



Instituto Tecnológico de Estudios Superiores de Monterrey

Campus Estado de México

## **Análisis de Biología Computacional**

**BT1013 – Grupo 202**

### **Profesores:**

Leonardo Mauricio Cañete Sifuentes

Aurora Antonio Pérez

### **Proyecto Integrador**

Equipo :

Frida Cano Falcón	A01752953
Gala Flores García	A01745037
Hortencia Alejandra Ramírez Vázquez	A01750150
Gabriela Cortés Olvera	A01751655
Harumi Cristal Manzano Yáñez	A01751568
Diana Paola López Espinosa	A01746530

### **Fecha de entrega:**

Miércoles 28 de abril de 2021

## **Abstract**

El objetivo del estudio en este reporte es emplear el análisis estadístico de datos enfocado en el cáncer de colon en el cual se usó la herramienta computacional llamada “RStudio” y de esta manera obtener un microarreglo de muestras de tejidos sanos y tumorales. Esto con el propósito de detectar los principales factores de riesgo del cáncer de colon (CC) y así tener una mejor comprensión de las causas del mismo para la detección temprana e innovar la técnica de tratamiento.

El análisis se realizó a través de una base de datos de muestras obtenida de “National Center for Biotechnology Information” (GEO dataset con ID “GSE161097”).

A partir de las características de la base de datos, se hizo la construcción de cuatro grupos a analizar “Cáncer de colon en etapa inicial con presencia de mucina” y “Cáncer de colon en etapa inicial sin presencia de mucina”, “Metástasis peritoneal con presencia de mucina” y “Metástasis peritoneal con ausencia de mucina”, obteniendo a partir de un análisis en R se obtuvo una lista de genes para cada tipo de tejido analizado, en el caso de cáncer de colon en etapa inicial la lista está constituida por trece genes y en el caso de metástasis peritoneal por veinte genes.

Finalmente, se realizó la obtención de los datos filtrados por R tomando los 33 genes sobreexpresados del microarreglo seleccionado y de esta manera fueron analizados mediante la base de datos biológica STRING, obteniendo de esta manera una red de asociación de proteínas funcionales. Así mismo, con la misma lista se realizó el análisis con Gene Ontology donde se obtuvo información de suma importancia para la relación de las funciones de tejidos con los genes y a su vez de las células cancerígenas. A partir de la interconexión obtenida por STRING se seleccionaron aquellos con una asociación entre proteínas, de las cuales se ejecutó una investigación para así encontrar su relación directa con el cáncer de colon.

## **Introducción**

De acuerdo con la Organización Mundial de la Salud el cáncer es una de las primeras causas de muerte a nivel mundial; en 2012 se le atribuyeron 8,2 millones de muertes [1]. De hecho existen más de 50 tipos de cáncer, esto debido a sus características que le permiten desarrollarse en cualquier parte del cuerpo, además es una enfermedad para la que aún no existe una cura, solamente tratamiento el cual es aplicado para un porcentaje de los casos. Entre los cinco tipos de cáncer más comunes en América el cáncer de colon se encuentra en el cuarto lugar [2]. En efecto, en los últimos años, el cáncer de colon (CC) ha aumentado su

incidencia; constituyendo un 10% de la tasa de mortalidad de los pacientes diagnosticados con cáncer en los países occidentales [3]. No obstante se ha demostrado que existe una tasa global de supervivencia de 90% en caso de que el CC sea diagnosticado a tiempo, en lo que es considerado “etapa 0”, ya que en esta etapa el tumor no ha alcanzado grandes dimensiones y sólo se requiere de una cirugía mínimamente invasiva para extraerlo [4]. Hasta la fecha los médicos no saben exactamente qué es lo que provoca el CC, se considera multifactorial.

Por ello, en América, una de las estrategias para reducir la incidencia de mortalidad de cáncer de colon consiste en la prevención a través de pruebas de tamizaje: sigmoidoscopia flexible o colonoscopia. A pesar de la efectividad de estos diagnósticos sus procesos resultan invasivos, lo cual provoca en los pacientes falta de iniciativa para tomar los estudios. Debido a esto en los últimos años se han buscado formas innovadoras para diagnosticar el CC, tales como el análisis de la expresión genética basado en el análisis bioinformático.

El objetivo de este reporte es poder realizar un estudio estadístico a través del análisis bioinformático a nivel molecular que permita detectar los principales factores de riesgo del CC.

En relación con los factores de riesgo, se ha descubierto que la tasa de supervivencia de los pacientes con CC disminuye considerablemente debido al desarrollo de metástasis en el peritoneo siendo aproximadamente 16.3 meses en comparación a aquellos con metástasis en el hígado (19.1 meses) y en el pulmón (24.6 meses) [5]. El desarrollo de metástasis en el peritoneo ocurre por la capacidad con la que las células cancerígenas se adhieren a la pared del peritoneo [6]. Se propone que una de las razones por las que ocurre esta adhesión es la presencia de adenocarcinoma mucina cuyo origen está en las células que constituyen el revestimiento interno de las glándulas de secreción externa.

