se comprueba la presencia de 34 bordes que conectan a la de regulación de la actividad de transcripción. Otro gen de muy alta interacción es el CDKN1A posee 26 bordes, por lo cual afirma que es un gen significativo dentro del CC.

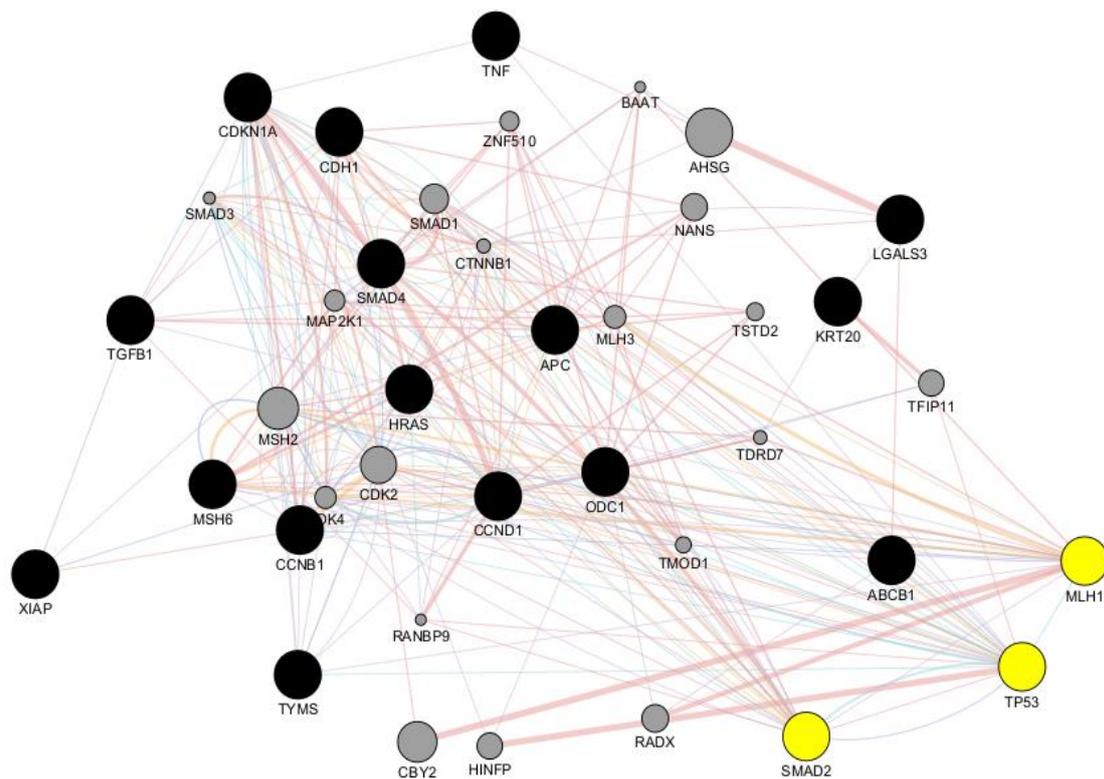
En el análisis de los genes que presentaron mayor interacción se encuentra el gen TP53, que regula la transcripción de diferentes genes involucrados en diversos procesos, lo cual incluye la reparación del ADN. Para el desarrollo del CC, se efectúa mutaciones del gen

TP53 que abarca las mutaciones de cambio de marco causadas por inserciones o mutaciones sin sentido (**Li et al., 2019**).

La principal interacción entre los genes TP53 y MLH1 ocurre en la regulación de la transcripción. Esto implica que el gen MLH1 juega un papel importante en el sistema de reparación del ADN conocido como "reparación de desajustes" (MMR, por sus siglas en inglés). El sistema MMR desencadena una serie de respuestas celulares, como la activación del punto de control del ciclo celular y la inducción de la apoptosis, cuando el daño del ADN alcanza un nivel crítico. La respuesta de las células al daño del ADN se basa en la activación de MLH1, que puede colaborar con el gen supresor de tumores p53 para facilitar la detención del ciclo celular y la muerte celular (**Sitepu et al., 2022**).

Cuando se presenta la inactivación del gen MLH1 se debe a la epimutación que se debe a los cambios epigenéticos adicionales que acompañan a la metilación de la citosina, incluidas las modificaciones represivas de las histonas. Es por ello que se asocia con el CC en aparición temprana que presenta MSI y a la pérdida de la expresión de MLH1 (**Zyla et al., 2021**).

Por otro lado, es crucial subrayar que la interacción señalización génica entre el gen TP53 y el gen SMAD2 se produce a través de la vía TGF- $\beta$ . Esta vía es de gran importancia, ya que cuando las células carecen de p53, se observa una respuesta citostática alterada frente a las señales de TGF- $\beta$ . En consecuencia, los complejos de proteínas Smad y p53 convergen en elementos de unión cis separados dentro de un promotor objetivo, lo que resulta en una activación sinérgica de la transcripción inducida por TGF- $\beta$  (**Cordenonsi et al., 2013**).



**Figura 9.** Red biológica de la regulación de transcripción

**Tabla 4.** Genes representativos de la interacción de la red biológica de la regulación de transcripción.

#	Gene name	Number Of Undirected Edges	Partner Of Multi Edged Node Pairs	Score
1	CDKN1A	26	7	0.5530
2	MSH6	16	2	0.8083
3	CCNB1	21	6	0.7095
4	APC	27	3	0.7048
5	CDH1	17	3	0.6873
6	MLH1	20	6	0.6756
7	SMAD2	26	5	0.6744
8	SMAD4	25	5	0.6165

9	TP53	34	8	0.5577
10	CCND1	24	3	0.5560

**Tabla 5.** *Características generales de la red biológica de la regulación de la transcripción.*

<b>Propiedades</b>	<b>Puntuación</b>
Numero de nodos	42
Numero de bordes	294
Medias Aproximaciones	10.00
Diámetro de red	4
Radio de red	2
Longitud de trazado	1.955
Coefficiente de agrupamiento	0.402
Densidad de red	0.244
Densidad heterogénea	0.574
Centralización de red	0.282
Componentes conectados	1
Tiempo de análisis	0.197

### 3.1.3.2. Activación anormal de la vía señalización MAPK

La figura 10, se puede apreciar la interacción de la actividad descontrolada de la vía de señalización de la quinasa activada por mitógenos (MAPK), en la cual se ha confirmado un marcado incremento en el score del gen PIK3CA. Además, se ha observado una interacción significativa entre los genes BRAF y KRAS, ambos mostrando una interacción muy elevada, según se muestra en la tabla 6. Asimismo, se ha generado una red biológica que consta de 28 nodos y 216 bordes, como se detalla en la tabla 7 y proporciona información general acerca de la estructura de la red.

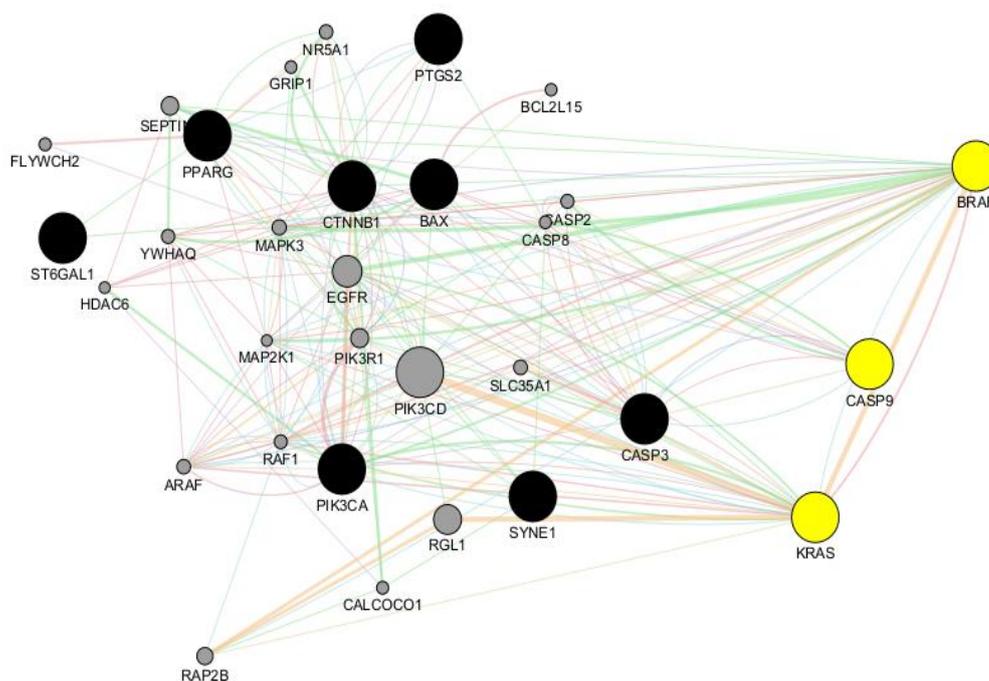
De los 28 genes obtenidos (anexo I) se escogieron 8 genes representativos tal y como se observa en la tabla 6 consta del nombre del gen, los numero de nodos, los pares de nodos aristas y el puntaje (score) de interacción de los genes. Del mismo modo, se evidencia en la figura 10 se comprueba la gran cantidad de bordes que abarca el gen KRAS, así

obteniendo 35 bordes que conecta en la vía de señalización de MAPK. Por otra parte, el gen BRAF igual presentó 26 bordes, esto confirma su participación en la enfermedad. El gen BRAF es responsable de codificar una enzima quinasa que desempeña un papel fundamental en una vía de señalización de la proteína quinasa activada por mitógenos (MAPK). Esta vía implica la participación de GTPasa, que activa a las proteínas RAF, incluyendo ARAF, BRAF y CRAF. Una vez activadas, estas proteínas RAF fosforilan y activan a las proteínas MEK1/2, las cuales, a su vez, fosforilan y activan a las proteínas ERK. Las ERK, al ser fosforiladas, llevan a cabo una amplia gama de acciones, incluyendo la regulación de diversos factores de transcripción y el control de múltiples funciones celulares (**Nakayama et al., 2020**).

Las mutaciones en el gen BRAF son resultado de la activación anormal de la vía MAPK. En el cáncer de colon metastásico con la mutación BRAF V600E, se ha observado la presencia de una mutación específica conocida como p.V600E. Esta mutación implica la sustitución de un residuo de valina por ácido glutámico en la posición 600 de la secuencia de aminoácidos de la proteína BRAF (**Proietti et al., 2020**). Las mutaciones oncogénicas en el gen BRAF promueven la activación espontánea de la proteína BRAF a través de diversos mecanismos. Estos incluyen la mejora de la dimerización asociada al RAS-GTP o la inducción de la dimerización de BRAF en ausencia del RAS-GTP (**Caputo et al., 2019**).

La proteína JNK (Cinasas c-Jun N-terminal) es una quinasa de la familia de las MAPK (Proteínas Quinasas Activadas por Mitógenos) que desempeña un papel importante en la transducción de señales intracelulares. JNK está involucrada en la respuesta celular a una variedad de estímulos, incluyendo el estrés, la inflamación y la apoptosis (**Yue & López, 2020**).

En el momento que las JNK son activadas por estrés, las MAPK p38 desempeñan la regulación de la apoptosis mediante la modulación de la expresión de genes pro-apoptóticos y antiapoptóticos, así como la fosforilación de proteínas clave en la vía apoptótica. Esta modulación puede afectar la actividad de la Caspasa 9 indirectamente a través de la regulación de los niveles de proteínas regulatorias en la vía de la apoptosis (**Parrish et al., 2013**).



**Figura 10.** Red biológica de la sobre activación de la vía de señalización del gen *MAPK*

**Tabla 6.** Genes representativos de la interacción de la red biológica de la hiperactivación de la vía de señalización de *MAPK*

#	Gene name	Number Of Undirected Edges	Partner Of Multi Edged Node Pairs	Score
1	PPARG	7	1	0.7944
2	CASP3	12	3	0.7645
3	CASP9	11	4	0.7368
4	PIK3CA	30	11	0.6653
5	KRAS	35	11	0.5399
6	BRAF	26	8	0.5040
7	PIK3CD	18	6	0.0825
8	EGFR	26	8	0.0533

**Tabla 7.** *Características generales de la red biológica de la hiperactivación de la vía de señalización de MAPK.*

<b>Propiedades</b>	<b>Puntuación</b>
Numero de nodos	28
Numero de bordes	216
Medias Aproximaciones	8.143
Diámetro de red	5
Radio de red	3
Longitud de trazado	2.116
Coefficiente de agrupamiento	0.524
Densidad de red	0.302
Densidad heterogénea	0.627
Centralización de red	0.274
Componentes conectados	1
Tiempo de análisis	0.052

#### **3.1.4. Análisis de la validación de los genes seleccionados**

Se han identificado varios genes relacionados con el CC. Se ha observado que algunos de estos genes están vinculados a fármacos utilizados en el tratamiento de diversas enfermedades. Para verificar la existencia de dianas en forma de fármacos para los genes seleccionados, se realizó una exhaustiva búsqueda en la base de datos de Drugbank.

En la tabla 8 se presentan los genes de élite seleccionados, así como los fármacos que están asociados a dichos genes y su posible aplicabilidad en el tratamiento del CC. Es relevante resaltar que algunos de estos medicamentos también se emplean en otros tipos de tratamientos no relacionados con el CC. Es importante tener en cuenta que existe una carencia de investigación en la comunidad científica en relación a estos medicamentos, y muchos de ellos aún se encuentran en etapas experimentales sin haber sido ampliamente validados. Adicionalmente, se ha encontrado que algunos de estos medicamentos son ligandos propuestos para realizar el acoplamiento molecular lo que representa una oportunidad para contrastar los resultados con el cribado virtual que se describe más adelante.

**Tabla 8.** Fármacos relacionados con los genes élite del CC.

#	Genes	Fármacos vinculados al gen	Diana	Aplica Al CC
1	APC	18	Si	Si
2	MSH6	No	No	No
3	CTNNB1	1	Si	No
4	KRAS	4	Si	Si
5	BAX	1	Si	No
6	TP53	8	Si	Si
7	SMAD4	No	No	No
8	PIK3CA	7	Si	Si
9	ABCB1	597	Si	Si
10	TYMS	11	Si	No
11	CDH1	1	Si	No
12	SMAD2	1	Si	No
13	PTGS2	112	Si	Si
14	ODC1	11	Si	Si
15	BRAF	12	Si	Si
16	KRT20	No	No	No
17	ST6GAL1	No	No	No
18	CDKN1A	2	Si	Si
19	SYNE1	No	No	No
20	LGALS3	2	Si	No
21	TNF	37	Si	Si
22	CCND1	4	Si	Si
23	CASP9	1	Si	No
24	CCNB1	7	Si	Si
25	TGFB1	6	Si	Si
26	MLH1	2	Si	No
27	PPARG	74	Si	Si
28	XIAP	1	Si	No
29	HRAS	6	Si	No
30	CASP3	18	Si	Si

En base a la tabla 8, los genes CCNB1, TP53, PIK3CA, KRAS, BRAF, APC, ABCB1, TGFB1, CASP3, PPARG y PTGS2 constan con sus respectivas dianas establecidas para el tratamiento del CC, no obstante, estos genes pueden generar complicaciones en el desarrollo del cáncer de colon .

En relación a los genes TYMS, ODC1, TNF, CCND1 y CDKN1A, se han identificado dianas específicas para cada uno de ellos. En el caso de ODC1, se han establecido diversas dianas para el tratamiento de diferentes enfermedades, siendo su principal objetivo la fase 3 para el tratamiento de la isquemia miocárdica. Por otro lado, en el gen TNF se ha establecido un fármaco, Bryostatín 1, para el tratamiento del CC, el cual se encuentra en fase 2 de complementación, y, además, este fármaco también está dirigido para otras enfermedades, como la enfermedad de Alzheimer (**Nelson *et al.*, 2017**)

De igual forma, los genes CCND1 y CDKN1A cuentan con dianas establecidas, como el trióxido de arsénico, que se encuentra en fase 2 en reclutamiento. Sin embargo, es importante mencionar que el gen TYMS también posee dianas, pero estas aún se encuentran en fase experimental.

Por otro lado, en el caso de genes como SMAD4, que no tienen fármacos asociados en Drugbank, se ha observado su relevancia en la enfermedad del CC. Esta proteína funciona como un trímero y forma complejos con SMAD2 y SMAD3, los cuales son fosforilados por el receptor. Estos complejos heteroméricos luego se desplazan desde el citoplasma hacia el núcleo, donde la unión con factores de unión al ADN facilita la transcripción de los genes diana (**Zhao *et al.*, 2018**).

El gen MSH6, que desafortunadamente no cuenta con objetivos terapéuticos registrados en la base de datos de Drugbank, desempeña un papel importante en el sistema de reparación de desajustes de ADN (conocido como MMR en inglés). Cuando se presenta una variante patógena en este gen, los nucleótidos no coincidentes persistirán como errores durante la siguiente ronda de replicación. Esto resulta en un fenotipo hipermutador, que incluye la inestabilidad de microsatélites y aumenta el riesgo de desarrollar cáncer de colon (**Pećina-Šlaus *et al.*, 2020**).

Igualmente, en la base de datos de Drugbank no se encuentra información de los genes KRT20, ST6GAL1 y SYNE1 puede ser que aún siguen en fase de experimental. Además, como parte del enfoque terapéutico, se ha considerado analizar las dianas de los genes que no estén directamente relacionadas con el tratamiento del CC. Es por ello que se seleccionó los genes CASP9, MLH1 y SMAD2, ya que sus dianas están dirigidas hacia otras enfermedades.

### **3.1.5. Evaluación de la viabilidad de las dianas terapéuticas**

Se realizó una evaluación con el objetivo de identificar el sitio de drugabilidad más destacado de una proteína utilizando un fármaco que se asemeja a un ligando en su estructura tridimensional. Para llevar a cabo esta evaluación, se analizaron los genes seleccionados utilizando la herramienta DoGSiteScorer del servicio web Protein.Plus. Gracias a esta herramienta computacional, fue posible identificar los sitios de unión al considerar las características generales de la proteína, como su forma, tamaño y propiedades químicas de las aperturas pronosticadas. Sin embargo, se priorizaron los genes que mostraron una mayor influencia en las redes biológicas relacionadas con las alteraciones bioquímicas.

El servicio web de DoGSiteScorer utiliza la predicción de bolsas en la superficie de la proteína como base para analizar la drogabilidad del sitio. Esto se logra mediante el uso de las coordenadas del átomo pesado de la proteína. La proteína se divide en una cuadrícula que utiliza puntos para etiquetar la superposición espacial con cualquier átomo de la proteína. Luego, se aplica un filtro llamado diferencia de Gaussiano (DoG) a esta cuadrícula. Esta operación permite identificar las posiciones en la superficie de la proteína donde la ubicación de un objeto con forma de esfera es favorable (**Volkamer *et al.*, 2012**).

Estas posiciones se agrupan en subbolsillos en función de un umbral de densidad. Los subbolsillos se describen mediante el recuento de átomos de bolsa o grupos funcionales, así como las composiciones de aminoácidos, lo que proporciona información sobre las características fisicoquímicas de la bolsa. Además, se aborda el carácter lipófilo de las bolsas mediante el análisis de la superficie lipófila y la relación global de hidrofobicidad.

Si se proporciona un ligando, también se calcula la superposición entre el ligando y el volumen del bolsillo (Fährrolfes *et al.*, 2017).

La Tabla 9 proporciona los códigos de las proteínas recopiladas junto con sus respectivos puntajes de drogabilidad. Para obtener las estructuras tridimensionales de las proteínas, se utilizó la base de datos PDB, que proporciona códigos para caracterizar estas estructuras. Sin embargo, es importante mencionar que algunas proteínas no han sido profundamente estudiadas, lo que significa que las formas tridimensionales disponibles en el PDB suelen ser fragmentos peptídicos en complejo con una macromolécula diferente a la proteína objetivo. Además, en algunos casos, no se cuenta con registros de las estructuras de las proteínas, lo cual podría afectar los resultados de la presente investigación.