Se busca comprobar la hipótesis a través del uso de biomarcadores moleculares, tomando como set de datos “GSE161097” el cual nos presenta 30 muestras de las cuales 15 son de “cáncer de colon en etapa inicial” y las restantes de “metástasis peritoneal”. A través de la plataforma RStudio [7] se busca desarrollar un análisis bioinformático estadístico el cual ayudará a identificar los principales genes expresados en los tejidos con presencia y ausencia de adenocarcinoma mucina. En seguida se identifican las interconexiones de dichos genes utilizando la plataforma STRING para conocer la forma en la que interactúan las moléculas para después analizar el comportamiento y función biológica de los genes.

Los resultados de este estudio pueden resultar de utilidad para el desarrollo de métodos de diagnóstico y seguimiento no invasivos con los cuáles se busquen identificar los marcadores moleculares, índices del desarrollo del cáncer de colon.

## **Metodología**

### *Selección de Dataset*

El dataset empleado se obtuvo de la base de datos de National Center for Biotechnology Information (NCBI) [8]. Para seleccionar el dataset adecuado se evaluó que las muestras estudiadas fueran realizadas en humanos (“Homo sapiens”) con cáncer de colon, que el perfil genético fuera expresado (“expression profiling by array”), por último que estuviera disponible en la base de datos GEO2R la cual nos permite analizar la secuencia de datos en el lenguaje de programación R. Gracias a la aplicación de estos filtros de búsqueda se obtuvo el dataset “GSE 161097” titulado “Early adaptation of colorectal cancer cells to the peritoneal cavity depends on activation of ‘stemness’ programs and local inflammation”. Donde los grupos de estudio para las 30 muestras son: tipos de tejidos (“primary colon cancer” o “peritoneal metastases”), estatus del paciente estudiado (“Alive” o “Dead”), historia del paciente (“non mucinous” y “mucinous”), edad (los diferentes rangos de edad de los pacientes sometidos a prueba), sexo (“Female” o “Male”), etapa del paciente (“IV”), supervivencia en meses.

En este estudio se busca comprobar la incidencia de la presencia de adenocarcinoma mucina en el desarrollo de metástasis en el peritoneo, por ello se analizan los siguientes grupos de estudio: el historial del paciente (“non mucinous” y “mucinous”) -presencia de adenocarcinoma mucina- y los tipos de tejidos (“primary colon cancer” o “peritoneal metastases”) -etapa inicial de CC y metástasis peritoneal-.

### *Filtrado de genes diferencialmente expresados (DEGs)*

La extracción de la secuencia de datos obtenida por GEO2R se procesó y filtró mediante la aplicación de las librerías ‘GEOquery’ y ‘limma’, paquetes disponibles en la plataforma RStudio. A partir del microarreglo obtenido se filtraron los cuatro grupos a analizar, agrupando los valores de expresión que tienen la característica perteneciente a “cáncer de colon en etapa inicial con adenocarcinoma mucina”, “cáncer de colon en etapa inicial sin

adenocarcinoma mucina”, “metástasis peritoneal con adenocarcinoma mucina”, “metástasis peritoneal sin adenocarcinoma mucina”.

Gracias a la función `cbind` se agruparon los cuatro grupos en un microarreglo. Posteriormente se calculó el promedio de los valores de expresión del microarreglo, lo cual permitió normalizar el conjunto de datos al aplicar la función `'sweep'`. Por cada grupo del microarreglo normalizado se calculó el promedio de los datos y nuevamente se agrupó en un nuevo dataframe.

Para analizar la proporción de expresión de mucina en cada tipo de tejido se asignó a dos arreglos diferentes la comparación de la presencia de mucina con su ausencia en tejidos de etapa inicial de cáncer de colon y en metástasis peritoneal respectivamente; dichas proporciones son asignadas a otro dataframe.

Para que la variación y significancia de los valores sea más evidente, los microarreglos obtenidos de la normalización, promedio y proporción fueron transformados a logaritmo de base dos ( $\log_2$ ).

A partir del microarreglo normalizado y con ayuda de la función creada: `'get_pvalue'` -cuyo funcionamiento está relacionado a la aplicación de un t-test para extraer el p-value- se obtuvieron los p-values del conjunto de datos que se está analizando, seleccionando aquellos genes que se encuentran por debajo del nivel de significancia observado ( $P < 0.05$ ).

Para hacer la selección de los genes diferencialmente expresados se aplican dos métodos para el filtrado: primero seleccionando aquellos con un p-value menor a 0.05 y con un tamaño del efecto mayor a 0.25 y menor a -0.25, obteniendo una lista de genes con un nivel de expresión 2 veces mayor en tejidos con presencia de mucina y 2 veces mayor con ausencia de mucina para cada tejido analizado.

```
# Seleccionar genes con menor p-value y con un tamaño del efecto "grande"
#selección de genes de tamaño del efecto grande
filtered_primary_colon_cancer_MA = primary_colon_cancer_p[primary_colon_cancer_p < 0.05 & microarray_ratios$primary_colon_cancer_ratios > 0.25]
filtered_primary_colon_cancer_ME = primary_colon_cancer_p[primary_colon_cancer_p < 0.05 & microarray_ratios$primary_colon_cancer_ratios < -0.25]
filtered_peritoneal_metastases_MA = primary_colon_cancer_p[peritoneal_metastases_p < 0.05 & microarray_ratios$peritoneal_metastases_ratios > 0.25]
filtered_peritoneal_metastases_ME = primary_colon_cancer_p[peritoneal_metastases_p < 0.05 & microarray_ratios$peritoneal_metastases_ratios < -0.25]
```

**Figura 1.** Selección de genes con p-value menor a 0.05 y con un tamaño del efecto grande.

Posteriormente los resultados obtenidos se ordenaron de menor a mayor en función del p-value.

```
#Ordenar datos
datos_pvalores_primary_colon_MA <- sort(filtred_primary_colon_cancer_MA)
datos_pvalores_primary_colon_ME <- sort(filtred_primary_colon_cancer_ME)

datos_pvalores_peritoneal_metastases_MA <- sort(filtred_peritoneal_metastases_MA)
datos_pvalores_peritoneal_metastases_ME <- sort(filtred_peritoneal_metastases_ME)
```

**Figura 2.** Organización de los genes de cada grupo en función de su p-value (menor a mayor)

De cada una de las listas anteriores se extrajeron los primeros diez genes, obteniendo así dos listas de diez genes con mayor expresión en un tejido mucinoso y con mayor expresión en un tejido con ausencia de mucina, esto para los dos casos analizados: cáncer de colon en etapa inicial y metástasis peritoneal. A estas listas obtenidas se les asignó el ID Gene para lograr una mejor identificación de ellos.

```
#Primary colon
#TOP10 MA

datos_primary_colonTop10_MA = datos_pvalores_primary_colon_MA[1:10]
df_datos_primary_colonTop10_MA<-data.frame(datos_primary_colonTop10_MA)

datos_primary_colonTop10_names_MA=row.names(df_datos_primary_colonTop10_MA)
Top10_genes_primary_MA=probes[datos_primary_colonTop10_names_MA,3] #Top
df_Top10_genes_primary_MA<-data.frame(Top10_genes_primary_MA)

#TOP10 ME

datos_primary_colonTop10_ME = datos_pvalores_primary_colon_ME[1:3]
df_datos_primary_colonTop10_ME<-data.frame(datos_primary_colonTop10_ME)

datos_primary_colonTop10_names_ME=row.names(df_datos_primary_colonTop10_ME)
Top10_genes_primary_ME=probes[datos_primary_colonTop10_names_ME,3] #Top
df_Top10_genes_primary_ME<-data.frame(Top10_genes_primary_ME)

#Peritoneal Metastases
#TOP10 MA

datos_peritoneal_Top10_MA = datos_pvalores_peritoneal_metastases_MA[1:10]
df_datos_peritoneal_Top10_MA<-data.frame(datos_peritoneal_Top10_MA)

datos_peritoneal_Top10_names_MA=row.names(df_datos_peritoneal_Top10_MA)
Top10_genes_peritoneal_MA=probes[datos_peritoneal_Top10_names_MA,3] #Top
df_Top10_genes_peritoneal_MA<-data.frame(Top10_genes_peritoneal_MA)

#TOP10 ME

datos_peritoneal_Top10_ME = datos_pvalores_peritoneal_metastases_ME[1:10]
df_datos_peritoneal_Top10_ME<-data.frame(datos_peritoneal_Top10_ME)

datos_peritoneal_Top10_names_ME=row.names(df_datos_peritoneal_Top10_ME)
Top10_genes_peritoneal_ME=probes[datos_peritoneal_Top10_names_ME,3] #Top
df_Top10_genes_peritoneal_ME<-data.frame(Top10_genes_peritoneal_ME)
```

**Figura 3.** Selección de los primeros diez genes con menor p-value de cada grupo.

### *Construcción de gráficas en RStudio*

El filtrado de datos realizado anteriormente, permitió la construcción de cinco gráficas (dendograma, mapa de calor, gráfica de dispersión, de relación de intensidad y de volcan) para mostrar la relación entre las muestras analizadas: cáncer de colon en etapa inicial con

presencia de tejido mucinoso y sin presencia de mucina, metástasis peritoneal con presencia de mucina y en su ausencia.

#### *Búsqueda del funcionamiento de los genes en plataformas GenBank, STRING.*

Para el enriquecimiento del análisis de los genes diferencialmente expresados se utilizaron herramientas bioinformáticas como Genbank [9] y “Search Tool for the Retrieval of Interacting Genes” (STRING) [10] que permitieron un análisis sistemático y biológico sobre el proceso molecular que realizan estos genes.

#### *Búsqueda de las interconexiones de los DEGs*

La plataforma STRING es un recurso bioinformático que permitió crear una interconexión proteína-proteína, la cual permitió conocer la forma de interacción que hay entre los genes diferencialmente expresados y genes relacionados a la función que realizan los DEGs. También se implementó el uso de la plataforma “Gene Ontology” [11] para complementar el análisis de la interacción que tienen los DEGs.