Con respecto a las estructuras ausentes en el PDB, se emplearon las predicciones de estructura del programa artificial AlphaFold. Este programa tiene la capacidad de predecir las estructuras de proteínas y proporciona coordenadas tanto para la cadena principal como para las cadenas laterales, las cuales son generalmente precisas en la estructura esperada cuando hay iones presentes, como en los sitios de unión de zinc y en presencia de cofactores (Scardino *et al.*, 2023).

**Tabla 9.** Puntuación de los sitios de unión de genes élite.

#	Genes	Protein.Plus	Puntaje de drogabilidad
1	APC	3NMZ	0.80
2	MSH6	Alphafold	0.80
3	CTNNB1	1LUJ	0.81
4	KRAS	4OBE	0.80
5	BAX	6EB6	0.68
6	TP53	4MZI	0.80
7	SMAD4	1YGS	0.75
8	PIK3CA	5ITD	0.80
9	ABCB1	Alphafold	0.78
10	TYMS	1HZW	0.80
11	CDH1	Alphafold	0.80

12	SMAD2	5XOD	0.77
13	PTGS2	5F19	0.81
14	ODC1	1D7K	0.80
15	BRAF	5C9C	0.81
16	KRT20	Alphafold	0.79
17	ST6GAL1	4JS1	0.82
18	CDKN1A	Alphafold	0.78
19	SYNE1	6XF2	0.81
20	LGALS3	2NMO	0.68
21	TNF	6RMJ	0.80
22	CCND1	6P88G	0.80
23	CASP9	Alphafold	0.79
24	CCNB1	2B9R	0.83
25	TGFB1	5VQP	0.81
26	MLH1	4P7A	0.79
27	PPARG	8BF1	0.81
28	XIAP	5OQW	0.82
29	HRAS	4XVR	0.80
30	CASP3	1RE1	0.79

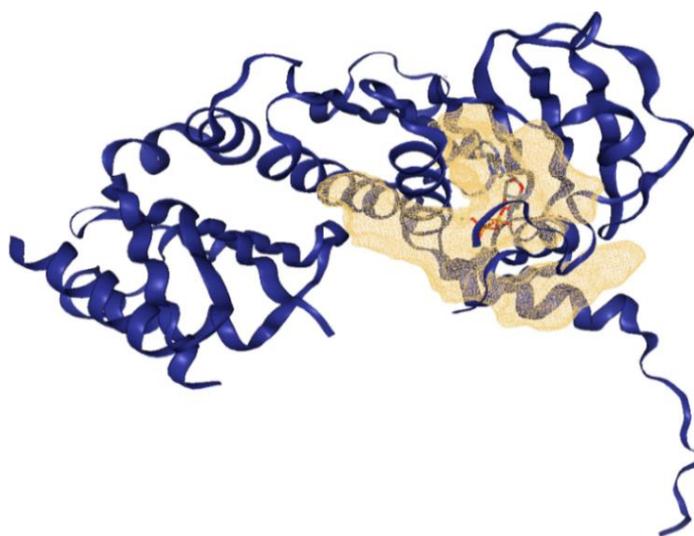
---

DoGSiteScorer es una herramienta computacional basada en la web que evalúa la capacidad de una proteína para ser "drugable" o susceptible de ser objetivo de fármacos. Proporciona una escala de puntuación que va desde 0 (baja drugabilidad) hasta 1 (máxima drugabilidad). Según un estudio **Michel et al. en 2019**, se considera que una proteína tiene una drugabilidad significativa si su puntuación está en el rango de 0.5 a 1, lo que la convierte en un posible objetivo terapéutico. En el caso de las proteínas seleccionadas, como MLH1, CASP9 con una puntuación de 0.79 y SMAD2 con una puntuación de 0.77, su drugabilidad se encuentra dentro del rango mencionado. Esto confirma que estas proteínas son consideradas como objetivos terapéuticos debido a su drugabilidad significativa.

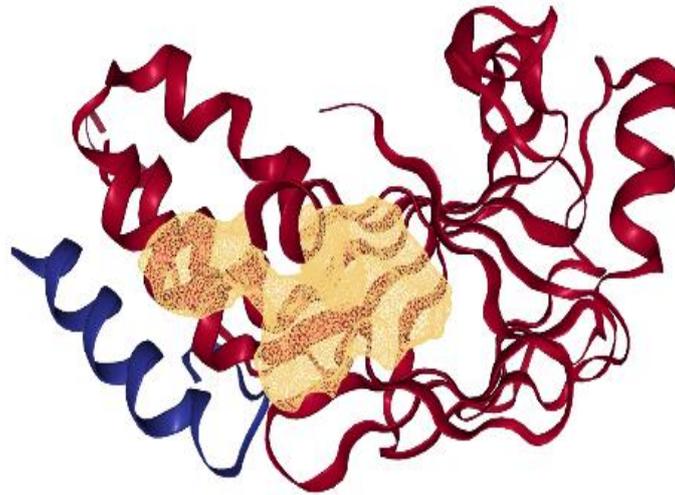
Los pockets (bolsillos) conllevan los siguientes parámetros que se analiza la profundidad, el volumen, el rango de drugabilidad y la composición de aminoácidos. Según el estudio de **Fitri et al., (2021)** un bolsillo con un gran volumen, una gran profundidad y una proporción alta de aminoácidos apolares se considera efectivo para determinar la capacidad de administración de fármacos.

Para el análisis de drugabilidad de las proteínas se escogió el "pocket\_0" debido a que mostraba un mayor potencial para la interacción con fármacos. Es por ello que en la figura 11, muestra la estructura tridimensional de la proteína MutL Homolog 1, con un volumen de cavidad de  $1580.8 \text{ \AA}^3$  y una profundidad de  $1921.99 \text{ \AA}^2$ , además una proporción de aminoácidos apolares de 0.53. Asimismo, en la figura 12 se observa la proteína Mothers Against Decapentaplegic Homolog 2 con un volumen de cavidad de  $515.71 \text{ \AA}^3$  y una profundidad de  $807.51 \text{ \AA}^2$  y una proporción de aminoácidos apolares de 0.44. Para la proteína Caspase 9 se observa en la figura 13, presenta un volumen de cavidad de  $1131.99 \text{ \AA}^3$  y una profundidad de  $1620.98 \text{ \AA}^2$  y aminoácidos apolares de 0.33.

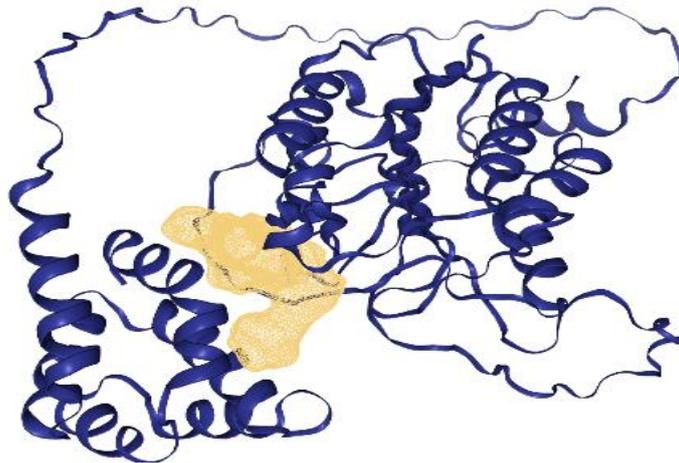
Al comparar los resultados obtenidos de los pockets con los resultados de la investigación de **Chowdhury et al. (2014)**, se encontraron que los bolsillos son similares, lo que sugiere una alta probabilidad de que las proteínas analizadas sean drugables. Además, estos bolsillos farmacológicos tienden a ser más grandes, más profundos, más hidrofílicos y de forma compleja.



**Figura 11.** *Pocket\_0 de la zona drugable de la proteína MutL Homolog 1*



**Figura 12.** *Pocket\_0* de la zona drugable de la proteína Mothers Against Decapentaplegic Homolog 2.



**Figura 13.** *Pocket\_0* de la zona drugable de la proteína Caspase 9

### 3.1.6 Análisis del cribado virtual

La evaluación del cribado virtual, se utilizó MTiOpenScreen, que permite predecir la interacción entre un fármaco y un ligando y proporcionó la afinidad de unión de interacción de las proteínas con los ligandos. Para este análisis, se procedió a cargar las estructuras de las proteínas MLH1, SMAD2 y CASP9 seleccionadas individualmente en formato pdb. Además, se utilizó la biblioteca de fármacos o ligandos comerciales conocida como Drugs-lib. A continuación, se seleccionaron las coordenadas de la Grid-Box obtenidas previamente en DogSiteScorer, ya que proporcionaron el *pocket\_0* que contiene los aminoácidos residuales que fueron transferidos a AutodockTools.

En MTiOpenScreen, se realizaron 1500 ensayos de interacción entre proteínas y ligandos obtenidos de Drugs-lib, específicos para cada proteína seleccionada. Se clasificaron los primeros 100 resultados con la mejor afinidad de interacción y se seleccionaron los resultados con las mejores energías de interacción. De estos 100 resultados, se eligieron los 4 primeros de cada cribado realizado. Sin embargo, es importante tener en cuenta que esta plataforma no proporciona información detallada sobre las vías de interacción entre el ligando y la proteína. Esta falta de información puede limitar la comprensión más profunda de los mecanismos de acción y las implicaciones terapéuticas.

La comprensión de las vías de interacción entre una proteína y un ligando es crucial para el desarrollo de fármacos. Es por esta razón que se generó el gráfico de STITCH, que integra predicciones de interacciones entre proteína-proteína y proteína-química. Al examinar los gráficos 11 y 12, se puede obtener información sobre las interacciones entre proteínas y medicamentos que podrían inhibir la alteración genética. Por lo tanto, se compararon los resultados obtenidos del cribado virtual con los medicamentos de las redes de STITCH y se encontró que ambos incluyen los medicamentos.

Es importante resaltar que en este estudio se seleccionaron los medicamentos de las redes de interacción. Esto se debe a que el papel crucial de la red de interacción es aún más relevante en el campo del desarrollo de fármacos, ya que las enfermedades suelen ser el resultado de múltiples cambios en la misma vía o complejo proteico. Además, es necesario considerar que la cercanía de las proteínas objetivo y la topología de la propia red pueden proporcionar una mejor comprensión del impacto que un fármaco tiene a nivel celular (Szklarczyk *et al.*, 2016).

En el contexto del análisis comparativo de los ligandos obtenidos de tres proteínas diferentes, tanto en el proceso de cribado como en el uso de redes. Se observó, que en la red de transcripción en el gen MLH1 dos interacciones indirectas con los fármacos MgATP y 5-fluororacilo. Sin embargo, al comparar estos resultados con los obtenidos en el cribado, no se encontraron los medicamentos recomendados por la red. Esta discrepancia podría explicarse mediante el método de acoplamiento molecular rígido empleado en el presente estudio, el cual puede afectar la flexibilidad conformacional de

los ligandos. Dicho fenómeno implica que los ligandos podrían experimentar cambios estructurales durante el proceso de unión, como una estrategia para optimizar sus interacciones con las proteínas objetivo (**Guedes et al., 2014**).

Otro aspecto relevante a considerar es la tensión inducida en los ligandos durante el acoplamiento molecular rígido, que se refiere a la imposición de una conformación que difiere significativamente de su estado natural. La presencia de tensión en los ligandos puede ejercer un efecto considerable en la afinidad de unión y, en algunos casos, conducir al rechazo de ciertos ligandos durante el proceso de acoplamiento (**Gu et al., 2021**).

La red STITCH proporciona la información con respecto de cómo actúa el medicamento etopósido indirectamente con las proteínas analizadas por lo cual en la figura 14 representa la red de señalización de MAPK. Esta red se analizó la interacción del gen CASP9 y presenta una interacción indirecta con el medicamento etopósido en la vía intrínseca de apoptosis. A la vez representa una interacción con la CASP 3, por lo cual tiene una relación de inhibición con etopósido y esto permite relacionar con la CaSP9.

Es importante conocer que la vía intrínseca es inducida por la liberación de las proteínas de la membrana externa mitocondrial debido a las proteínas proapoptóticas Bcl-2 (linfoma de células B2) estas median la permeabilización mitocondrial. Hay que señalar que las proteínas tanto proapoptóticas como antiapoptóticas están en control de las cascadas de las cascadas JNK o p38 MAPK a nivel transcripcional o postranscripcional. En niveles postranscripcionales la fosforilación de CASP9 en Thr125 por ERK1/2 (quinasa regulada extracelular) restringe la apoptosis intrínseca en la mitosis. Además, la activación transitoria de JNK retrasa la activación de CASP9 (**Montecucco et al., 2015**).

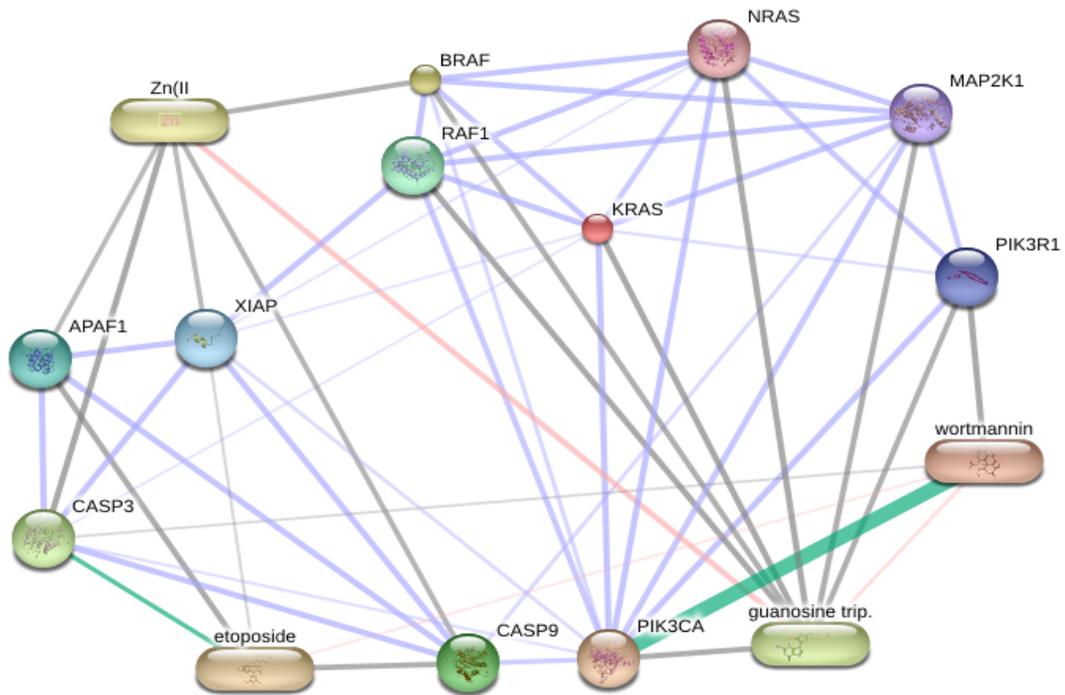
Cuando el gen CASP9 pierde la funcionalidad disminuye su función apoptótica, es por ello que en la investigación de **Cillessen et al., (2017)** menciona el uso de líneas celulares Nalm6 (líneas celulares de leucemia) se trató con el etopósido, lo cual probó la funcionalidad de la vía de apoptosis intrínseca que es mediada por CASP9. Después de la inducción con etopósido, se encontró un fuerte aumento en la actividad de CASP9.

La interacción que se presenta en la figura 15 entre los genes TP53 y SMAD2 está regulada por la vía de señalización del factor de crecimiento transformante-beta (TGF- $\beta$ ). Para cierta interacción menciona **Adorno et al., (2019)** que la isoforma de TP53 (p53AS)

junto con FAST-1, TGF- $\beta$  induce un fuerte aumento en la transcripción de Mix2, estos genes dianas son reguladas por las proteínas SMAD.

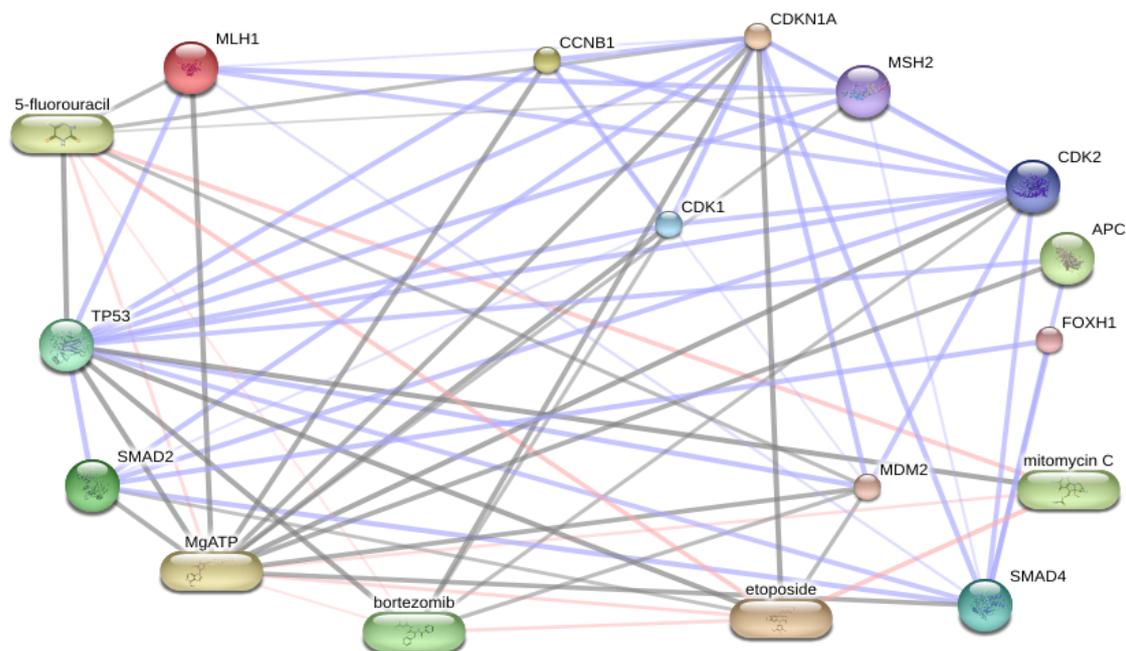
Sin embargo, si se presenta una mutación en el gen TP53, se observa una represión del gen TGF- $\beta$ RII, lo que conlleva a un retraso o disminución en la fosforilación de la proteína Smad2 por parte del receptor TGF-BRI. Tras la mutación de TP53, la fosforilación del gen SMAD2 se ve afectada. Es por ello que la relación que se tendría entre el gen SMAD2 y el medicamento etopósido es que permitiría una mayor lo que resulta en una mayor fosforilación en las cascadas de abajo en la unión del TP53 mutado **(Nemade *et al.*, 2018)**. Además, esta mutación interfiere con la formación del complejo Smad2/3 y Smad4, lo que inhibe la translocación de estas proteínas al núcleo celular. La mutación en TP53 puede llevar a una pérdida de la capacidad de unión al ADN, lo que, a su vez, inhibe la interacción con SMAD en los promotores específicos de genes, disminuyendo así la inducción transcripcional **(Elston & Inman, 2012)**.

Los resultados de los mejores candidatos a dianas terapéuticas debido a su alta estabilidad energética se encuentran en las tablas 10, 11 y 12. Con el objetivo de un enfoque más riguroso, se escogieron los 4 primeros ensayos de los 3 genes analizados, con el propósito de evaluar las posibles propiedades de la energía de unión y, posteriormente, realizar una evaluación de la toxicidad.



**Figura 14.** Red de los genes involucrados la vía de señalización de MAPK con interacción de medicamentos de inhibición.