#### *Selección de genes principales*

A partir de las gráficas de interconexión realizadas en STRING se extrajeron aquellos genes relacionados, con mayor susceptibilidad a mutar y, por fines de investigación, se analizaron aquellos cuyo expediente estuviera disponible en la plataforma COSMIC.

#### *Análisis de mutaciones de los genes principales*

Con ayuda de las plataformas Catalogue Of Somatic Mutations In Cancer (COSMIC) [12] y The Genome Aggregation Database (genomAD) [13] se enriqueció el análisis de los genes identificados con el fin de explorar el impacto de las mutaciones en el cáncer de colon y hacer una mayor identificación de las características que presentan los genes principales.

## **Resultados**

#### *Filtrado de genes diferencialmente expresados (DEGs)*

Aplicando los filtros anteriormente mencionados en la sección de metodología sobre cada uno de nuestros grupos, se redujo significativamente el número de genes; para el grupo cáncer de colon en etapa inicial que contiene originalmente 780 genes se redujo a 13 con presencia de mucina y 3 genes sin mucina, mientras que en el tejido de metástasis peritoneal

se redujo de 780 genes a 12 con mucina y 23 sin mucina. A partir de la selección de los 10 primeros genes por cada grupo se obtuvieron las siguientes tablas:

**Tabla 1.** Genes diferencialmente expresados para cáncer de colon en etapa inicial

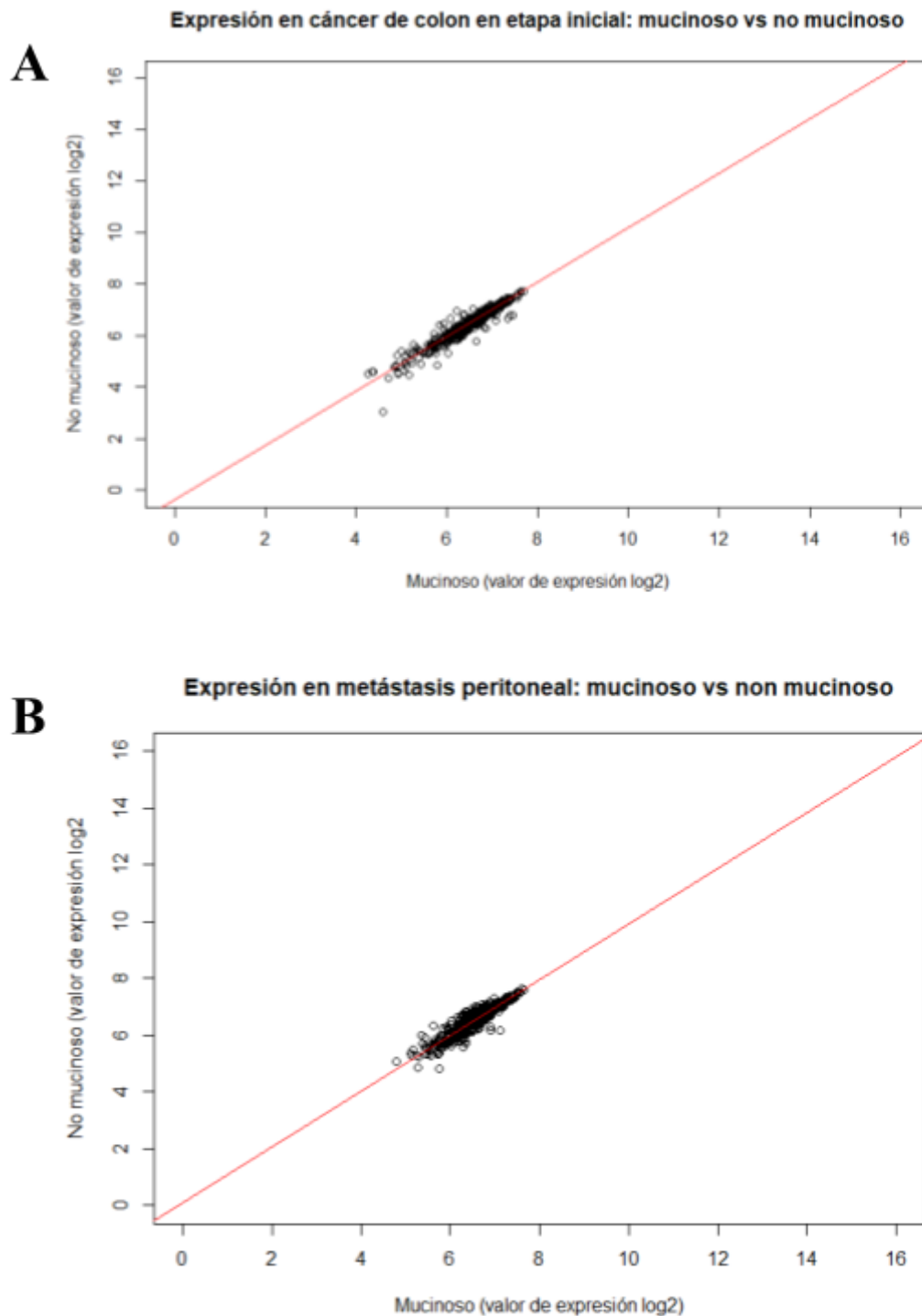
Sobreexpresados en presencia de mucina	Sobreexpresados sin mucina
TCN1	WNT11
SERPINA1	AXIN2
DLL1	EREG
COL4A5	
FZD9	
MUC2	
SPINK4	
REG4	
CD79B	
TWIST1	

**Tabla 2.** Genes diferencialmente expresados para metástasis peritoneal

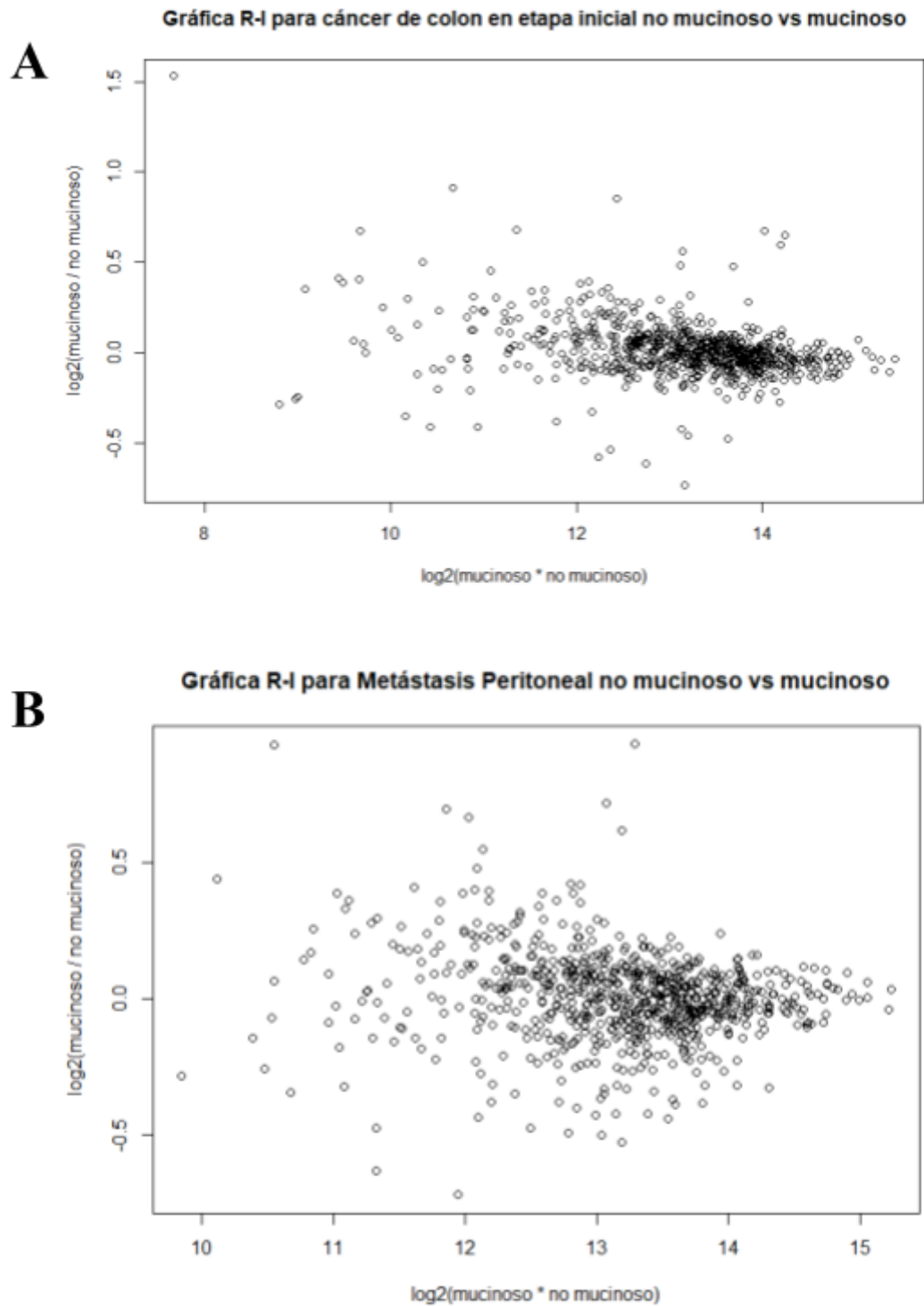
Sobreexpresados en presencia de mucina	Sobreexpresados sin mucina
MUC2	AXIN2
SPINK4	MET
CD79B	F2RL1
CD244	LAMC2
CD79A	CDH1
FAM30A	ALDOC
ZAP70	GNG4
GZMM	AQP8
CD28	CDC25C
TCL1A	HDAC11