Nota: La red de MAPK consta de bordes de diferentes colores que representan diferentes acciones: inhibición del medicamento a la proteína (verde), interacciones entre proteínas (morado), interacción medicamento-medamento (rosado) e interacción indirecta (gris). Los nodos son de forma de esferas, por lo que cuando son pequeñas sus estructuras son proteína de estructura 3D desconocida y cuando son grandes constan de la estructura 3D y finalmente los medicamentos se representan en forma de píldoras. La red muestra la interacción de la caspasa 3 con etopósido activa la vía intrínseca de apoptosis. Mientras que la la wortmanina interrumpe los procesos celulares que dependen de la señalización de PI3K.



**Figura 15.** Red de los genes involucrados en la regulación de transcripción con interacción de medicamentos de inhibición.

Nota: La red consta de bordes de diferentes colores que representan diferentes acciones: interacciones entre proteínas (morado), interacción medicamento-medicamento (rosado) e interacción indirecta (gris). Los nodos son de forma de esferas, por lo que cuando son pequeñas sus estructuras son proteína de estructura 3D desconocida y cuando son grandes constan de la estructura 3D y finalmente los medicamentos se representan en forma de píldoras. La red muestra interacciones indirectas, es por ello que el gen TP53 se relacionara con la inhibición indirecta en el etopósido.

**Tabla 10.** *Fármacos con mejor energía de afinidad como posibles dianas terapéuticas relacionadas con la proteína MutL Homolog 1.*

#	Ligando	Zinc CODE	Energía (kcal/mol)	nRot	Código SMILES
1	Zosuquidar	ZINC000100029945	-9.5	7	<chem>O[C@@H](COc1cccc2ncccc12)CN1CCN([C@@H]2c3cccc3[C@@H]3[C@H](c4cccc42)C3(F)F)CC1</chem>
2	Telcagepant	ZINC000028827350	-9.5	5	<chem>O=C(N[C@@H]1CC[C@@H](c2cccc(F)c2F)CN(CC(F)(F)F)C1=O)N1CCC(n2c(=O)[nH]c3ncccc32)CC1</chem>
3	Merestinib	ZINC000095926668	-9.4	6	<chem>Cc1ccc(C(=O)Nc2ccc(Oc3cc4cnn(C)c4cc3c3cn[nH]c3)c(F)c2)c(=O)n1-c1ccc(F)cc1</chem>
4	Lopinavir	ZINC000101007591	-9.3	16	<chem>Cc1cccc(C)c1OCC(=O)N[C@@H](Cc1cccc1)[C@@H](O)C[C@@H](Cc1cccc1)NC(=O)[C@H](C(C)C)N1CCCNC1=O</chem>
5	Enpiroline	ZINC000004628938	-9.2	6	<chem>O[C@H](c1cc(c2ccc(C(F)(F)F)cc2)nc(C(F)(F)F)c1)[C@@H]1CCCCN1</chem>

**Tabla 11.** Fármacos obtenidos con mejor energía de afinidad como posibles dianas terapéuticas relacionadas con la proteína Mothers Against Decapentaplegic Homolog 2

#	Ligando	Zinc CODE	Energía (kcal/mol)	nRot	Código SMILES
1	R428	ZINC000051951669	-9.3	5	<chem>Nc1nc(Nc2ccc3c(c2)CC[C@@H](N2CCCC2)CC3)nn1-c1cc2c(nn1)-c1cccc1CCC2</chem>
2	Bagrosin	ZINC000026663185	-9.3	1	<chem>C[C@]1(c2ccc3ccc4ccccc4c3c2)NC(=O)NC1=O</chem>
3	Mk3207	ZINC000103760981	-9.0	4	<chem>O=C(CN1C(=O)C2(CCCC2)NC[C@@H]1c1cc(F)cc(F)c1)Nc1ccc2c(c1)C[C@@]1(C2)C(=O)Nc2ncccc21</chem>
4	Liafensine	ZINC000095598450	-9.0	3	<chem>CN1Cc2cc(-c3ccc(N)nn3)ccc2[C@H](c2ccc3ccccc3c2)C1</chem>
5	Etoposide	ZINC000011615744	-8.8	8	<chem>COc1cc([C@@H]2c3cc4c(cc3[C@H](O[C@@H]3O[C@@H]5CO[C@@H](C)O[C@@H]5[C@@H](O)[C@@H]3O)[C@@H]3COC(=O)[C@@H]32)OCO4)cc(OC)c1O</chem>

**Tabla 12.** Fármacos con mejor energía de afinidad como posibles dianas terapéuticas relacionadas con la proteína Caspasa 9.

#	Ligando	Zinc CODE	Energía (kcal/mol)	nRot	Código SMILES
1	Venetoclax	ZINC000150338755	-10.8	13	<chem>CC1(C)CCC(CN2CCN(c3ccc(C(=O)NS(=O)(=O)c4ccc(NCC5CCOC5)c([N+](=O)[O])c4)c(Oc4cnc5[nH]ccc5c4)c3)CC2)=C(c2ccc(Cl)cc2)C1</chem>
2	Ravidasvir	ZINC000150607150	-10.5	12	<chem>COC(=O)N[C@H](C(=O)N1CCC[C@H]1c1ncc(c2ccc3cc(c4ccc5nc([C@@H]6CCCN6C(=O)[C@@H](NC(=O)OC)C(C)C)[nH]c5c4)ccc3c2)[nH]1)C(C)C</chem>
3	Elbasvir	ZINC000150588351	-10.4	13	<chem>COC(=O)N[C@H](C(=O)N1CCC[C@H]1c1nc(c2ccc3c(c2)O[C@@H](c2ccccc2)n2c3cc3cc(c4c[nH]c([C@@H]5CCCN5C(=O)[C@@H](NC(=O)OC)C(C)C)n4)ccc32)c[nH]1)C(C)C</chem>
4	Samatasvir	ZINC000150588806	-10.3	13	<chem>COC(=O)N[C@H](C(=O)N1CCC[C@H]1c1nc2cc(c3csc4c(c5ccc(c6c[nH]c([C@@H]7CCCN7C(=O)[C@H](NC(=O)OC)c7ccccc7)n6)cc5)csc34)ccc2[nH]1)C(C)C</chem>
5	Etoposide	ZINC000015449127	-8.8	8	<chem>COc1cc([C@@H]2c3cc4c(cc3[C@H](O[C@@H]3O[C@@H]5CO[C@@H](C)O[C@H]5[C@H](O)[C@H]3O)[C@@H]3COC(=O)[C@@H]32)OCO4)cc(OC)c1O</chem>

### 3.1.7. Análisis de interacción molecular

Se realizó la búsqueda bibliográfica de las dianas terapéuticas que ya están propuestas en los genes para tomarlas como controles para la comparación con los ligandos obtenidos en relación con la afinidad de unión. Es por ello que en el caso del gen SMAD2 obtiene una afinidad de unión de  $-7.88$  kcal/mol con el ligando imiquimod, este fármaco se usa para el tratamiento de verrugas anogenitales externas, queratosis actínicas y carcinomas basocelulares superficiales (**Jayanthi, et al., 2023**).

En el caso del gen CASP9, se tomó como control el ligando betulin que presentó una afinidad de unión de  $-8.3$  kcal/mol, este compuesto fitoquímico se está usando para el tratamiento carcinoma hepatocelular (**Mustafa et al., 2023**).

Estos resultados son valores estándar, por lo cual se comparó con el medicamento seleccionado en las redes de STITCH. Del medicamento que se seleccionó se realizó análisis de acoplamiento molecular para las dos proteínas, para su posterior análisis de toxicidad de las moléculas. Además, se realizó para todos los ligandos obtenidos tanto en el cribado como en las redes para obtener los aminoácidos que interactúan en las proteínas seleccionadas

La interacción molecular se obtuvo los aminoácidos residuales de las interacciones ligando-proteína que se utilizó el software Discovery Studio la cual permitió la visualización de patrones 2D y 3D. Por otra parte, los cinco compuestos seleccionados del cribado virtual se examinaron por separado para cada proteína receptora. Es por ello que se tomó los principales fármacos que presentaron menor valor a sus afinidades de unión tal y como se presenta en la tabla 13.

**Tabla 13.** *Afinidad de unión con los principales genes seleccionados.*

#	Ligando	Receptor	Afinidad de unión (Kcal/mol)	Aminoácidos que interactúan
Control	Betulin		-8.30	Arg:408 y Lys:409
1	Venetoclax		-11.98	HisA:243, Ser 259, AsnA:265, ArgA:10, ArgA:11, LeuA:14, AlaA:209, LysA:210, ValA:213, ValA:258, PheA:319, ProA:257, LysA: 262.
2	Samatasvir	CASP9	-11.04	ArgA:10, ArgA:11, ArgA:13, LeuA:14, CysA:256, ArgA:324, LeuA:323, ThrA:325
3	Ravidasvir		-9.94	ArgA:10, ArgA:11, LeuA:14, ArgA:15, GluA:18, ProA:257, His:243, ArgA:324.
4	Elbasvir		-9.12	ArgA:11, CysA:12, LeuA:14, ArgA:15, CysA:256, AlaA:209, LysA:262
5	Etoposide		-9.10	ArgA:7, ArgA:10, ArgA:11, LeuA:14, LeuA:177
control	Imiquimod		-7.88	His:12, Glu: 6 y Leu:7
1	R428		-10.34	ThrA:372, ValA:373, CysA:374, LysA:375, GluA:425
2	Etoposide	SMAD2	-8.85	CysA:362, ArgA:365, LysA:372, CysA:374, IleA:376, LysA:383
3	Liafensine		-8.50	AlaA:371, ThrA:372, ValA:373, CysA:374, LysA:375, TrpA: 422, GluA: 425
4	Mk3207		-8,39	GluA:425, ThrA:372, LysA:375, ProA:377, LeuA:382
5	Bagrosin		-7.70	ValA: 373, LysA:375, AlaA:424, GluA:425

En el análisis de las interacciones proteína-ligando, se comparó la afinidad de unión entre la interacción del medicamento de control y el medicamento seleccionado de la red STITCH. Como resultado en la interacción de las dos proteínas, el medicamento seleccionado de la red mostró una mayor afinidad de unión. Esta mayor afinidad que presentaron las interacciones se puede atribuirse a las diversas fuerzas interactivas, que incluyen interacciones hidrofóbicas, fuerzas de Van der Waals, enlaces de hidrógeno e interacciones electrostáticas.

Como factores principales que otorgaron mayor afinidad fueron las interacciones intermoleculares débiles, como los enlaces de hidrógeno y las interacciones hidrofóbicas, son actores clave en la estabilización de ligandos energéticamente favorecidos, en un entorno conformacional abierto de estructuras proteicas. Los enlaces de hidrógeno promueven la afinidad de unión del ligando al desplazar las moléculas de agua unidas a la proteína (**Chen et al., 2016**). Mientras que cuando exista mayores interacciones hidrofóbicas en el núcleo activo de la interfase del fármaco-objetivo aumenta aún más la actividad biológica del fármaco (**Varma et al., 2020**).

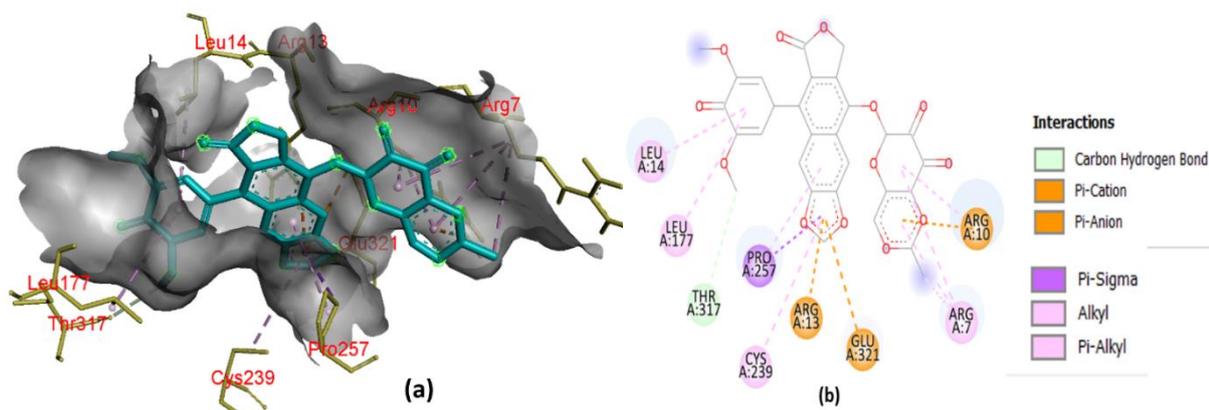
Por otro lado, en la interacción del ligando etopósido con la proteína CASP9 que presentó una afinidad de -9.10 kcal/mol. En la figura 16, se observa que presentó un enlace carbono-hidrogeno convencional en el aminoácido ThrA:317, esta es una interacción de carácter sencillo, sin embargo, su baja reactividad frente a otros compuestos lo hacen un enlace sólido y difícil de romper y se forma debido a que el átomo de carbono completa los 8 electrones que necesita para estabilidad al enlazarse con los hidrógenos circundantes (**Peterson, 2019**).

De igual manera, se encontró un enlace Pi-anión electrostático con GluA:321. También en el dioxano presento un enlace de Pi-cación formándose con el aminoácido ArgA:13 y a la vez el aminoácido ArgA:10 presenta dos enlaces formadores de Pi-cación y Pi-donador de hidrogeno, el enlace pi cación se caracteriza por ser de naturaleza no covalente y se forma en anillos que presentan aromaticidad y un grupo funcional con carga positiva de la estructura del ligando estudiado, pese a que en el presente estudio se ha detectado múltiples enlaces de este tipo, según **Alabugin et al., (2019)** es inusual debido a que los anillos de benceno que participan, no presentan un momento dipolar definido porque la tendencia polar de los carbonos componentes se anula debido a la simetría de la molécula,

sin embargo, electrones presentes encima y debajo de la estructura pueden simular una carga negativa parcial lo que genera su interacción con los iones positivos del ligando, mientras que el enlace pi anión se genera por condiciones similares pero con un grupo funcional con carga negativa reaccionando con una nube de electrones con característica positiva.

Otras interacciones hidrofóbicas observadas como es alkyl con LeuA:14, LeuA:177, CysA:239 y ArgA:7 con triple enlace de alkyl e interacciones Pi-alkyl con Pi-sigma con residuo de aminoácido de ProA:257. Las interacciones hidrofóbicas se dan típicamente en el interior de la proteína que es donde se ubica su región hidrofóbica y se da debido a que los grupos funcionales hidrofóbicos del ligando y las regiones de dicha naturaleza en la proteína se juntan debido a un desplazamiento del solvente o agua en el medio donde existe afinidad polar, mediante una reacción de exclusión lo que genera el aumento de la entropía en el medio y por consiguiente un mejor acomodamiento del ligando al interior del sitio activo diana (Xiao et al., 2020).

Se identificó que las interacciones Pi-cation son interacciones no covalentes de corto alcance entre cationes y sistemas  $\pi$  cercanos las cuales presentaron ArgA:10, ArgA:13, HisA:243 (Infield et al., 2021).



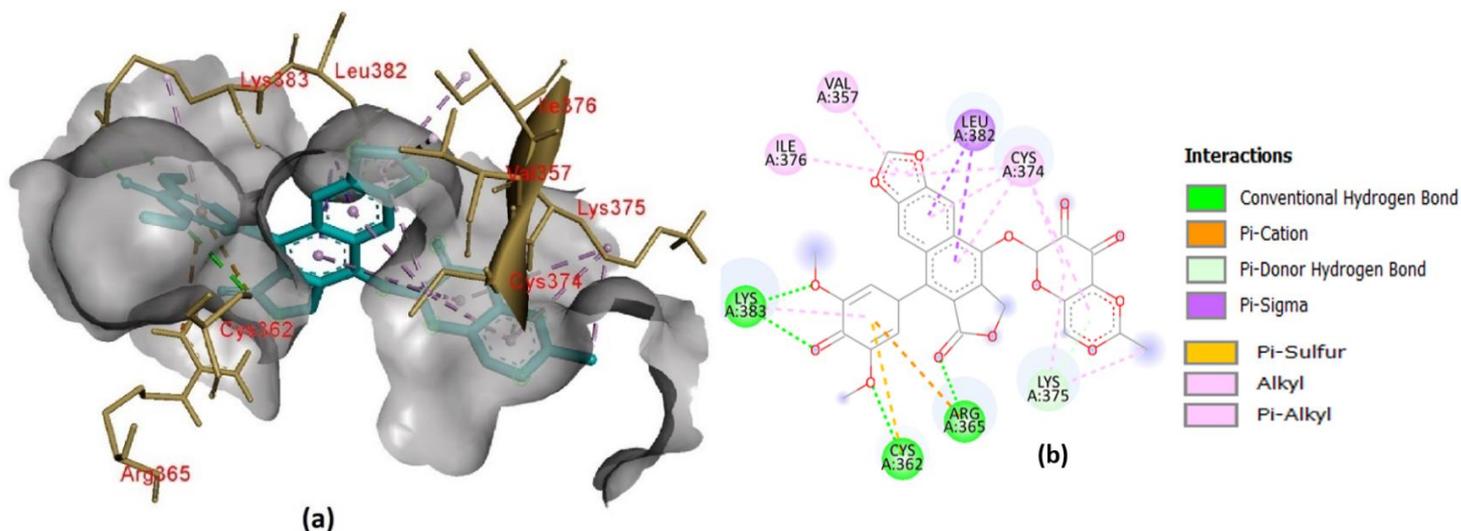
**Figura 16.** Pose acoplada de la estructura de la proteína Caspasa 9 (AlphaFold)

Nota: (a) Etopósido 3D con aminoácidos circundantes de predicción de la proteína AlphaFold (b) Vista en 2D de la interacción de etoposide con los aminoácidos circundantes de proteína AlphaFold.

La interacción de la proteína SMAD2 con el ligando etopósido presentó una afinidad de unión de -8.85 kcal/mol. En la figura 17, se observa la presencia de tres enlaces convencionales de hidrógeno (ArgA:365, CysA:362, LysA:383) debido a que estos aminoácidos de característica polar reaccionaron con los oxígenos exteriores del ligando, con una interacción del tipo electrostática dada por el hidrógeno que tiene una alta electronegatividad con otro de las mismas características, generando una fuerza de atracción del tipo dipolo, este enlace es en términos de energía menor a otros más fuertes como el enlace covalente, sin embargo puede ser considerado como una fuerza de interacción notable cuando se encuentran en alta cantidad en la zona, tal como se observa en (Figura 17) donde múltiples enlaces de hidrógeno permiten la eficiente unión del ligando con la proteína SMAD2.

Por otra parte, se ha detectado un enlace de Pi-donador de hidrogeno de LysA: 375, la característica de estos enlaces se centra en su estructura orbital, donde presentan dos orbitales p completos de los cuales uno puede funcionar como un donador sigma y como donador pi, en este caso la Lisina al tener grupo funcional de característica hidrofílica polar, es capaz de formar un enlace del tipo puente de hidrógeno con el anillo de benceno del ligando que participa como donador pi.

Además, se evidencia la interacción de los enlaces Pi-sulfur que presenta la CysA:362 con el enlace Pi-cación del aminoácido ArgA:365, esto da lugar a una estable con una fuerte interacción entre el grupo sulfóxido (que se forma gracias a la cisteína contiene un grupo tiol, que puede oxidarse para formar un grupo sulfóxido) y la arginina (**Ringer et al., 2017**). Asimismo, la estabilidad de unión que presenta la interacción, se atribuye a las interacciones Pi-Pi, son interacciones debidas a la superposición de las nubes electrónicas pi, lo que da lugar a una interacción de tipo apilamiento, determinado por el solapamiento de dos lóbulos de un orbital electrónico en los dos lóbulos del orbital electrónico contiguo, es por ello que hay interacción entre la Pi-sigma CysA:374 con los Pi-alquilo que es IleA:376, ValA:357. El gran número de interacciones Pi-sigma (Pi-alquilo y Pi-sulfur) que implican en gran medida la transferencia de carga ayuda a intercalar el fármaco en el sitio de unión del receptor (**Abdullahi et al., 2022**).