## Construcción de gráficas

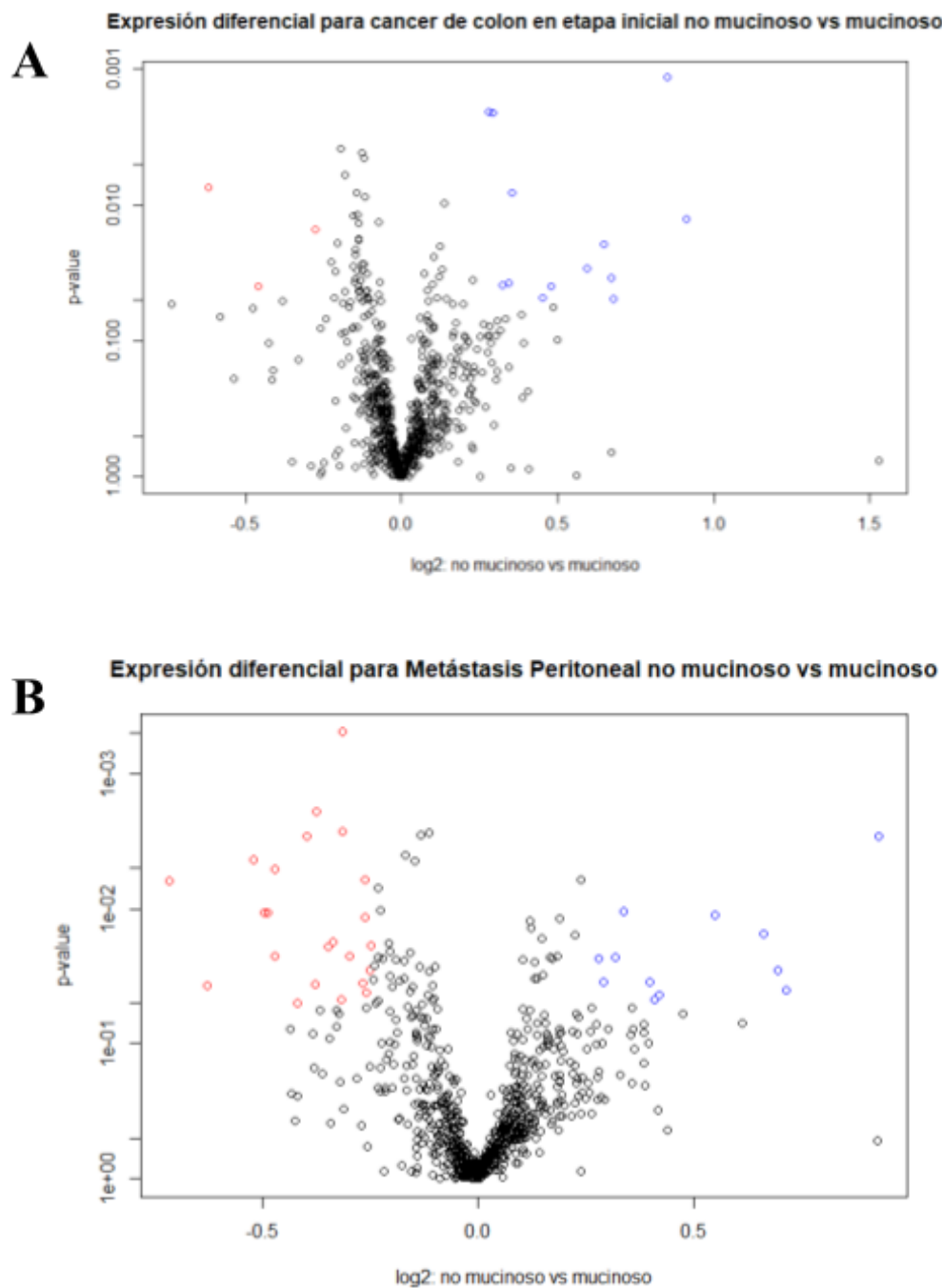


**Figura 4.** Gráficas de expresión de las muestras en tejidos con cáncer de colon en etapa inicial (A) y en tejidos con metástasis peritoneal (B). Estas gráficas hacen uso de los valores de expresión normalizados de las muestras en presencia de mucina y en ausencia de mucina, la línea roja representa el promedio de los valores, en el eje x se observan los valores de expresión para genes con mayor expresión en presencia de mucina, mientras que en el eje y los genes con mayor expresión en ausencia de mucina.