**Figura 17.** Pose acoplada de la estructura de la proteína SMAD2 (PDB:5XOD).

Nota: (a) Etopósido 3D con aminoácidos circundantes de 5XOD; (b) Vista en 2D de la interacción de etopósido con los aminoácidos circundantes de 5XOD.

### 3.1.8. Análisis de la toxicidad de la molécula

En la evaluación de la toxicidad del fármaco etopósido, se utilizó el software VEGA, específicamente el modelo QSAR *in silico*. Este software desempeña un papel analítico al proporcionar análisis de toxicidad mediante predicciones computacionales. Para interpretar los análisis se presenta en la tabla 14. Además, los parámetros que se evalúa al compuesto es la mutagenicidad, toxicidad del desarrollo, carcinogenicidad. A continuación, se demuestra la tabla 15 parámetros de toxicidad para el fármaco seleccionado de las redes de STITTCH. Es importante tener en cuenta que los resultados de análisis de la toxicidad que se obtuvo del etopósido que es aplicado en los genes CASP9 como SMAD2 se obtuvo los mismos resultados.

Los resultados que se obtuvieron en la tabla 15 es el análisis de toxicidad del del fármaco se observa que en la hepatotoxicidad y mutagenicidad presenta una interpretación segura. No obstante, en la carcinogénesis implica la transición de células normales a células cancerosas a través de una secuencia de etapas que implican tanto alteraciones genéticas (es decir, mutaciones) como eventos no genéticos. En los resultados se obtuvo poco confiable, sin embargo, el compuesto se encuentra en el rango de 1 esto puede ser que es probable que exista un mecanismo genotóxico para la carcinogenicidad.

Sin embargo, estos descriptores empleados en los modelos QSAR se derivan de cálculos de estimación y predicción, lo que indica que es necesario realizar estudios posteriores de ensayos *in vivo* con el objetivo de obtener resultados experimentales sobre la seguridad del fármaco obtenido de la red de STITCH.

**Tabla 14.** Interpretación para los datos farmacológicos.

Elemento	Interpretación
	Seguro
	Intermedio
	Muy peligroso
1	Poco confiable
2	Moderadamente confiable
3	Muy Confiable

**Tabla 15.** Prueba de toxicidad de los fármacos seleccionado

Parámetros	Fármaco	
	Etopósido	
	Predicción	Fiabilidad
Mutagenicidad		1
Carcinogénesis		1
Hepatotoxicidad		2

Para un mayor enfoque de análisis de toxicidad se realizó las pruebas ADMET que realiza la web pKCSM que es análisis *in silico* donde predice las propiedades farmacocinéticas de los parámetros de absorción, distribución, excreción y toxicidad. Este análisis es esencial en el proceso de descubrimiento de medicamentos, ya que permite predecir el comportamiento del fármaco en el organismo, como su absorción después de la administración oral o en el tracto gastrointestinal. Si la absorción es deficiente, podría afectar su distribución y metabolismo, potencialmente llevando a problemas como neurotoxicidad o nefrotoxicidad (Mitra *et al.*, 2022).

En última instancia, el propósito del estudio ADMET es entender cómo la molécula de fármaco se comporta dentro del cuerpo. Por consiguiente, es una parte fundamental en el diseño computacional de fármacos.

La prueba de absorción se muestra en la tabla 16, el etopósido presenta baja solubilidad a 25°C en agua, es por ello que para obtener una mejor solubilidad menciona **Hempel, (2020)** que se debe formular con polietilenglicol, Tween 80, etanol y alcohol bencílico. No obstante, para obtener una mejor biodisponibilidad y mejor acción del fármaco el etopósido se en su forma de fosfato, se logra una mayor solubilidad en agua. Al ser más soluble en agua, el etopósido fosfato puede administrarse de manera más eficiente y rápida.

La permeabilidad de la piel es un parámetro importante la cual se define como la propiedad que tiene el fármaco para atravesar la membrana biológica, por lo cual obtuvo un resultado desfavorable logkp (distribución de la permeabilidad logarítmica) ya que presentó un resultado mayor a -2.5 que presenta lo cual no es interés para el desarrollo de la administración transdérmica de fármaco (**Firdausy et al., 2020**).

La absorción intestinal presentó 87.389% al comparar con lo propuesto de **Pratama et al., (2020)** menciona que una molécula con una absorción intestinal inferior al 30% es considerada pobre de absorción por lo cual el compuesto está dentro del rango. Se considera que un compuesto que tiene una alta permeabilidad al Caco-2 por encima del rango de  $8 \times 10^{-6}$  cm/s es aceptable para el desarrollo de forma de dosificación oral del compuesto (**Peres et al., 2021**).

La glicoproteína P (P-gp) es un transportador de membrana bien identificado con capacidad para expulsar moléculas de fármaco fuera de la célula, lo que reduce la eficacia de la terapia con fármaco (**Robinson & Tiriveedhi, 2020**). Como resultado predice que si es actúa como un sustrato de la glicoproteína P. Además, los sustratos de la glicoproteína P pueden actuar potencialmente como inhibidores o inductores de su función. Al realizar la prueba de inhibición de las glicoproteínas I y II no actúan como inhibidor. Puede actuar como inductores de la glicoproteína P que reduce la biodisponibilidad (**Cox, 2021**).

**Tabla 16. Propiedades de absorción del Etopósido**

Parámetros	Fármaco	
	Etopósido	
	Valor previsto	
<b>Absorción</b>	Solubilidad del agua (mol/L)	-3.524
	Permeabilidad Caco-2 (Log Papp in 10 <sup>-6</sup> cm/s)	0.375
	Absorción intestinal (% Absorbido)	87.389
	Permeabilidad de la piel (Log Kp)	-2.735
	Inhibidor de la glicoproteína P I	No
	Inhibidor de la glicoproteína P II	No
	Sustrato de glicoproteína P	Si

Por otro lado, la prueba de distribución como se observa en la tabla 17, se llevó a cabo el análisis de la barrera hematoencefálica, lo cual es importante saber la capacidad que tiene el fármaco para cruzar al cerebro, esta característica es importante porque permite reducir los efectos secundarios y toxicidad. Se considera que un logBBB (para la barrera hematoencefálica) mayor de 0,3 para un fármaco determinado cruza fácilmente la BBB, mientras que las moléculas con logBBB menor a -1 se distribuyen mal en el cerebro, al comparar se observa que no se distribuiría de forma deficiente (Pires *et al.*, 2015).

En cambio, en el parámetro volumen de distribución en estado estacionario, representa el volumen teórico que requiere un fármaco para que se distribuya en la sangre, más alto sea el VD se distribuye en el tejido que en el plasma, al observar el rango se encuentra en un estado normal (Mansoor & Mahabadi, 2022.). El parámetro Fracción no ligada predice la fracción que se liberará en el plasma. Este valor interviene en la filtración glomerular renal y en el metabolismo hepático, es por ello que afecta al volumen de distribución y al aclaramiento total que es una prueba de excreción. Es necesario obtener un valor bajo ya

que cuando se obtiene valores altos menos eficiente puede atravesar las membranas celulares, como nuestro resultado dio bajo valor se predice que podrá cruzar la membrana (Watanabe *et al.*, 2018).

El parámetro permeabilidad del SNC (sistema nervioso central), se basa en el producto del área superficie de permeabilidad hematoencefálica (logPS) es una medida directa para predecir si podrá entrar al sistema nervioso, su rango optimo es logPS menor a -3, al comparar con el resultado obtenido, se encuentra en el rango de que no podrá entrar al SNC (Jain *et al.*, 2017).

**Tabla 17.** *Propiedades de distribución del etopósido.*

		<b>Fármaco</b> Etopósido
	<b>Parámetros</b>	<b>Valor Previsto</b>
<b>Distribución</b>	Permeabilidad BBB (log BB)	-1.578
	Fracción no ligada (Fu)	0.112
	Permeabilidad del SNC (log PS)	-3.815
	Volumen de distribución en estado estacionario (log L/kg)	0.594

La prueba de metabolismo representa la tabla 18, se debe tener en cuenta que el citocromo P450 es una enzima que cumple la función de desintoxicación en el cuerpo. Si es probable que el compuesto sea un sustrato del citocromo P450 ya que la isoforma 3A4 presenta la predicción. Por otra parte, en los inhibidores del citocromo P450 se puede observar que en todos los parámetros no son inhibidor de la función de desintoxicación (Pires *et al.*, 2015).

**Tabla 18. Propiedades del metabolismo del etopósido**

		Fármaco
		Etopósido
	Parámetros	Valor predicho
<b>Metabolismo</b>	Inhibidor CYP2C19	No
	Sustrato CYP2D6	No
	Inhibidor CYP2C9	No
	Inhibidor CYP2D6	No
	Inhibidor CYP3A4	No
	Sustrato CYP3A4	Si

La prueba de excreción presenta dos parámetros tal como lo muestra la tabla 19, el primero es el sustrato OCT2 renal que representa como un transportador de captación renal que cumple la función de eliminación renal de fármacos, por lo que en este caso se predice que no se comporta como sustrato de OCT2 (**Vian *et al.*, 2019**). El otro parámetro es el aclaramiento del fármaco es una medida de la capacidad de todo el cuerpo para prescindir del fármaco. Esta se produce en composición del aclaramiento hepático y renal que está relacionado con la biodisponibilidad, al observar el resultado obtenido se puede decir que se tiene una velocidad rápida de eliminación del fármaco (**Watanabe *et al.*, 2019**).

**Tabla 19. Propiedades de excreción del etopósido**

		Fármaco
		Etopósido
	Parámetros	Valor predicho
<b>Excreción</b>	Sustrato OCT2 renal	No
	Aclaramiento total	0.006

En cuanto a la prueba de toxicidad está representada por la tabla 20 se tiene la prueba en la toxicidad por AMES que se considera como una metodología ampliamente utilizada que se considera para verificar el potencial mutagénico de un determinado fármaco utilizando bacterias, lo que indica que cuando los resultados son positivos, el compuesto estudiado será mutagénico y podría comportarse como carcinógeno (**Benigni, 2021**). Comparando con lo obtenido las predicciones son negativas.

En la máxima dosis tolerada proporciona el valor umbral para los efectos secundarios relacionados con la dosis. Este parámetro se basa en los efectos terapéuticos que se desean obtener por el fármaco y los efectos secundarios adversos o tóxicos que no se desean. A la vez está en función en la promiscuidad como la pseudopotencia. Según **Sur et al., (2021)**, indica que si se encuentra inferior o igual a  $0,477 \log (\text{mg/kg/día})$  se considera bajo en dosis, es por ello que el resultado presenta un valor bajo de dosis. Esto puede ser por la característica que tiene dicho fármaco que presenta una alta promiscuidad y esto se debe a que puede unirse a diferentes proteínas humanas y muy alta en pseudopotencia. Sin embargo, estas características de estar en equilibrio para obtener interacciones óptimas con las proteínas que se relaciona. Hay que tener en cuenta que la dosis no se puede aumentar por que conduciría a efectos no deseados (**T. Liu et al., 2016**).

El parámetro de inhibidores Inhibidor hERG (gen humano relacionado con el éter-a-go-go) I y II obtuvo un resultado negativo, es importante mencionar que este parámetro analiza las inhibiciones canales del potasio codificadas por hERG. Es importante saber que los efectos secundarios al inhibir hERG, produciría las arritmias cardíacas por lo cual llevaría a efectos no deseados por ende al fracaso farmacéutico (**Garrido et al., 2020**).

La toxicidad aguda oral (LD50) en ratas representa la simulación de la dosis letal que matara al 50% de un grupo de animales de prueba en un período determinado. Se obtuvo un valor considerable  $3.037 \text{ mol/kg}$ , lo cual al comparar el modelo establecido por **Gadaleta et al., (2019)**, sugiere los productos químicos seguros ( $\text{LD50} > 5000 \text{ mg/kg}$ ) están incluidos en la Categoría IV, lo cual momento de cambiar las unidades otorga un valor de  $178 \times 10^4 \text{ mg/kg}$  óptimo de producto químico. Mientras que la toxicidad crónica oral en ratas mide la dosis más baja del compuesto que da como resultado un efecto adverso (LOAEL) y está relacionada con la duración del tratamiento y la concentración bioactiva lo cual otorga un valor similar al obtenido del LD50.

Por otra parte, en la sensibilización de la piel otorgo un valor negativo. En cuanto a la toxicidad de los pececillos representan la concentración de una molécula necesaria para causar la muerte 50% de los Flathead Minnows (especie de peces). El resultado proporciono un valor mayor al 0,5 mM lo cual no produce una toxicidad aguda. En cambio, en la toxicidad de por *Tetrahymena pyriformis* es una bacteria protozoaria cuya toxicidad se utiliza a menudo como prueba de toxicidad, lo cual otorgó un valor de no toxicidad ya que se encuentra menor que lo sugerido que es de 0.5 (ug/L) (Han *et al.*, 2019)

La lesión hepática inducida por fármacos es una preocupación de seguridad importante para el desarrollo de fármacos. La hepatotoxicidad está asociada con la alteración de la función normal del hígado y los valores previstos para este caso es negativo.

**Tabla 20.** *Propiedades de toxicidad del etopósido.*

Parámetros	Fármaco
	Etopósido
	Valor predicho
<b>Toxicidad</b>	
AMES toxicidad	No
Máxima dosis tolerada (log mg/kg/día)	-0.331
Inhibidor hERG I	No
Inhibidor hERG II	No
Toxicidad aguda oral en ratas (LD50) (mol/kg)	3.037
Toxicidad crónica oral en ratas (LOAEL) (log mg/kg_bw/day)	3.268
Sensibilización de la piel	No
Toxicidad de los pececillos (log mM)	4.399

---

Toxicidad por <i>T.Pyriformis</i>	0.285
(log ug/L)	
Hepatotoxicidad	No

---

Es importante destacar que el resultado final del análisis del etopósido otorgo negativo en las pruebas de toxicidad y hepatotoxicidad es por ello que para contrastar con estudios posteriores y ensayos se podría realizar pruebas tanto *in vitro* como *in vivo*, esto con la finalidad de confirmar experimentalmente que la administración del etopósido sea con total seguridad y así obtener datos clínicos que permitan mejorar tanto el principio activo como la farmacodinámica.

## CAPITULO IV

### CONCLUSIONES Y RECOMENDACIONES

#### 4.1. Conclusiones

Se analizó las diferentes bases genéticas del cáncer de colon con la finalidad de encontrar posibles dianas terapéuticas que están involucradas en la alteración de la transcripción, sistema de mecanismo de reparación, vía de señalización TGF- $\beta$  y alteración apoptótica en la vía de señalización de MAPK. Los genes identificados presentaron diversas funciones debido a que se involucra diferentes rutas metabólicas y vías moleculares de señalización. Se tomó en cuenta los genes que tuvieron mayor participación en disfunción los cuales son SMAD2 y CASP9.

Se estudio la expresión diferencial de los genes, la inhibición y expresión de la vía MAPK, cuando el gen KRAS pierde su funcionalidad por mutación o inhibición, lo que desencadena la expresión de MAPK, MAPK14, JNK y facilita el desarrollo de la enfermedad del cáncer de colon. El gen PI3K actúa como inhibidor de la función del gen KRAS, el gen YAP1 actúa como cofactor transcripcional, para reactivar la ruta del gen MAPK.

Se analizaron 30 genes élite en MalaCards, que presentaban alteraciones bioquímicas relacionadas con la enfermedad del cáncer de colon (CC). Estos genes fueron validados utilizando bases de datos como OMIM, Harmonizome y KEGG. De los 30 genes, se seleccionaron aquellos que no estaban vinculados como objetivos terapéuticos para la enfermedad, lo que llevó a la elección de los genes SMAD2, MLH1 y CASP9 para cada alteración bioquímica identificada. En el proceso de selección, se descartaron genes que ya habían sido objeto de estudios con aplicaciones terapéuticas para el CC, como KRAS, BRAF, TP53 y APC. Asimismo, se excluyeron aquellos genes que tenían evidencia insuficiente de relación con la enfermedad.

La viabilidad de las dianas terapéuticas se efectuó analizando la drugabilidad, mediante la plataforma Protein Plus con el plugin integrado DoGSiteScorer. Se utilizó las proteínas en formato .pdb provenientes de Protein Data Bank, este proceso se enfocó en los genes CASP9, y SMAD2 los que presentaron excelentes condiciones para interacción ligando-

molécula. Por otra parte, se identificó en las redes PPI de STITCH el medicamento etopósido que está relacionado con los genes mencionados.

Las propiedades toxicológicas evaluadas al medicamento etopósido por medio de predicciones de plataformas bioinformáticas como es QSAR VEGA y pkCSM permitió conocer las propiedades farmacocinéticas, los parámetros de toxicidad y hepatotoxicidad que se analizaron se obtuvo un resultado negativo, estos parámetros son esenciales para el desarrollo y utilización del fármaco.

#### **4.2. Recomendaciones**

La efectividad de los componentes activos presentes en los medicamentos actuales para cualquier tipo de tratamiento requiere un amplio estudio y pruebas experimentales, que a menudo implican un largo período de seguimiento. Los modelos matemáticos, predicciones mediante química computacional, análisis complejo realizado mediante computadora sigue siendo una oportunidad valiosa en este campo.

El cribado virtual se sugiere realizarlo con diferentes herramientas, se podría utilizar PyRx, BindingDB, combinando los resultados con la web STITCH que proporciona la red PP. Con esta información se puede establecer una interacción entre las proteínas con los compuestos químicos.

Las proteínas en su proceso de cristalización en varios casos presentan residuos y fragmentos de moléculas no deseadas, ligandos o incluso algunas de estas son susceptibles de grandes cambios al trabajar como monómeros o dímeros, por lo tanto, se sugiere el análisis bioinformático de tantas maneras sea posible y siempre tomando en consideración la limpieza y preparación del modelo para eliminar cualquier tipo de impureza que pueda interferir en la búsqueda de dianas terapéuticas mediante química computacional.

## REFERENCIAS BIBLIOGRÁFICAS

- Abdullahi, S. H., Uzairu, A., Shallangwa, G. A., Uba, S., & Umar, A. B. (2022). In-silico activity prediction, structure-based drug design, molecular docking and pharmacokinetic studies of selected quinazoline derivatives for their antiproliferative activity against triple negative breast cancer (MDA-MB231) cell line. *Bulletin of the National Research Centre*, 46(1). <https://doi.org/10.1186/s42269-021-00690-z>
- Abbas, M., & El-Manzalawy, Y. (2020). Machine learning based refined differential gene expression analysis of pediatric sepsis. *BMC Medical Genomics*, 13(1). <https://doi.org/10.1186/s12920-020-00771-4>
- Adorno, M., Cordenonsi, M., Montagner, M., Dupont, S., Wong, C., Hann, B., Solari, A., Bobisse, S., Rondina, M. B., Guzzardo, V., Parenti, A. R., Rosato, A., Bicciato, S., Balmain, A., & Piccolo, S. (2019). A Mutant-p53/Smad Complex Opposes p63 to Empower TGF $\beta$ -Induced Metastasis. *Cell*, 137(1), 87–98. <https://doi.org/10.1016/j.cell.2009.01.039>
- Agarwal, S., & Mehrotra, R. (2016). An Overview of Molecular Simulation. *JSM Chemistry*, 4(2), 1024–1028. [https://www.researchgate.net/profile/Shweta-Agarwal7/publication/303897563\\_Mini\\_Review\\_An\\_overview\\_of\\_Molecular\\_Docking/links/575b9fb108aed884620da00a/Mini-Review-An-overview-of-Molecular-Docking.pdf](https://www.researchgate.net/profile/Shweta-Agarwal7/publication/303897563_Mini_Review_An_overview_of_Molecular_Docking/links/575b9fb108aed884620da00a/Mini-Review-An-overview-of-Molecular-Docking.pdf)
- Aghabozorgi, A. S., Bahreyni, A., Soleimani, A., Bahrami, A., Khazaei, M., Ferns, G. A., Avan, A., & Hassanian, S. M. (2019). Role of adenomatous polyposis coli (APC) gene mutations in the pathogenesis of colorectal cancer; current status and perspectives. *Biochimie*, 157, 64–71. <https://doi.org/10.1016/j.biochi.2018.11.003>
- Alabugin, I. V., dos Passos Gomes, G., & Abdo, M. A. (2019). Hyperconjugation. In *Wiley Interdisciplinary Reviews: Computational Molecular Science* (Vol. 9, Issue 2). Blackwell Publishing Inc. <https://doi.org/10.1002/wcms.1389>
- André, T., Shiu, K.-K., Kim, T. W., Jensen, B. V., Jensen, L. H., Punt, C., Smith, D., Garcia-Carbonero, R., Benavides, M., Gibbs, P., de la Fouchardiere, C., Rivera, F., Elez, E., Bendell, J., Le, D. T., Yoshino, T., Van Cutsem, E., Yang, P., Farooqui, M. Z. H., ... Diaz, L. A. (2020). Pembrolizumab in Microsatellite-Instability–High Advanced Colorectal Cancer. *New England Journal of Medicine*, 383(23), 2207–2218. <https://doi.org/10.1056/nejmoa2017699>

- Ballón Paucara, W. G., & Grados Torrez, R. E. (2019). Acomplamiento molecular:: criterios prácticos para la selección de ligandos biológicamente activos e identificación de nuevos blancos terapéuticos. *Revista CON-CIENCIA*, 7(2), 55–72. [http://www.scielo.org.bo/scielo.php?script=sci\\_arttext&pid=S2310-02652019000200006&lng=es&nrm=iso&tlng=es](http://www.scielo.org.bo/scielo.php?script=sci_arttext&pid=S2310-02652019000200006&lng=es&nrm=iso&tlng=es).
- Bellio, H., Fumet, J. D., & Ghiringhelli, F. (2021). Targeting BRAF and RAS in colorectal cancer. In *Cancers* (Vol. 13, Issue 9). MDPI AG. <https://doi.org/10.3390/cancers13092201>.
- Bender, B. J., Gahbauer, S., Lutgens, A., Lyu, J., Webb, C. M., Stein, R. M., Fink, E. A., Balius, T. E., Carlsson, J., Irwin, J. J., & Shoichet, B. K. (2021). A practical guide to large-scale docking. *Nature Protocols* 2021 16:10, 16(10), 4799–4832. <https://doi.org/10.1038/s41596-021-00597-z>
- Benson, A. B., Venook, A. P., Al-Hawary, M. M., Arain, M. A., Chen, Y. J., Ciombor, K. K., Cohen, S., Cooper, H. S., Deming, D., Farkas, L., Garrido-Laguna, I., Grem, J. L., Gunn, A., Hecht, J. R., Hoffe, S., Hubbard, J., Hunt, S., Johung, K. L., Kirilcuk, N., ... Gurski, L. A. (2021). Colon cancer, Version 2.2021. *JNCCN Journal of the National Comprehensive Cancer Network*, 19(3), 329–359. <https://doi.org/10.6004/jnccn.2021.0012>.
- Benigni, R. (2021). In silico assessment of genotoxicity. Combinations of sensitive structural alerts minimize false negative predictions for all genotoxicity endpoints and can single out chemicals for which experimentation can be avoided. *Regulatory Toxicology and Pharmacology*, 126. <https://doi.org/10.1016/j.yrtph.2021.105042>
- Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 68(6), 394–424. <https://doi.org/10.3322/CAAC.21492>.
- Caputo, F., Santini, C., Bardasi, C., Cerma, K., Casadei-Gardini, A., Spallanzani, A., Andrikou, K., Cascinu, S., & Gelsomino, F. (2019). BRAF-mutated colorectal cancer: Clinical and molecular insights. In *International Journal of Molecular Sciences* (Vol. 20, Issue 21). MDPI AG. <https://doi.org/10.3390/ijms20215369>
- Chakrabarti, S., Peterson, C. Y., Sriram, D., & Mahipal, A. (2020). Early stage colon cancer: Current treatment standards, evolving paradigms, and future directions. *World Journal of Gastrointestinal Oncology*, 12(8), 808. <https://doi.org/10.4251/WJGO.V12.I8.808>

- Cheng, X., Xu, X., Chen, D., Zhao, F., & Wang, W. (2019). Therapeutic potential of targeting the Wnt/ $\beta$ -catenin signaling pathway in colorectal cancer. *Biomedicine & Pharmacotherapy*, *110*, 473–481. <https://doi.org/10.1016/J.BIOPHA.2018.11.082>.
- Chen, C. Y., Chen, J., He, L., & Stiles, B. L. (2018). PTEN: Tumor suppressor and metabolic regulator. In *Frontiers in Endocrinology* (Vol. 9, Issue JUL). Frontiers Media S.A. <https://doi.org/10.3389/fendo.2018.00338>
- Chen, D., Oezguen, N., Urvil, P., Ferguson, C., Dann, S. M., & Savidge, T. C. (2016). Regulation of protein-ligand binding affinity by hydrogen bond pairing. *Science Advances*, *2*(3). <https://doi.org/10.1126/sciadv.1501240>
- Chen, J., & Sadowski, I. (2015). Identification of the mismatch repair genes PMS2 and MLH1 as p53 target genes by using serial analysis of binding elements. [www.pnas.org/cgi/doi/10.1073/pnas.0407069102](http://www.pnas.org/cgi/doi/10.1073/pnas.0407069102)
- Chen, Y., Chen, M., & Deng, K. (2023). Blocking the Wnt/ $\beta$ -catenin signaling pathway to treat colorectal cancer: Strategies to improve current therapies (Review). In *International Journal of Oncology* (Vol. 62, Issue 2). Spandidos Publications. <https://doi.org/10.3892/IJO.2022.5472>
- Chen, Y., Lun, A. T. L., & Smyth, G. K. (2016). From reads to genes to pathways: Differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. *F1000Research*, *5*. <https://doi.org/10.12688/F1000RESEARCH.8987.2>
- Cillessen, S. A. G. M., Hess, C. J., Hooijberg, E., Castricum, K. C. M., Kortman, P., Denkers, F., Vos, W., Van De Wiel, M. A., Schuurhuis, G. J., Ossenkoppele, G. J., Meijer, C. J. L. M., & Oudejans, J. J. (2017). Inhibition of the intrinsic apoptosis pathway downstream of caspase-9 activation causes chemotherapy resistance in diffuse large B-cell lymphoma. *Clinical Cancer Research*, *13*(23), 7012–7021. <https://doi.org/10.1158/1078-0432.CCR-06-2891>
- Chowdhury, M. R. H., Bhuiyan, M. I. K., Saha, A., Mosleh, I. M., Mondol, S., & Ahmed, C. M. S. (2014). Identification and analysis of potential targets in *Streptococcus sanguinis* using computer aided protein data analysis. *Advances and Applications in Bioinformatics and Chemistry*, *7*(1), 45–54. <https://doi.org/10.2147/AABC.S67336>
- Cohen, Y., Merhavi-Shoham, E., Avraham, R. B., Frenkel, S., Pe'er, J., & Goldenberg-Cohen, N. (2018). Hypermethylation of CpG island loci of multiple tumor suppressor genes in retinoblastoma. *Experimental Eye Research*, *86*(2), 201–206. <https://doi.org/10.1016/J.EXER.2007.10.010>

- Cordenonsi, M., Dupont, S., Maretto, S., Insinga, A., Imbriano, C., Piccolo, S., & Activin, N. (2013). Links between Tumor Suppressors: p53 Is Required for TGF-Gene Responses by Cooperating with Smads phosphorylation and activation of the Smad family of signal transducers. Two different Smad signaling branches have been described: TGF-like signals, including TGF. In *Cell* (Vol. 113). <http://www.cell.com/cgi/content/full/113/>
- Cox, D. (2021). How not to discover a drug - integrins. In *Expert Opinion on Drug Discovery* (Vol. 16, Issue 2, pp. 197–211). *Taylor and Francis Ltd.* <https://doi.org/10.1080/17460441.2020.1819234>
- Dalal, N., Jalandra, R., Sharma, M., Prakash, H., Makharia, G. K., Solanki, P. R., Singh, R., & Kumar, A. (2020). Omics technologies for improved diagnosis and treatment of colorectal cancer: Technical advancement and major perspectives. *Biomedicine & Pharmacotherapy*, 131, 110648. <https://doi.org/10.1016/J.BIOPHA.2020.110648>
- Dar, A. M., & Mir, S. (2017). Molecular Docking: Approaches, Types, Applications and Basic Challenges. *Journal of Analytical & Bioanalytical Techniques* 2017 8:2, 8(2), 1–3. <https://doi.org/10.4172/2155-9872.1000356>
- De'angelis, G. L., Bottarelli, L., Azzoni, C., De'angelis, N., Leandro, G., Di Mario, F., Gaiani, F., & Negri, F. (2018). Microsatellite instability in colorectal cancer. *Acta Biomedica*, 89(6), 97–101. <https://doi.org/10.23750/abm.v89i9-S.7960>
- De Vivo, M., Ensing, B., & Klein, M. L. (2005). Computational study of phosphatase activity in soluble epoxide hydrolase: High efficiency through a water bridge mediated proton shuttle. *Journal of the American Chemical Society*, 127(32), 11226–11227. [https://doi.org/10.1021/JA053049J/SUPPL\\_FILE/JA053049JSI20050707\\_093259](https://doi.org/10.1021/JA053049J/SUPPL_FILE/JA053049JSI20050707_093259)
- Dong, B., Chai, M., Chen, H., Feng, Q., Jin, R., & Hu, S. (2020). Screening and verifying key genes with poor prognosis in colon cancer through bioinformatics analysis. *Translational Cancer Research*, 9(11), 6720–6732. <https://doi.org/10.21037/TCR-20-2309>
- Duraiyaran, S., Adefuye, M., Manjunatha, N., Ganduri, V., & Rajasekaran, K. (2022). Colon Cancer and Obesity: A Narrative Review. *Cureus*, 14(8), 1–9. <https://doi.org/10.7759/cureus.27589>
- ECIS. (2022). Incidence and mortality estimates 2020. *European Cancer Information System*, 72(2), <https://ecis.jrc.ec.europa.eu/>.
- Elston, R., & Inman, G. J. (2012). Crosstalk between p53 and TGF-  $\beta$  Signalling . *Journal*

- of Signal Transduction, 2012, 1–10. <https://doi.org/10.1155/2012/294097>.
- Esmeeta, A., Adhikary, S., Dharshnaa, V., Swarnamughi, P., Ummul Maqsummiya, Z., Banerjee, A., Pathak, S., & Duttaroy, A. K. (2022). Plant-derived bioactive compounds in colon cancer treatment: An updated review. *Biomedicine & Pharmacotherapy = Biomedecine & Pharmacotherapie*, 153. <https://doi.org/10.1016/J.BIOPHA.2022.113384>
- Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Židek, A., Bates, R., Blackwell, S., Yim, J., Ronneberger, O., Bodenstein, S., Zielinski, M., Bridgland, A., Potapenko, A., Cowie, A., Tunyasuvunakool, K., Jain, R., Clancy, E., Hassabis, D. (2022). Protein complex prediction with AlphaFold-Multimer. *BioRxiv*, 2021.10.04.463034. <https://doi.org/10.1101/2021.10.04.463034>
- Evans, C., Hardin, J., & Stoebel, D. M. (2018). Selecting between-sample RNA-Seq normalization methods from the perspective of their assumptions. *Briefings in Bioinformatics*, 19(5), 776–792. <https://doi.org/10.1093/bib/bbx008>
- Fan, J., Fu, A., & Zhang, L. (2019). Progress in molecular docking. *Quantitative Biology*, 7(2), 83–89. <https://doi.org/10.1007/S40484-019-0172-Y>
- Fährrolfes, R., Bietz, S., Flachsenberg, F., Meyder, A., Nittinger, E., Otto, T., Volkamer, A., & Rarey, M. (2017). Proteins Plus: A web portal for structure analysis of macromolecules. *Nucleic Acids Research*, 45(W1), W337–W343. <https://doi.org/10.1093/nar/gkx333>
- Ferlay, J., Colombet, M., Soerjomataram, I., Mathers, C., Parkin, D. M., Piñeros, M., Znaor, A., & Bray, F. (2019). Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *International Journal of Cancer*, 144(8), 1941–1953. <https://doi.org/10.1002/ijc.31937>
- Firdausy, A. F., Muti'ah, R., & Rahmawati, E. K. (2020). Prediction Pharmacokinetic profiles of sunflower's (*Helianthus annuus* L.) Active Compunds using in silico approach. *Journal of Medicine*, 4(1), 1–7. <http://biosig.unimelb.edu.au/pkcsmp/predi>
- Fitri, A., Basultan, H., & Iryani. (2021). Hydrophobic Pocket of SARS-Cov-2 Spike Glycoprotein are Potential as Binding Pocket. *Journal of Physics: Conference Series*, 1788(1). <https://doi.org/10.1088/1742-6596/1788/1/012021>
- Furlan, D., Casati, B., Cerutti, R., Facco, C., Terracciano, L., Capella, C., & Chiaravalli, A. M. (2002). Genetic progression in sporadic endometrial and gastrointestinal cancers with high microsatellite instability. *Journal of Pathology*, 197(5), 603–609. <https://doi.org/10.1002/path.1162>

- Gadaleta, D., Vuković, K., Toma, C., Lavado, G. J., Karmaus, A. L., Mansouri, K., Kleinstreuer, N. C., Benfenati, E., & Roncaglioni, A. (2019). SAR and QSAR modeling of a large collection of LD50 rat acute oral toxicity data. *Journal of Cheminformatics*, 11(1). <https://doi.org/10.1186/s13321-019-0383-2>
- Garrido, A., Lepailleur, A., Mignani, S. M., Dallemagne, P., & Rochais, C. (2020). hERG toxicity assessment: useful guidelines for drug design.
- Gore, S., Sanz García, E., Hendrickx, P. M. S., Gutmanas, A., Westbrook, J. D., Yang, H., Feng, Z., Baskaran, K., Berrisford, J. M., Hudson, B. P., Ikegawa, Y., Kobayashi, N., Lawson, C. L., Mading, S., Mak, L., Mukhopadhyay, A., Oldfield, T. J., Patwardhan, A., Peisach, E., Kleywegt, G. J. (2017). *Validation of Structures in the Protein Data Bank. Structure* (London, England:1993), 25(12), 1916. <https://doi.org/10.1016/J.STR.2017.10.009>
- Granados-Romero, J. J., Valderrama-Treviño, A. I., Contreras-Flores, E. H., Barrera-Mera, B., Herrera Enríquez, M., Uriarte-Ruiz, K., Ceballos-Villalba, J. C., Estrada-Mata, A. G., Alvarado Rodríguez, C., & Arauz-Peña, G. (2017). Colorectal cancer: a review. *International Journal of Research in Medical Sciences*, 5(11), 4667. <https://doi.org/10.18203/2320-6012.ijrms20174914>
- Grogan, S., Charles, & Preuss, V. (2022). Pharmacokinetics. <https://www.ncbi.nlm.nih.gov/books/NBK557744/>
- Guide, P. P. (2015). *Schrodinger suite 2015-2, Protein Preparation Wizard Epik*.
- Gu, S., Smith, M. S., Yang, Y., Irwin, J. J., & Shoichet, B. K. (2021). Ligand Strain Energy in Large Library Docking. *Journal of Chemical Information and Modeling*. <https://doi.org/10.1021/acs.jcim.1c00368>
- Guo, T., Hou, D., & Yu, D. (2019). Bioinformatics analysis of gene expression profile data to screen key genes involved in intracranial aneurysms. *Molecular Medicine Reports*, 20(5), 4415–4424. <https://doi.org/10.3892/mmr.2019.10696>
- Guedes, I. A., de Magalhães, C. S., & Dardenne, L. E. (2014). Receptor-ligand molecular docking. *In Biophysical Reviews* (Vol. 6, Issue 1, pp. 75–87). <https://doi.org/10.1007/s12551-013-0130-2>
- Guan, D., & Tian, H. (2017). Integrated network analysis to explore the key genes regulated by parathyroid hormone receptor 1 in osteosarcoma. *World Journal of Surgical Oncology*, 15(1). <https://doi.org/10.1186/s12957-017-1242-0>
- Han, Y., Zhang, J., Hu, C. Q., Zhang, X., Ma, B., & Zhang, P. (2019). In silico ADME and toxicity prediction of ceftazidime and its impurities. *Frontiers in Pharmacology*,

- 10(APR). <https://doi.org/10.3389/fphar.2019.00434>.
- Hempel, G. (2020). Dose and therapy individualization in cancer chemotherapy. In *Handbook of Analytical Separations (Vol. 7, pp. 291–319)*. Elsevier B.V. <https://doi.org/10.1016/B978-0-444-64066-6.00013-7>
- Hong, S. N. (2018). Genetic and epigenetic alterations of colorectal cancer. *Intestinal Research, 16(3)*, 327. <https://doi.org/10.5217/IR.2018.16.3.327>
- Hossain, M. S., Karuniawati, H., Jairoun, A. A., Urbi, Z., Ooi, D. J., John, A., Lim, Y. C., Kaderi Kibria, K. M., Mohiuddin, A. K. M., Ming, L. C., Goh, K. W., & Hadi, M. A. (2022). Colorectal Cancer: A Review of Carcinogenesis, Global Epidemiology, Current Challenges, Risk Factors, Preventive and Treatment Strategies. *Cancers, 14(7)*, 1732. <https://doi.org/10.3390/CANCERS14071732>
- Huang, Z., & Yang, M. (2022). Molecular Network of Colorectal Cancer and Current Therapeutic Options. *Frontiers in Oncology, 12(April)*, 1–13. <https://doi.org/10.3389/fonc.2022.852927>
- Huggins, D. J., Sherman, W., & Tidor, B. (2012). Rational approaches to improving selectivity in drug design. *Journal of Medicinal Chemistry, 55(4)*, 1424–1444. <https://doi.org/10.1021/JM2010332>
- Iqbal, M. S., Ahmad, W., Alizadehsani, R., Hussain, S., & Rehman, R. (2022). Breast Cancer Dataset, Classification and Detection Using Deep Learning. *Healthcare 2022, Vol. 10, Page 2395, 10(12), 2395*. <https://doi.org/10.3390/HEALTHCARE10122395>
- Infield, D. T., Rasouli, A., Galles, G. D., Chipot, C., Tajkhorshid, E., & Ahern, C. A. (2021). Cation- $\pi$  Interactions and their Functional Roles in Membrane Proteins: Cation- $\pi$  interactions in membrane proteins. In *Journal of Molecular Biology (Vol. 433, Issue 17)*. Academic Press. <https://doi.org/10.1016/j.jmb.2021.167035>
- Islam, M. E., & Kalimuthu, R. (2019). Assumptions Of Theory Z: A Tool For Managing People At Work. *The 1St Multidisciplinary Academic Research International Conference (Maric) 2019, 1(1)*, 364–375.
- Jain, T., Kumar, P., Thomas, & High, J. (2017). In Silico Prediction of Blood-Brain Barrier Permeability of Chemical Compounds through Molecular Feature Modeling.
- Jayanthi Pazhani, Vishnu Priya Veeraraghavan & Selvaraj Jayaraman. (30 de April de 2023). Molecular docking analysis of imiquimod with the TGF- $\beta$  targets for oral carcinoma. *Biomedical Informations, 4*.