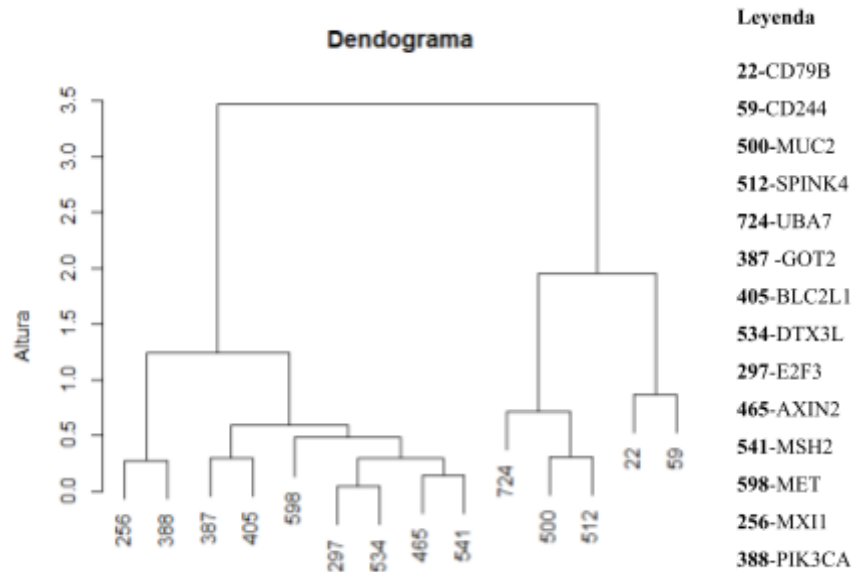


**Figura 5.** Gráficas de relación de intensidad para cáncer de colon en etapa inicial (A) y en tejidos con metástasis peritoneal (B). Estas gráficas hacen una relación entre las muestras en presencia de mucina y en ausencia de mucina, permitiendo observar en la mayor agrupación de puntos aquellos genes que se comportan de manera similar para ambas muestras, mientras que los puntos que se encuentran más dispersos corresponden a los genes con mayor diferencia de expresión.

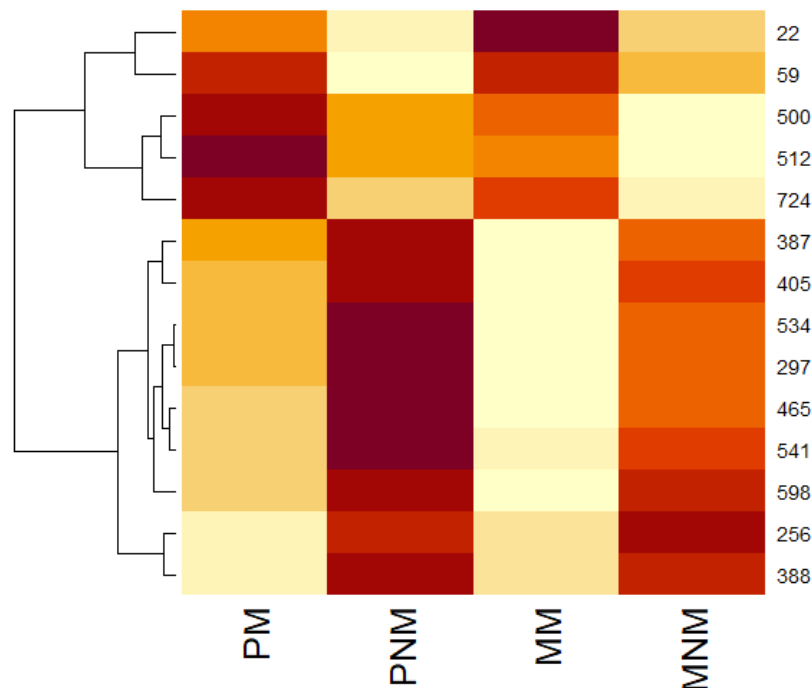
Para destacar la distribución de cada gen en función de su tamaño ( $\log_2$ ) y p-value en cada uno de los tipos de tejido estudiados se utiliza la gráfica de volcano (Figura 4).



**Figura 6.** Gráficas de volcano de los genes sobreexpresados en tejidos con cáncer de colon en etapa inicial (A) y en tejidos con metástasis peritoneal (B). Los puntos rojos representan la sobreexpresión de los genes en su respectivo tejido sin tener este rastros de mucina y los puntos azules representan la sobreexpresión de los genes en su respectivo tejido con presencia de mucina.



**Figura 7.** Dendrograma de los genes diferenciales. Esta gráfica nos permite observar el nivel de similitud que hay entre cada uno de los genes, este nivel de similitud se puede observar en el eje vertical, mientras que en el eje horizontal se muestran los genes relacionados, a menor altura hay mayor similitud entre los genes, lo cual se puede observar en la agrupación que tienen.



**Figura 8.** Mapa de calor donde se puede observar la expresión de los genes para los cuatro grupos analizados: cáncer de colon en etapa inicial con presencia de mucina (PM), cáncer de colon en etapa inicial en ausencia de mucina (PNM), metástasis peritoneal con presencia de mucina (MM) y metástasis peritoneal en ausencia de mucina (MNM), este mapa se construyó con la ayuda del dendrograma generado en la figura 7 lo cual permitió disponer de los genes diferencialmente expresados en una jerarquía según la distancia y similitud que hay entre ellos.

Para poder identificar con mayor facilidad los genes sobreexpresados en cada grupo en el mapa de calor, se realizó una tabla traduciendo el número relacionado al gen por su ID-gene.