- Jones, J. C., Renfro, L. A., Al-Shamsi, H. O., Schrock, A. B., Rankin, A., Zhang, B. Y., Kasi, P. M., Voss, J. S., Leal, A. D., Sun, J., Ross, J., Ali, S. M., Hubbard, J. M., Kipp, B. R., McWilliams, R. R., Kopetz, S., Wolff, R. A., & Grothey, A. (2017). Non-V600 BRAF Mutations Define a Clinically Distinct Molecular Subtype of Metastatic Colorectal Cancer. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology*, 35(23), 2624–2630. <https://doi.org/10.1200/JCO.2016.71.4394>
- Karim, S., Brennan, K., Nanji, S., Berry, S. R., & Booth, C. M. (2017). Association Between Prognosis and Tumor Laterality in Early-Stage Colon Cancer. *JAMA Oncology*, 3(10), 1386–1392. <https://doi.org/10.1001/JAMAONCOL.2017.1016>
- Keum, N. N., & Giovannucci, E. (2019). Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies. *Nature Reviews Gastroenterology & Hepatology* 2019 16:12, 16(12), 713–732. <https://doi.org/10.1038/s41575-019-0189-8>
- Kim, S. H., Park, K. H., Shin, S. J., Lee, K. Y., Kim, T. Il, Kim, N. K., Rha, S. Y., & Ahn, J. B. (2018). CpG Island Methylator Phenotype and Methylation of Wnt Pathway Genes Together Predict Survival in Patients with Colorectal Cancer. 59(5), 588–594.
- Labbé, C. M., Rey, J., Lagorce, D., Vavruša, M., Becot, J., Sperandio, O., Villoutreix, B. O., Tufféry, P., & Miteva, M. A. (2015). MTiOpenScreen: a web server for structure-based virtual screening. *Nucleic Acids Research*, 43(Web Server issue), W448. <https://doi.org/10.1093/NAR/GKV306>
- Li, H., Huang, K., Liu, X., Liu, J., Lu, X., Tao, K., Wang, G., & Wang, J. (2014). Lithium chloride suppresses colorectal cancer cell survival and proliferation through ROS/GSK-3 $\beta$ /NF- $\kappa$ B signaling pathway. *Oxidative Medicine and Cellular Longevity*, 2014. <https://doi.org/10.1155/2014/241864>.
- Li, X., Wu, Y., & Tian, T. (2022). TGF- $\beta$  Signaling in Metastatic Colorectal Cancer (mCRC): From Underlying Mechanism to Potential Applications in Clinical Development. In *International Journal of Molecular Sciences* (Vol. 23, Issue 22). MDPI. <https://doi.org/10.3390/ijms232214436>.
- Liu, Y. P., Zheng, C. C., Huang, Y. N., He, M. L., Xu, W. W., & Li, B. (2021). Molecular mechanisms of chemo- and radiotherapy resistance and the potential implications for cancer treatment. *MedComm*, 2(3), 315. <https://doi.org/10.1002/MCO2.55>
- Liu, L., Chen, F., Jiang, S., Zhong, B., Li, W., Xu, K., Wang, Q., Wang, Y., & Cao, J.

- (2022). Analysis of gene expression profile of peripheral blood in alveolar and cystic echinococcosis. *Frontiers in Cellular and Infection Microbiology*, 12. <https://doi.org/10.3389/fcimb.2022.913393>
- Liu, T., Oprea, T., Ursu, O., Hasselgren, C., & Altman, R. B. (2016). Estimation of Maximum Recommended Therapeutic Dose Using Predicted Promiscuity and Potency. *Clinical and Translational Science*, 9(6), 311–320. <https://doi.org/10.1111/cts.12422>
- Maden, S. F., Sezer, S., Acuner, S. E., Maden, S. F., Sezer, S., & Acuner, S. E. (2022). *Fundamentals of Molecular Docking and Comparative Analysis of Protein–Small-Molecule Docking Approaches*. <https://doi.org/10.5772/INTECHOPEN.105815>
- Madhavi Sastry, G., Adzhigirey, M., Day, T., Annabhimoju, R., & Sherman, W. (2013). Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments. *Journal of Computer-Aided Molecular Design*, 27(3), 221–234. <https://doi.org/10.1007/s10822-013-9644-8>
- Maia, E. H. B., Assis, L. C., de Oliveira, T. A., da Silva, A. M., & Taranto, A. G. (2020). Structure-Based Virtual Screening: From Classical to Artificial Intelligence. In *Frontiers in Chemistry* (Vol. 8). *Frontiers Media S.A.* <https://doi.org/10.3389/fchem.2020.00343>
- Malki, A., Elruz, R. A., Gupta, I., Allouch, A., Vranic, S., & Al Moustafa, A. E. (2021). Molecular mechanisms of colon cancer progression and metastasis: Recent insights and advancements. *International Journal of Molecular Sciences*, 22(1), 1–24. <https://doi.org/10.3390/ijms22010130>
- Mansoor, A., & Mahabadi, N. (2022). Volume of Distribution. <https://www.ncbi.nlm.nih.gov/books/NBK545280/>
- Michel, M., Visnes, T., Homan, E. J., Seashore-Ludlow, B., Hedenström, M., Wiita, E., Vallin, K., Paulin, C. B. J., Zhang, J., Wallner, O., Scobie, M., Schmidt, A., Jenmalm-Jensen, A., Warpman Berglund, U., & Helleday, T. (2019). Computational and Experimental Druggability Assessment of Human DNA Glycosylases. *ACS Omega*, 4(7), 11642–11656. <https://doi.org/10.1021/acsomega.9b00162>
- Mitra, S., Chakraborty, A. J., Tareq, A. M., Emran, T. Bin, Nainu, F., Khusro, A., Idris, A. M., Khandaker, M. U., Osman, H., Alhumaydhi, F. A., & Simal-Gandara, J. (2022). Impact of heavy metals on the environment and human health: Novel therapeutic insights to counter the toxicity. *Journal of King Saud University - Science*, 34(3). <https://doi.org/10.1016/j.jksus.2022.101865>.

- Montecucco, A., Zanetta, F., & Biamonti, G. (2015). Molecular mechanisms of etoposide. In *EXCLI Journal* (Vol. 14, pp. 95–108). Leibniz Research Centre for Working Environment and Human Factors. <https://doi.org/10.17179/excli2014-561>
- Mustafa, G., Younas, S., Mahrosh, H. S., Albeshr, M. F., & Bhat, E. A. (2023). Molecular Docking and Simulation-Binding Analysis of Plant Phytochemicals with the Hepatocellular Carcinoma Targets Epidermal Growth Factor Receptor and Caspase-9. *Molecules*, 28(8). <https://doi.org/10.3390/molecules28083583>
- Nakayama, I., Hirota, T., & Shinozaki, E. (2020). BRAF Mutation in Colorectal Cancers: From Prognostic Marker to Targetable Mutation. *Cancers*, 12(11), 1–29. <https://doi.org/10.3390/CANCERS12113236>
- Nemade, H., Chaudhari, U., Acharya, A., Hescheler, J., Hengstler, J. G., Papadopoulos, S., & Sachinidis, A. (2018). Cell death mechanisms of the anti-cancer drug etoposide on human cardiomyocytes isolated from pluripotent stem cells. *Archives of Toxicology*, 92(4), 1507–1524. <https://doi.org/10.1007/s00204-018-2170-7>
- Nelson, T. J., Sun, M. K., Lim, C., Sen, A., Khan, T., Chirila, F. V., & Alkon, D. L. (2017). Bryostatin Effects on Cognitive Function and PKC $\epsilon$  in Alzheimer's Disease Phase IIa and Expanded Access Trials. *Journal of Alzheimer's Disease*, 58(2), 521–535. <https://doi.org/10.3233/JAD-170161>
- Nojadeh, J. N., Sharif, S. B., & Sakhinia, E. (2018). Microsatellite instability in colorectal cancer. *EXCLI Journal*, 17, 159. <https://doi.org/10.17179/EXCLI2017-948>.
- Olkinuora, A. P., Peltomäki, P. T., Aaltonen, L. A., & Rajamäki, K. (2021). From APC to the genetics of hereditary and familial colon cancer syndromes. *Human Molecular Genetics*, 30(R2), R206–R224. <https://doi.org/10.1093/HMG/DDAB208>
- Orlov, Y. L., Anashkina, A. A., Klimontov, V. V., & Baranova, A. V. (2021). Medical Genetics, Genomics and Bioinformatics Aid in Understanding Molecular Mechanisms of Human Diseases. *International Journal of Molecular Sciences* 2021.. <https://doi.org/10.3390/IJMS22189962>
- Papadatos-Pastos, D., Rabbie, R., Ross, P., & Sarker, D. (2015). The role of the PI3K pathway in colorectal cancer. In *Critical Reviews in Oncology/Hematology* (Vol. 94, Issue 1, pp. 18–30). Elsevier Ireland Ltd. <https://doi.org/10.1016/j.critrevonc.2014.12.006>
- Palermo, G., & Diego, S. (2016). Encyclopedia of Nanotechnology. *Encyclopedia of Nanotechnology*, September, 0–15. <https://doi.org/10.1007/978-94-007-6178-0>
- Paper, O. (2017). *Characterization of TP53 polymorphisms in the Romania colorectal*

*cancer patients. x.*

- Parrish, A. B., Freil, C. D., & Kornbluth, S. (2013). Cellular mechanisms controlling caspase activation and function. *Cold Spring Harbor Perspectives in Biology*, 5(6).
- Peres, R. B., Fiuza, L. F. de A., da Silva, P. B., Batista, M. M., Camillo, F. da C., Marques, A. M., Brito, L. de C., Figueiredo, M. R., & Soeiro, M. de N. C. (2021). In vitro phenotypic activity and in silico analysis of natural products from brazilian biodiversity on trypanosoma cruzi. *Molecules*, 26(18). <https://doi.org/10.3390/molecules26185676>
- Peterson, L. E. (2019). Small molecule docking of DNA repair proteins associated with cancer survival following PCNA metagene adjustment: *A potential novel class of repair inhibitors. Molecules*, 24(3). <https://doi.org/10.3390/molecules24030645>
- Pino, M. S., & Chung, D. C. (2014). *THE CHROMOSOMAL INSTABILITY PATHWAY IN COLON. 138(6)*, 2059–2072. <https://doi.org/10.1053/j.gastro.2009.12.065>.
- Pires, D. E. V., Blundell, T. L., & Ascher, D. B. (2015). pkCSM: Predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. *Journal of Medicinal Chemistry*, 58(9), 4066–4072. <https://doi.org/10.1021/acs.jmedchem.5b00104>
- Pećina-Šlaus, N., Kafka, A., Salamon, I., & Bukovac, A. (2020). Mismatch Repair Pathway, Genome Stability and Cancer. In *Frontiers in Molecular Biosciences (Vol. 7)*. Frontiers Media S.A. <https://doi.org/10.3389/fmolb.2020.00122>
- Proietti, I., Skroza, N., Michelini, S., Mambrin, A., Balduzzi, V., Bernardini, N., Marchesiello, A., Tolino, E., Volpe, S., Maddalena, P., Fraia, M., Mangino, G., Romeo, G., & Potenza, C. (2020). BRAF inhibitors: Molecular targeting and immunomodulatory actions. In *Cancers (Vol. 12, Issue 7, pp. 1–13)*. MDPI AG. <https://doi.org/10.3390/cancers12071823>
- Raval, K., & Ganatra, T. (2022). Basics, types and applications of molecular docking: A review. *IP International Journal of Comprehensive and Advanced Pharmacology*, 7(1), 12–16. <https://doi.org/10.18231/J.IJCAAP.2022.003>
- Rhee, Y. Y., Kim, K. J., & Kang, G. H. (2017). CpG Island Methylator Phenotype-High Colorectal Cancers and Their Prognostic Implications and Relationships with the Serrated Neoplasia Pathway. *Gut and Liver*, 11(1), 38. <https://doi.org/10.5009/GNL15535>
- Ringer, A. L., Senenko, A., & Sherrill, C. D. (2017). Models of S/ $\pi$  interactions in protein structures: Comparison of the H<sub>2</sub> S-benzene complex with PDB data . *Protein*

- Science*, 16(10), 2216–2223. <https://doi.org/10.1110/ps.073002307>
- Robinson, K., & Tiriveedhi, V. (2020). Perplexing Role of P-Glycoprotein in Tumor Microenvironment. In *Frontiers in Oncology* (Vol. 10). *Frontiers Media S.A.* <https://doi.org/10.3389/fonc.2020.00265>
- Sawicki, T., Ruszkowska, M., Danielewicz, A., Niedźwiedzka, E., Arłukowicz, T., & Przybyłowicz, K. E. (2021). A Review of Colorectal Cancer in Terms of Epidemiology, Risk Factors, Development, Symptoms and Diagnosis. *Cancers* 2021, Vol. 13, Page 2025, 13(9), 2025. <https://doi.org/10.3390/CANCERS13092025>
- Scardino, V., Di Filippo, J. I., & Cavasotto, C. N. (2023). How good are AlphaFold models for docking-based virtual screening? *IScience*, 26(1). <https://doi.org/10.1016/j.isci.2022.105920>
- Shen, A., Liu, L., Huang, Y., Shen, Z., Wu, M., Chen, X., Wu, X., Lin, X., Chen, Y., Li, L., Cheng, Y., Chu, J., Sferra, T. J., Wei, L., Zhuang, Q., & Peng, J. (2022). Down-Regulating HAUS6 Suppresses Cell Proliferation by Activating the p53/p21 Pathway in Colorectal Cancer. *Frontiers in Cell and Developmental Biology*, 9. <https://doi.org/10.3389/fcell.2021.772077>
- Shen, L., Toyota, M., Kondo, Y., Lin, E., Zhang, L., Guo, Y., Hernandez, N. S., Chen, X., Ahmed, S., Konishi, K., Hamilton, S. R., & Issa, J. P. J. (2017). Integrated genetic and epigenetic analysis identifies three different subclasses of colon cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 104(47), 18654–18659. <https://doi.org/10.1073/PNAS.0704652104>
- Simon, K. (2016). Colorectal cancer development and advances in screening. *Clinical Interventions in Aging*, 11, 967–976. <https://doi.org/10.2147/CIA.S109285>
- Sitepu, R. K., Natzir, R., Warsinggih, Hatta, M., Rudiman, R., Labeda, I., Lusikooy, R. E., Bukhari, A., Miskad, U. A., & Bahar, B. (2022). Relation between expression of hMLH1 and p53 mRNA genes, in the feces of patients with colorectal carcinoma. Cross-sectional study. *Annals of Medicine and Surgery*, 73. <https://doi.org/10.1016/j.amsu.2021.103237>
- Sliwoski, G., Kothiwale, S., Meiler, J., & Lowe, E. W. (2014). Computational Methods in Drug Discovery. *Pharmacological Reviews*, 66(1), 334. <https://doi.org/10.1124/PR.112.007336>
- Smit, W. L., Spaan, C. N., De Boer, R. J., Ramesh, P., Garcia, T. M., Meijer, B. J., Vermeulen, J. L. M., Lezzerini, M., MacInnes, A. W., Koster, J., Medema, J. P., Van

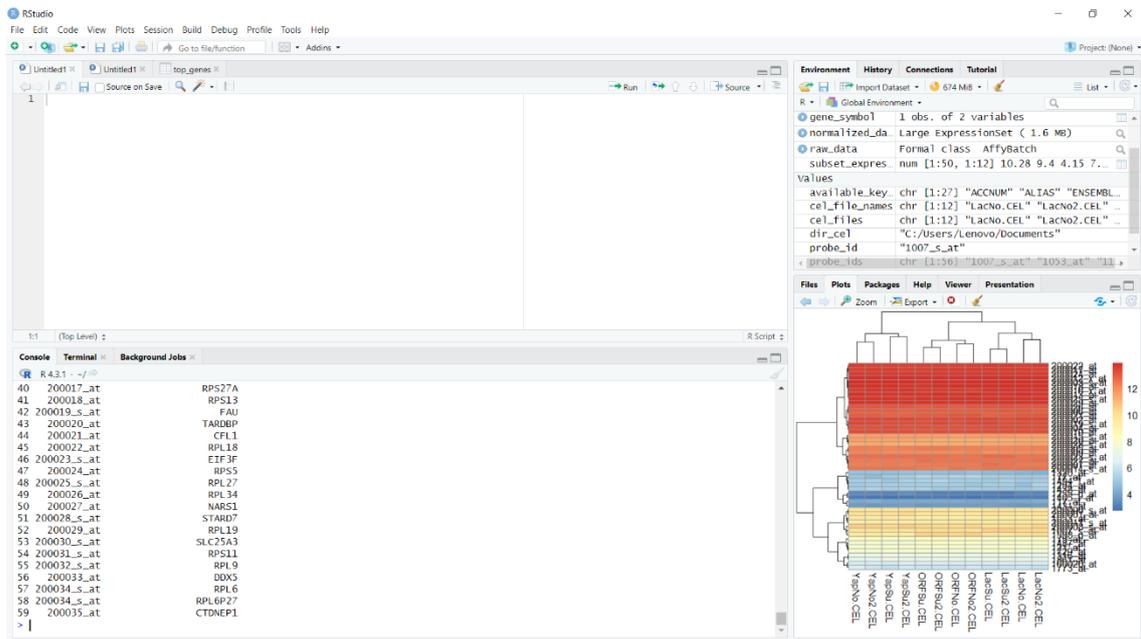
- Den Brink, G. R., Muncan, V., & Heijmans, J. (2020). Driver mutations of the adenoma-carcinoma sequence govern the intestinal epithelial global translational capacity. *Proceedings of the National Academy of Sciences of the United States of America*,  
[https://doi.org/10.1073/PNAS.1912772117/SUPPL\\_FILE/PNAS.1912772117.SM02.MP4](https://doi.org/10.1073/PNAS.1912772117/SUPPL_FILE/PNAS.1912772117.SM02.MP4).
- Stefani, C., Miricescu, D., Stanescu-Spinu, I. I., Nica, R. I., Greabu, M., Totan, A. R., & Jinga, M. (2021). Growth factors, pi3k/akt/mTOR and MAPK signaling pathways in colorectal cancer pathogenesis: Where are we now? *International Journal of Molecular Sciences*, 22(19). <https://doi.org/10.3390/ijms221910260>
- Szklarczyk, D., Santos, A., Von Mering, C., Jensen, L. J., Bork, P., & Kuhn, M. (2016). STITCH 5: Augmenting protein-chemical interaction networks with tissue and affinity data. *Nucleic Acids Research*, 44(D1), D380–D384. <https://doi.org/10.1093/nar/gkv1277>
- Testa, U., Pelosi, E., & Castelli, G. (2018). Colorectal Cancer: Genetic Abnormalities, Tumor Progression, Tumor Heterogeneity, Clonal Evolution and Tumor-Initiating Cells. *Medical Sciences*. <https://doi.org/10.3390/MEDSCI6020031>.
- Varma, A. K., Patil, R., Das, S., Stanley, A., Yadav, L., & Sudhakar, A. (2020). Optimized hydrophobic interactions and hydrogen bonding at the target-ligand interface leads the pathways of Drug-Designing. *PLoS ONE*, 5(8). <https://doi.org/10.1371/journal.pone.0012029>
- Vian, M., Raitano, G., Roncaglioni, A., & Benfenati, E. (2019). In silico model for mutagenicity (Ames test), taking into account metabolism. *Mutagenesis*, 34(1), 55–65. <https://doi.org/10.1093/mutage/gey045>
- Volkamer, A., Kuhn, D., Rippmann, F., & Rarey, M. (2012). Dogsitescorer: A web server for automatic binding site prediction, analysis and druggability assessment. *Bioinformatics*, 28(15), 2074–2075. <https://doi.org/10.1093/bioinformatics/bts310>
- Wang, Z. (2018). Antisense RNA and Cancer. In *Cancer and Noncoding RNAs*. Elsevier. <https://doi.org/10.1016/B978-0-12-811022-5.00012-7>.
- Wang, J.-C., & Lin, J.-H. (2017). Scoring functions for prediction of protein-ligand interactions. *Current Pharmaceutical Design*. <https://doi.org/10.2174/1381612811319120005>
- Watanabe, R., Esaki, T., Kawashima, H., Natsume-Kitatani, Y., Nagao, C., Ohashi, R., & Mizuguchi, K. (2018). Predicting Fraction Unbound in Human Plasma from

- Chemical Structure: Improved Accuracy in the Low Value Ranges. *Molecular Pharmaceutics*, 15(11), 5302–5311. <https://doi.org/10.1021/acs.molpharmaceut.8b00785>
- Watanabe, R., Ohashi, R., Esaki, T., Kawashima, H., Natsume-Kitatani, Y., Nagao, C., & Mizuguchi, K. (2019). Development of an in silico prediction system of human renal excretion and clearance from chemical structure information incorporating fraction unbound in plasma as a descriptor. *Scientific Reports*, 9(1). <https://doi.org/10.1038/s41598-019-55325-1>
- Williams, A. B., & Schumacher, B. (2016). p53 in the DNA-damage-repair process. *Cold Spring Harbor Perspectives in Medicine*, 6(5). <https://doi.org/10.1101/cshperspect.a026070>
- Wise, H. C., Iyer, G. V., Moore, K., Temkin, S. M., Gordon, S., Aghajanian, C., & Grisham, R. N. (2019). Activity of M3814, an Oral DNA-PK Inhibitor, In Combination with Topoisomerase II Inhibitors in Ovarian Cancer Models. *Scientific Reports*, 9(1). <https://doi.org/10.1038/s41598-019-54796-6>
- World Health Organization, International Agency for Ressearch on Cnacer . (3 de mayo de 2023).GLOBOCAN. [https://gco.iarc.fr/today/online-analysis-pie?v=2020&mode=cancer&mode\\_population=continents&population=900&populations=218&key=total&sex=0&cancer=39&type=0&statistic=5&prevalence=0&population\\_group=0&ages\\_group%5B%5D=0&ages\\_group%5B%5D=17&nb\\_items=7&group\\_cancer=1&include\\_nmssc=1&include\\_nmssc\\_other=1&half\\_pie=0&donut=0#collapse-group-0-1](https://gco.iarc.fr/today/online-analysis-pie?v=2020&mode=cancer&mode_population=continents&population=900&populations=218&key=total&sex=0&cancer=39&type=0&statistic=5&prevalence=0&population_group=0&ages_group%5B%5D=0&ages_group%5B%5D=17&nb_items=7&group_cancer=1&include_nmssc=1&include_nmssc_other=1&half_pie=0&donut=0#collapse-group-0-1).
- Wu, X. L., Wang, L. K., Yang, D. D., Qu, M., Yang, Y. J., Guo, F., Han, L., & Xue, J. (2018). Effects of Glut1 gene silencing on proliferation, differentiation, and apoptosis of colorectal cancer cells by targeting the TGF- $\beta$ /PI3K-AKT-mTOR signaling pathway. In *Journal of Cellular Biochemistry*. Wiley-Liss Inc. <https://doi.org/10.1002/jcb.26399>
- Xi, & Xu, P. (2021). Global colorectal cancer burden in 2020 and projections to 2040. *Translational Oncology*, 14(10), 101174. <https://doi.org/10.1016/j.tranon.2021.101174>
- Xie, Y. H., Chen, Y. X., & Fang, J. Y. (2020). Comprehensive review of targeted therapy for colorectal cancer. *Signal Transduction and Targeted Therapy* 2020 5:1, 5(1), 1–30. <https://doi.org/10.1038/s41392-020-0116-z>
- Xiao, F., Chen, Z., Wei, Z., & Tian, L. (2020). Hydrophobic Interaction: A Promising

- Driving Force for the Biomedical Applications of Nucleic Acids. In *Advanced Science* (Vol. 7, Issue 16). John Wiley and Sons Inc. <https://doi.org/10.1002/advs.202001048>
- Yang, J., Nie, J., Ma, X., Wei, Y., Peng, Y., & Wei, X. (2019). Targeting PI3K in cancer: Mechanisms and advances in clinical trials. In *Molecular Cancer*. BioMed Central Ltd. <https://doi.org/10.1186/s12943-019-0954-x>
- Yang, W., Ma, J., Zhou, W., Li, Z., Zhou, X., Cao, B., Zhang, Y., Liu, J., Yang, Z., Zhang, H., Zhao, Q., Hong, L., & Fan, D. (2019). Identification of hub genes and outcome in colon cancer based on bioinformatics analysis. *Cancer Management and Research*, *11*, 323–338. <https://doi.org/10.2147/CMAR.S173240>.
- Yeoh, Y., Low, T. Y., Abu, N., & Lee, P. Y. (2021). Regulation of signal transduction pathways in colorectal cancer: Implications for therapeutic resistance. *PeerJ*, *9*. <https://doi.org/10.7717/peerj.12338>.
- Yin, Y., Hua, H., Li, M., Liu, S., Kong, Q., Shao, T., Wang, J., Luo, Y., Wang, Q., Luo, T., & Jiang, Y. (2016). mTORC2 promotes type I insulin-like growth factor receptor and insulin receptor activation through the tyrosine kinase activity of mTOR. *Cell Research*. <https://doi.org/10.1038/cr.2015.133> .
- Yu, C., Chen, F., Jiang, J., Zhang, H., & Zhou, M. (2019). Screening key genes and signaling pathways in colorectal cancer by integrated bioinformatics analysis. *Molecular Medicine Reports*, *20*(2), 1259–1269. <https://doi.org/10.3892/mmr.2019.10336>
- Yue, J., & López, J. M. (2020). Understanding MAPK signaling pathways in apoptosis. In *International Journal of Molecular Sciences* (Vol. 21, Issue 7). MDPI AG. <https://doi.org/10.3390/ijms21072346>
- Zhang, B., Li, H., Yu, K., & Jin, Z. (2022). Molecular docking-based computational platform for high-throughput virtual screening. *Ccf Transactions on High Performance Computing*, *4*(1), 63. <https://doi.org/10.1007/S42514-021-00086-5>
- Zhang, L., & Shay, J. W. (2017). Multiple Roles of APC and its Therapeutic Implications in Colorectal Cancer. *Journal of the National Cancer Institute*, *109*(8), 1–10. <https://doi.org/10.1093/jnci/djw332>
- Zhang, X., Zhang, W., & Cao, P. (2021). Advances in CpG Island Methylator Phenotype Colorectal Cancer Therapies. *Frontiers in Oncology*, *11*, 193. <https://doi.org/10.3389/FONC.2021.629390/BIBTEX>.
- Zhao, H., Ming, T., Tang, S., Ren, S., Yang, H., Liu, M., Tao, Q., & Xu, H. (2022). Wnt

- signaling in colorectal cancer: pathogenic role and therapeutic target. *Molecular Cancer*, 21(1), 1–34. <https://doi.org/10.1186/s12943-022-01616-7>.
- Zhou, Y., Zhu, J., Tong, T., Wang, J., Lin, B., & Zhang, J. (2019). A statistical normalization method and differential expression analysis for RNA-seq data between different species. *BMC Bioinformatics*, 20. <https://doi.org/10.1186/s12859>
- Zyla, R., Graham, T., Aronson, M., Velsher, L., Mrkonjic, M., & Turashvili, G. (2021). MLH1 epimutation is a rare mechanism for Lynch syndrome: A case report and review of the literature. *Genes Chromosomes and Cancer*, 60(9), 635–639. <https://doi.org/10.1002/gcc.22957>

# ANEXOS



Anexo A. Dataset en RStudio para el reconocimiento de genes de expresión

The screenshot shows the MalaCards search results for "Adenocarcinoma colon". The search results are displayed in a table with columns for #, Family, MCID, Name, MIFTS, and Score. The top results include:

#	Family	MCID	Name	MIFTS	Score
1		DFF043	Diffuse Gastric and Lobular Breast Cancer Syndrome	69	162.257
2		ADN016	Adenocarcinoma	64	155.821
3		CLR023	Colorectal Cancer	100	148.389
4		PLY162	Polyposis Syndrome, Hereditary Mixed_1	62	148.188
5		LNG064	Lung Cancer Susceptibility 3	74	146.273
6		CLR087	Colorectal Cancer 12	35	142.348
7		GST040	Gastric Adenocarcinoma	67	135.837
8		SML009	Small Intestine Adenocarcinoma	61	127.022
9		CLN045	Colonic Benign Neoplasm	62	122.972
10		LNG032	Lung Cancer	98	111.826
11		CLN015	Colon Adenocarcinoma	55	107.076

Below the table, there is a section for "Adenocarcinoma colon" with a header, aliases & classifications, summaries, and publications. The header is "colon adenocarcinoma". The aliases & classifications section lists "adenocarcinoma of colon", "colon adenocarcinoma", "adenocarcinoma of the colon", and "colonic adenocarcinoma". The summaries section states: "A colon carcinoma that derives from epithelial cells of glandular origin." The publications section includes a link to a paper: "ARL9 is upregulated and serves as a biomarker for a poor prognosis in colon adenocarcinoma. (PMID: 36823542)".

Anexo B. Búsqueda en MalaCards para obtener genes elite con la entrada Adenocarcinoma colon.

Location	Phenotype	Phenotype MIM number	Inheritance	Phenotype mapping key	Gene/Locus	Gene/Locus MIM number
1p36.13	[Colorectal cancer, susceptibility to]	114500	AD, SMu	3	PLA2G2A	172411
1p13.2	Colorectal cancer, somatic	114500		3	NRAS	164790
2q13	Colorectal cancer with chromosomal instability, somatic	114500		3	BUB1	602452
3p22.1	Colorectal cancer, somatic	114500		3	CTNNB1	116896
5q36.32	Colorectal cancer, somatic	114500		3	PIK3CA	171834
4p16.3	Colorectal cancer, somatic	114500		3	FGFR3	134934
4q11.3	[Colorectal cancer, susceptibility to]	114500	AD, SMu	3	TLR2	603028
5q22.2	Colorectal cancer, somatic	114500		3	APC	611731
5q22.2	Colorectal cancer, somatic	114500		3	MCC	159350
7q11.23	Colorectal cancer, somatic	114500		3	PITPN2	600079
7q34	Colorectal cancer, somatic	114500		3	BRCA1	164757
8p22	Colorectal cancer, somatic	114500		3	DLC1	604258
8p22	Colorectal cancer, somatic	114500		3	PDGFRL	604584
8q21.1	Colorectal cancer, somatic	114500		3	RAD50	604289
11q11.2	Colorectal cancer, somatic	114500		3	PITP1	600925
11q13.3	[Colorectal cancer, susceptibility to]	114500	AD, SMu	3	CCND1	168461
14q24.3	Colorectal cancer, somatic	114500		3	MLF3	604395
14q32.33	Colorectal cancer, somatic	114500		3	AKT1	164730
15q13.1	Colorectal cancer, somatic	114500		3	BUB3B	602860
17p13.1	[Colorectal cancer]	114500	AD, SMu	3	TP53	191170
17p11.2	Colorectal cancer, somatic	114500		3	FLCN	607273
17q24.1	Colorectal cancer, somatic	114500		3	AXIN2	604025
18q11.2	Colorectal cancer, somatic	114500		3	DCC	120470
19q13.33	Colorectal cancer, somatic	114500		3	BAX	600040
20q11.23	Colorectal cancer, advanced, somatic	114500		3	SRC	190090
20q13.2	[Colorectal cancer, susceptibility to]	114500	AD, SMu	3	AURKA	603872
22q12.1	[Colorectal cancer, susceptibility to]	114500	AD, SMu	3	CHEK2	604373
22q13.2	Colorectal cancer, somatic	114500		3	EP300	602700

### Anexo C. Análisis de los genes asociados con la enfermedad del CC.

Harmonizome 3.0

SEARCH DOWNLOAD VISUALIZE PREDICT API MOBILE ABOUT

### Colon Adenocarcinoma Gene Set

Dataset: DISEASES Text-mining Gene-Disease Association Evidence Scores

Category: Disease or phenotype associations

Type: Disease

Description: A colon carcinoma that derives from epithelial cells of glandular origin. (Human Disease Ontology, DOID\_234)

Similar Terms:

Downloads & Tools:  

### Genes

316 genes co-occurring with the disease colon adenocarcinoma in abstracts of biomedical publications from the DISEASES Text-mining Gene-Disease Association Evidence Scores dataset.

Show  entries Filter

Symbol	Name	Standardized Value
KRT20	keratin 20, type I	1.31799
KRT7	keratin 7, type II	1.21738
CDX2	caudal type homeobox 2	1.20223
TYMS	thymidylate synthetase	1.12698
HRAS	Harvey rat sarcoma viral oncogene homolog	1.05219
TP53	tumor protein p53	1.02187
IL2	interleukin 2	1.01807

### Anexo D. Genes de élite obtenidos en la base de datos Harmonizome

**Anexo E.** Genes relacionados a la aparición de cáncer de colon obtenidos de las bases de datos con relación gen-enfermedad.

#	Genes	Alteración Bioquímica	Proteína	DATABASE			
				MalaCards	OMIM	Harmonizome	KEGG
1	APC	Regulación transcripcional-supresor de tumores	APC Regulator of WNT Signaling Pathway	420.39	si	1.66375	si
2	MSH6	Reparación del ADN	MutS Homolog 6	406.48	si	1.65655	si
3	RAD54L	Reparación del ADN	RAD54 Like	402.81	no	no	si
4	SNHG6	Pertenece a lncRNA	Small Nucleolar RNA Host Gene 6	153.98	no	no	no
5	LINC02582	Pertenece a lncRNA	Long Intergenic Non-Protein Coding RNA 2582	150	no	no	no
6	CTNNB1	Actúa como coactivador de los factores de transcripción.	Catenin Beta 1	26.94	si	1.11982	si
7	KRAS	Activación anormal de la vía MAPK	KRAS Proto-Oncogene, GTPase	26.3	si	1.00963	si
8	BAX	Silenciamiento de la expresión en la transcripción.	BCL2 Associated X, Apoptosis Regulator	23.53	si	0.698313	si
9	TP53	Regulación transcripcional-supresor de tumores	Tumor Protein P53	23.15	si	1.02187	si
10	SMAD4	Regulación transcripcional	SMAD Family Member 4	23.01	si	0.95996	si
11	PIK3CA	Activación anormal de la vía MAPK	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha	20.77	si	1.27083	si
12	ABCB1	Regulación transcripcional	ATP Binding Cassette Subfamily B Member 1	20.69	no	0.833968	no
13	TYMS	Factor Transcripcional	Thymidylate Synthetase	20.34	no	1.12698	no

14	CDH1	Regulación transcripcional	Cadherin 1 Mothers Against Decapentaplegic Homolog	20.29	no	0.785165	si
15	SMAD2	Regulación transcripcional	2	20.27	no	0.596012	si
16	CEACAM5	Regulador de apoptosis celular	CEA Cell Adhesion Molecule 5	20.22	no	no	no
17	PTGS2	Activación anormal de la vía MAPK	Prostaglandin-Endoperoxide Synthase 2	20.20	si	0.713198	si
18	CASP3	Alteración en la vía MAPK	Caspase 3 Acyl-CoA Synthetase Long Chain Family	20.15	si	0.828639	si
19	ACSL4	Regulador de apoptosis celular	Member 4	19.70	no	no	no
20	ODC1	Regulación transcripcional	Ornithine Decarboxylase 1 B-Raf Proto-Oncogene, Serine/Threonine	19.42	no	0.749644	si
21	BRAF	Activación anormal de la vía MAPK	Kinase	19.38	si	0.858543	si
22	KRT20	Alteración en la transcripción	Keratin 20	19.28	no	1.31799	no
23	MTHFD1L	Regulación del metabolismo de folato y pterinas	Methylenetetrahydrofolate Dehydrogenase (NADP+ Dependent) 1 Like	18.90	no	no	no
24	CASP8	Inducir la apoptosis celular	Caspase 8	18.65	si	0.438282	si
25	FUT4	Síntesis de antígenos en la superficie celular	Fucosyltransferase 4	18.62	no	0.318635	no
26	BCL2	Regulador de apoptosis celular	BCL2 Apoptosis Regulator	18.61	si	no	no
27	ST6GAL1	Alteración en la transcripción	ST6 Beta-Galactoside Alpha-2,6-Sialyltransferase 1	18.44	si	0.560616	si
28	ANXA5	Regulación de la actividad de CFTR	Annexin A5	18.29	no	0.437906	si

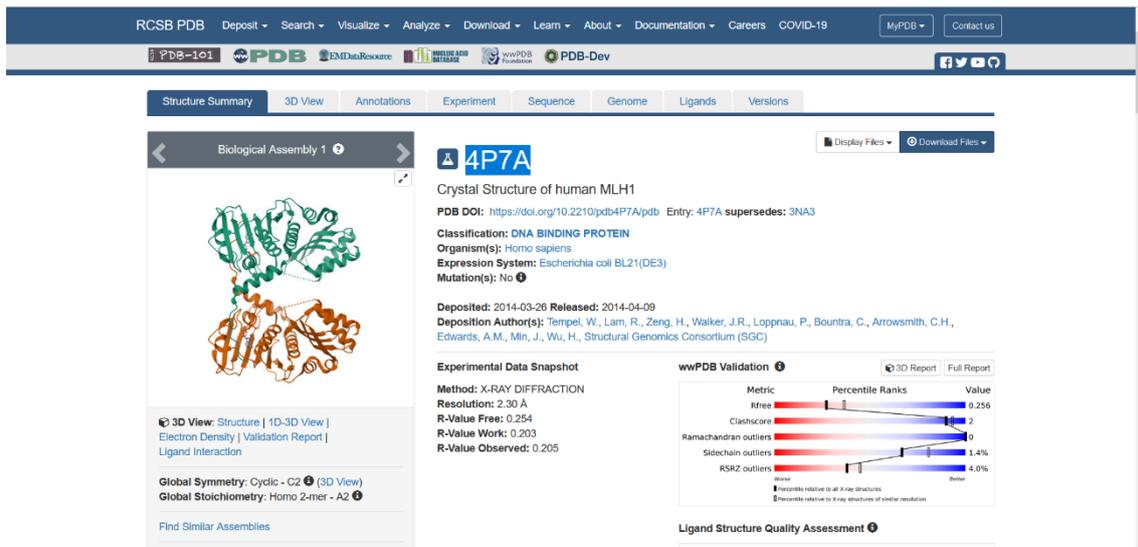
29	GTDC1	Regulación de la actividad de glicosiltransferasa	Glycosyltransferase Like Domain Containing 1	18.27	no	0.178126	si
30	NQO1	Regulación en la protección celular contra el estrés oxidativo.	NAD(P)H Quinone Dehydrogenase 1 ATP Binding Cassette Subfamily C Member 1	18.03	no	0.299452	no
31	ABCC1	Promueven la proliferación celular	Calbindin 2	17.90	no	no	no
32	CALB2	Promueven la proliferación celular	Cyclin Dependent Kinase Inhibitor 1A	17.85	no	0.490072	si
33	CDKN1A	Regulación transcripcional	BCL2 Like 1	17.71	si	1.33152	si
34	BCL2L1	Regulador de apoptosis celular	6-Phosphofructo-2-Kinase/Fructose-2,6-Biphosphatase 3	17.65	si	0.384079	si
35	PFKFB3	Regulación del ciclo celular	TNF Superfamily Member 10	17.65	si	0.474877	si
36	TNFSF10	Regulación de glucocorticoides mediada por MIF	Jun Proto-Oncogene, AP-1 Transcription Factor Subunit	17.25	no	0.354868	no
37	JUN	Regulación como factor transcripcional	Spectrin Repeat Containing Nuclear Envelope Protein 1	17.07	si	0.24996	si
38	SYNE1	Activación anormal de la vía MAPK	Fas Ligand	16.98	no	0.596012	si
39	FASLG	Regulador de apoptosis celular	Protein Tyrosine Kinase 2	16.93	si	0.289744	si
40	PTK2	Regulación del ciclo celular	Nuclear Receptor Subfamily 1 Group I Member 2	16.80	si	no	si
41	NR1I2	Regulación transcripcional	Mitogen-Activated Protein Kinase 8	16.79	no	0.491597	no
42	MAPK8	Regulación transcripcional	Epidermal Growth Factor	16.61	si	0.265849	si
43	EGF	Regulación transcripcional	Poly(ADP-Ribose) Polymerase 1	16.55	si	no	si
44	PARP1	Alteración en la reparación del ADN	Dipeptidyl Peptidase 4	16.51	si	0.163111	si
45	DPP4	Regulación del polipéptido insulínico dependiente de glucosa		16.44	no	no	no

46	ITGA6	Promueven la proliferación celular	Integrin Subunit Alpha 6	16.43	no	0.391141	no
47	ESR2	Regulación transcripcional	Estrogen Receptor 2	16.27	no	0.273308	no
48	ANXA1	Promueven la proliferación celular	Annexin A1	16.24	no	no	si
49	EIF4E	Regulación transcripcional	Eukaryotic Translation Initiation Factor 4E	16.21	no	no	si
50	FAS	Regulador de apoptosis celular	Fas Cell Surface Death Receptor	16.10	no	0.172291	si
51	CFLAR	Regulador de apoptosis celular	CASP8 And FADD Like Apoptosis Regulator	15.96	no	no	si
52	SI	Regulador de la sacarosa	Sucrase-Isomaltase	15.90	no	0.352398	si
53	E2F1	Regulación transcripcional	E2F Transcription Factor 1	15.84	no	no	si
54	CCNA2	Regulación del ciclo celular	Cyclin A2	15.82	no	0.323002	no
55	BID	Regulación transcripcional	BH3 Interacting Domain Death Agonist	15.63	no	no	no
56	FURIN	Regula el crecimiento celular	Furin, Paired Basic Amino Acid Cleaving Enzyme	15.56	si	0.466172	si
57	CYP3A4	Regulación transcripcional	Cytochrome P450 Family 3 Subfamily A Member 4	15.54	no	0.303781	si
58	GUSB	Regulación en la degradación de glucosaminoglucanos	Glucuronidase Beta	15.54	si	0.309566	si
59	ABCC3	Regulación transcripcional	ATP Binding Cassette Subfamily C Member 3	15.47	no	0.43227	si
60	MAPK14	Activación anormal de la vía MAPK	Mitogen-Activated Protein Kinase 14	15.45	si	no	si
61	LGALS3	Regulación transcripcional	Galectin 3	15.20	si	1.11099	si
62	CTNND1	Regulación transcripcional	Catenin Delta 1	15.08	si	0.273308	si
63	ANPEP	Regulación del metabolismo del glutación	Alanyl Aminopeptidase, Membrane	14.72	si	0.208245	si
64	TNFRSF1A	Regulador de apoptosis celular	TNF Receptor Superfamily Member 1A	14.67	si	0.170544	si

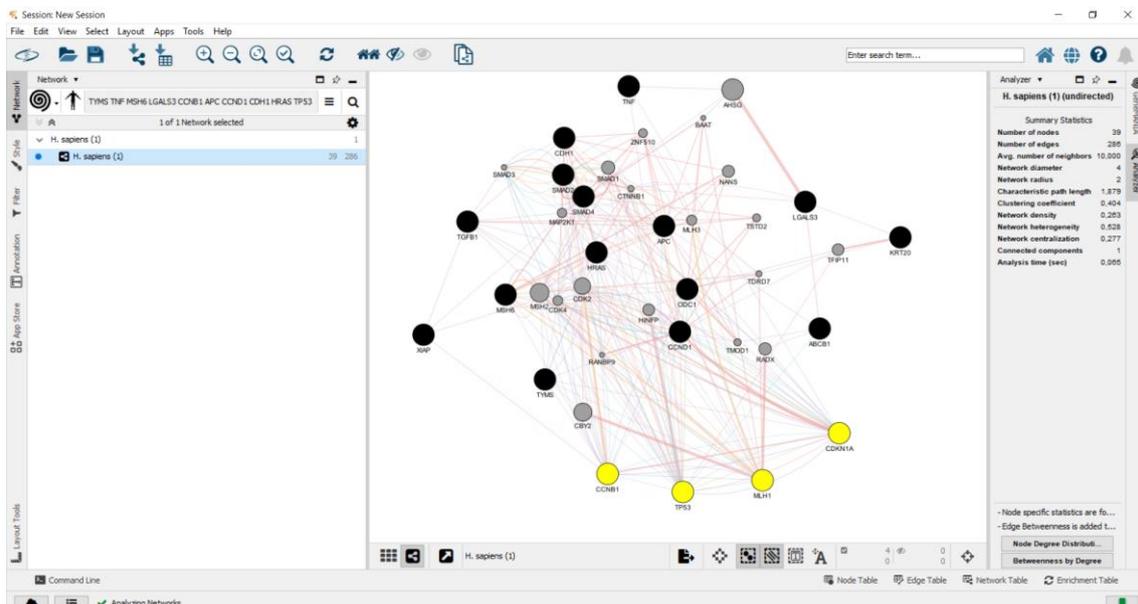
65	TGFA	Promueven la proliferación celular	Transforming Growth Factor Alpha	14.66	no	no	no
66	HPSE	Promueven la proliferación celular	Heparanase	14.58	no	no	si
67	S100A4	Regulación transcripcional	S100 Calcium Binding Protein A4	14.52	no	no	si
68	CAV1	Regulación transcripcional	Caveolin 1	14.49	si	0.489691	si
69	GRP	Síntesis, secreción e inactivación del péptido 1 similar al glucagón (GLP-1)	Gastrin Releasing Peptide	14.39	no	0.235616	si
70	VIP	Alteración en la señalización por GPCR	Vasoactive Intestinal Peptide	14.25	no	0.307395	si
71	TCN2	Regulación del metabolismo de folato y pterinas	Transcobalamin 2	14.08	no	0.224884	si
72	MYC	Regulación transcripcional	MYC Proto-Oncogene, BHLH Transcription Factor	14.06	si	0.20658	si
73	FOS	Regulación transcripcional	Fos Proto-Oncogene, AP-1 Transcription Factor Subunit	14.02	si	0.383337	si
74	TKT	Regulación en la expresión y actividad de la transcetolasa	Transketolase	13.99	no	no	no
75	EGFR	Induce la proliferación y migración de células endoteliales vasculares	Epidermal Growth Factor Receptor	13.98	si	0.366301	si
76	LGALS4	Promueven la proliferación celular	Galectin 4	13.76	si	0.165874	si
77	CEACAM7	Alteración postraduccional de proteínas	CEA Cell Adhesion Molecule 7	13.56	si	no	si
78	VEGFA	Induce la proliferación y migración de células endoteliales vasculares	Vascular Endothelial Growth Factor A	13.54	si	0.377773	si
79	MGAM	Regulación de la descomposición oligosacáridos de almidón	Maltase-Glucoamylase	13.47	no	no	no
80	AREG	Inhibición del crecimiento celular	Amphiregulin	13.35	no	no	no

81	PRKCA	Promueven la proliferación celular	Protein Kinase C Alpha	13.17	no	no	si
82	TNF	Regulación transcripcional	Tumor Necrosis Factor	13.10	si	0.565666	si
83	CCND1	Regulación transcripcional	Cyclin D1	12.99	si	0.6608	si
84	CYCS	Regula la apoptosis celular	Cytochrome C, Somatic	12.99	no	no	no
85	CASP9	Activación anormal de la vía MAPK	Caspase 9	12.69	no	0.540442	si
86	CCNB1	Regulación transcripcional	Cyclin B1	12.53	no	1.18582	si
87	INS	Reparación del ADN	Insulin	12.42	no	no	no
88	METTL3	Regulación transcripcional	Methyltransferase 3, N6-Adenosine-Methyltransferase Complex Catalytic Subunit	12.42	no	no	no
89	ESR1	Promueven la proliferación celular	Estrogen Receptor 1	12.38	no	no	si
90	EPHX2	Síntesis de epoxi (EET)	Epoxide Hydrolase 2	12.21	no	no	si
91	TGFB1	Regulación transcripcional	Transforming Growth Factor Beta 1	12.00	si	0.728547	si
92	MLH1	Regulación transcripcional	MutL Homolog 1	11.89	si	2.09524	si
93	FBXW7	Promueven la proliferación celular	F-Box and WD Repeat Domain Containing 7	11.87	no	no	no
94	ARID1A	Regulación transcripcional	AT-Rich Interaction Domain 1A	11.83	no	no	no
95	MAPK1	Activación anormal de la vía MAPK	Mitogen-Activated Protein Kinase 1	11.82	si	0.355235	si
96	PPARG	Activación anormal de la vía MAPK	Peroxisome Proliferator Activated Receptor Gamma	11.82	si	1.13163	si
97	SP1	Regulación del crecimiento celular	Sp1 Transcription Factor	11.76	no	0.373328	si
98	SMAD3	Regulación transcripcional	SMAD Family Member 3	11.74	si	no	si
99	ANXA2	Regulación del crecimiento celular	Annexin A2	11.57	no	no	si
100	XIAP	Participa en la transcripción	X-Linked Inhibitor of Apoptosis	11.57	no	1.02524	si

101	TIMP1	Promueven la proliferación celular	IMP Metallopeptidase Inhibitor 1	11.47	no	no	si
102	HRAS	Regulación transcripcional	HRas Proto-Oncogene, GTPase	11.42	si	1.05219	si
103	SOX9	Regulación transcripcional	SRY-Box Transcription Factor 9	11.38	no	0.301615	si
104	HDAC1	Promueven la proliferación celular	Histone Deacetylase 1	11.31	no	no	no
105	PXN	Regula el citoesqueleto de actina	Paxillin	11.26	no	0.359289	no
106	VCL	Regula la expresión de E-cadherina.	Vinculin	11.23	si	no	si
107	FOXO1	Regulación transcripcional	Forkhead Box O1	11.22	no	no	si
108	TCF7L2	Regulación transcripcional	Transcription Factor 7 Like 2	11.13	si	0.324458	si
109	MKI67	Actúa como un surfactante permitiendo la motilidad cromosómica	Marker Of Proliferation Ki-67	11.01	no	no	no
110	GNAS	Promueven la proliferación celular	GNAS Complex Locus	10.99	no	no	si
111	HNRNPDL	Regulación transcripcional	Heterogeneous Nuclear Ribonucleoprotein D Like	10.97	no	no	no
112	KLK1	Regulación de la insulina	Kallikrein 1	10.97	si	0.280082	si
113	CYP1A1	Promueven la proliferación celular	Cytochrome P450 Family 1 Subfamily A Member 1	10.87	no	0.286877	si
114	FGF19	Regula el crecimiento celular	Fibroblast Growth Factor 19	10.83	no	no	no
115	AHR	Regulación transcripcional	Aryl Hydrocarbon Receptor	10.78	si	0.25207	si
116	NFKB1	Regulación transcripcional	Nuclear Factor Kappa B Subunit 1	10.75	no	no	si
117	MAP2K4	Regulación del ciclo celular	Mitogen-Activated Protein Kinase 4	10.69	si	no	si
118	AMER1	Degradación de CTNNB1	APC Membrane Recruitment Protein 1	10.64	no	no	si
119	PTK2B	Regula la reorganización del citoesqueleto de actina	Protein Tyrosine Kinase 2 Beta	10.61	si	no	si



**Anexo F. Búsqueda y análisis de la cristalografía tridimensional del gen MLH1.**



**Anexo G. Construcción de la red biológica de la alteración bioquímica de la regulación de transcripción.**

**Anexo H.** Genes representativos de la interacción de la red biológica de la regulación de transcripción.

#	Gene name	Number Of Undirected Edges	Partner Of Multi Edged Node Pairs	Score
1	KRT20	2	0	0.8650
2	ST6GAL1	3	0	0.8433
3	TYMS	8	1	0.8379
4	ABCB1	4	0	0.8266
5	XIAP	4	0	0.8108
6	MSH6	16	2	0.8083
7	BAX	11	1	0.7907
8	TNF	3	0	0.7645
9	ODC1	14	1	0.7538
10	CCNB1	21	6	0.7095
11	APC	27	3	0.7048
12	LGALS3	5	1	0.6954
13	CDH1	17	3	0.6873
14	HRAS	21	1	0.6871
15	MLH1	20	6	0.6756
16	SMAD2	26	5	0.6744
17	TGFB1	11	1	0.6506
18	SMAD4	25	5	0.6165
19	TP53	34	8	0.5577
20	CCND1	24	3	0.5560
21	CDKN1A	26	7	0.5530
22	CTNNB1	29	7	0.4962
23	AHSG	4	0	0.0683
24	CDK2	39	12	0.0449
25	MSH2	23	4	0.0415
26	CBY2	2	0	0.0402
27	SMAD1	17	3	0.0393
28	LEF1	13	2	0.0364
29	RADX	2	0	0.0294

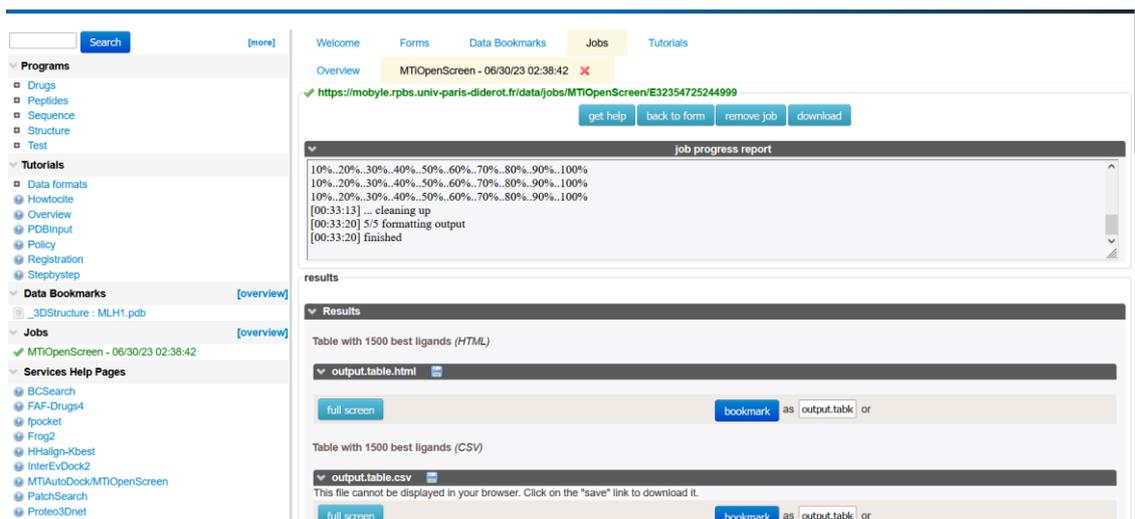
30	HINFP	3	2	0.02940
31	TFIP11	4	0	0.0282
32	MAP2K1	15	1	0.0279
33	NANS	5	0	0.0256
34	CDK4	29	8	0.0255
35	MLH3	13	2	0.0250
36	AMER1	2	0	0.0244
37	GSK3B	20	4	0.0214
38	ZNF510	10	1	0.0203
39	SMAD3	16	4	0.0194
40	TSTD2	5	0	0.0191
41	RANBP9	8	2	0.0183
42	TMOD1	7	1	0.0179

---

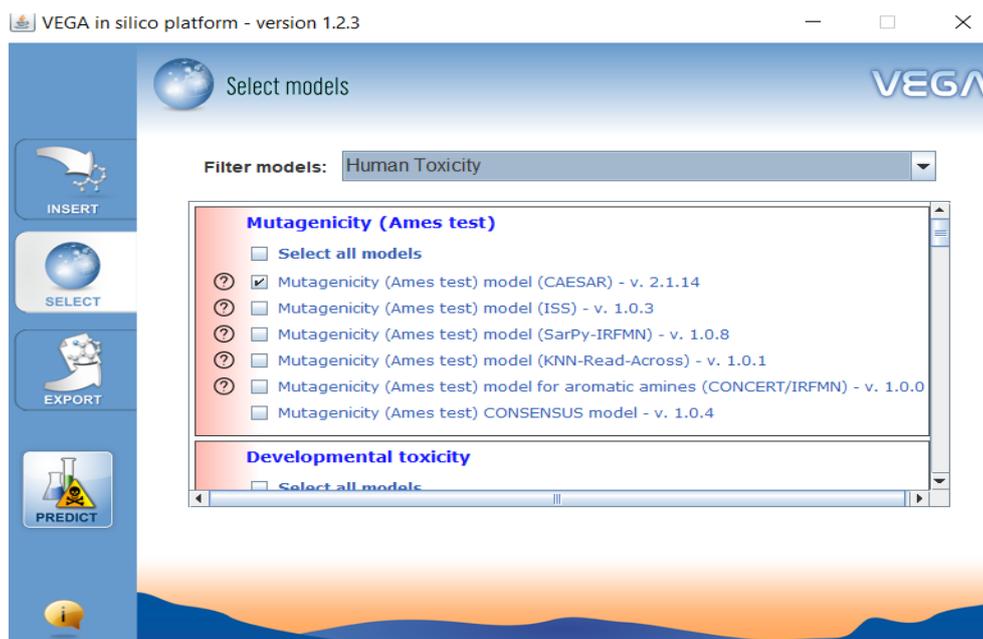
*Anexo I. Genes representativos de la interacción de la red biológica de la vía de señalización de MAPK*

#	Gene name	Number Of Undirected Edges	Partner Of Multi Edged Node Pairs	Score
1	SYNE1	1	0	0.8782
2	PTGS2	3	0	0.8463
3	PPARG	7	1	0.7944
4	CASP3	12	3	0.7645
5	CASP9	11	4	0.7368
6	PIK3CA	30	11	0.6653
7	KRAS	35	11	0.5399
8	BRAF	26	8	0.5040
9	PIK3CD	18	6	0.0825
10	EGFR	26	8	0.0533
11	RGL1	3	0	0.0455
12	MAPK3	25	8	0.0307
13	CASP2	9	3	0.0277
14	SLC35A1	1	0	0.0273
15	RAF1	30	11	0.0247
16	PIK3R1	22	6	0.0242
17	CASP8	15	6	0.0240
18	FLYWCH2	2	0	0.0237
19	RAP2B	5	1	0.0228
20	ARAF	16	5	0.0224
21	MAP2K1	26	7	0.0199
22	SOS1	25	9	0.0187
23	MAPK1	26	8	0.0185
24	LMAN2L	1	0	0.0180
25	IFT57	3	1	0.0167
26	HDAC6	4	0	0.0156
27	SHC1	24	10	0.0152
28	GRB2	26	9	0.0144

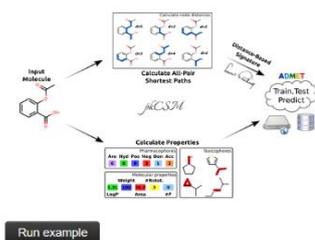




Anexo I. Web de cribado virtual basado en la estructura tridimensional y las coordenadas de la GRI-BOX



Anexo M. QSAR-VEGA in silico evaluación de mutagenicidad, carcinogénesis y hepatotoxicidad para las dianas obtenidas.



Run example

**Disclaimer** ✕  
 No molecule information will be retained on the system after being uploaded by the user.

### Step 1: Please provide a set of molecules (SMILES format)

Description

Upload your SMILES file:  Ningún archivo seleccionado.

OR

Provide a SMILES string:

Example:  
CC(=O)OC1=CC=CC=C1C(=O)O

Files are expected to have headers identifying the columns [File limits](#)

### Step 2: Please choose the prediction mode

Description

Prediction of pharmacokinetic properties

Anexo N. pkCSM para la evaluación de pruebas ADMET para la diana seleccionada