**Tabla 3.** Muestra la relación entre color obtenido en el mapa de calor con su gen correspondiente.

Cáncer de colon en etapa inicial con presencia de mucina	Cáncer de colon en etapa inicial en ausencia de mucina	Metástasis peritoneal con presencia de mucina	Metástasis peritoneal en ausencia de mucina
CD79B	CD79B	CD79B	CD79B
CD244	CD244	CD244	CD244
MUC2	MUC2	MUC2	MUC2
SPINK4	SPINK4	SPINK4	SPINK4
UBA7	UBA7	UBA7	UBA7
GOT2	GOT2	GOT2	GOT2
BLC2L1	BLC2L1	BLC2L1	BLC2L1
DTX3L	DTX3L	DTX3L	DTX3L
E2F3	E2F3	E2F3	E2F3
AXIN2	AXIN2	AXIN2	AXIN2
MSH2	MSH2	MSH2	MSH2
MET	MET	MET	MET
MXI1	MXI1	MXI1	MXI1
PIK3CA	PIK3CA	PIK3CA	PIK3CA

*Búsqueda del funcionamiento de los genes en plataformas GenBank, STRING.*

A partir de las tablas 1 y 2, se presenta una comparativa del funcionamiento y características biológicas presentes en los genes diferencialmente expresados para tejidos de cáncer de colon en etapa inicial y metástasis peritoneal.

**Tabla 4.** Comparativa de genes relevantes diferencialmente expresados para cáncer de colon en etapa inicial.

Genes	Localización	Función	Expediente de la secuencia	Número de exones	Serie de nucleótidos repetidos	Tejido en el que está más expresado
<b>TCN1</b> (Transcobalam in 1)	11q12.1  <b>Inicio/Fin:</b> NC_000011.10 (59852808.. 59866487, complement)  <b>Longitud del gen:</b> 13,679.	Codifica un miembro de las proteínas de unión a vitamina B12. También son conocidas como aglutinantes R y se encuentran en tejidos y secreciones. Facilita el transporte de cobalamina en el interior de las células.	<a href="https://www.ncbi.nlm.nih.gov/nucleotide/NC_000011.10?report=genbank&amp;from=59852808&amp;to=59866487&amp;strand=true">https://www.ncbi.nlm.nih.gov/nucleotide/NC_000011.10?report=genbank&amp;from=59852808&amp;to=59866487&amp;strand=true</a>	9	WindowMasker: 962 nt  RepeatMasker: 5,381 nt	Glándula salival
<b>SERPINA1</b> (Serp family A member 1)	14q32.13  <b>Inicio/Fin:</b> NC_000014.9 (94376747.. 94390654, complement)  <b>Longitud del gen:</b> 13,907.	La proteína codificada por este gen es de la familia de serpinas. Se produce en el hígado, médula ósea, células linfocíticas. Se le asocia con enfermedad crónica pulmonar y enfermedad hepática crónica.	<a href="https://www.ncbi.nlm.nih.gov/nucleotide/NG_008290.1?from=7091&amp;to=18946&amp;report=genbank">https://www.ncbi.nlm.nih.gov/nucleotide/NG_008290.1?from=7091&amp;to=18946&amp;report=genbank</a>	7	WindowMasker: 308 nt  RepeatMasker: 296 nt	Hígado
<b>DLL1</b> (delta like canonical Notch ligand 1)	6q27  <b>Inicio/Fin</b> NC_000006.12 (170282200..	Se encarga de la mediación de las decisiones del destino celular durante la hematopoyesis. Puede desempeñar una parte importante en la comunicación de célula a célula.  El gen DLL1 se encarga de la mediación de las decisiones del destino celular y su desarrollo esquelético durante la	<a href="https://www.ncbi.nlm.nih.gov/nucleotide/NG_027940.1?from=4532&amp;to=13404&amp;report=genbank">https://www.ncbi.nlm.nih.gov/nucleotide/NG_027940.1?from=4532&amp;to=13404&amp;report=genbank</a>	11	WindowMasker 3,652 nt  RepeatMasker	Bazo

	170291078, complement) <b>Longitud del gen:</b> 8,878	hematopoyesis (la producción de las células sanguíneas) por lo que su mutación causa muchas represalias al ser una parte importante de la comunicación entre notch. Se alterarían estas señales creando una anomalía en la regulación, especificación, la proliferación y la morfogénesis de las células. [15]			1,517 nt	
<b>COL4A5</b> (collagen type IV alpha 5 chain)	Xq22.3  <b>Inicio/Fin</b> NC_000023.11 (108439838.. 108697545)  <b>Longitud del gen:</b> 257,707	Este gen codifica una de las seis subunidades del colágeno tipo IV, el principal componente estructural de las membranas basales. Las mutaciones en este gen están asociadas con el síndrome de Alport ligado al cromosoma X, también conocido como nefritis hereditaria. Al igual que los otros miembros de la familia de genes del colágeno tipo IV, este gen está organizado en una conformación cabeza a cabeza con otro gen del colágeno tipo IV, de modo que cada par de genes comparte un promotor común. Alternativamente, se han identificado variantes de transcripción empalmadas para este gen.  El gen COL4A5 crea redes de proteína complejas que son parte importante de las membranas basales que separan y sostienen las células de tejidos y este caso en específico son importantes para el riñón, oído interno y el ojo. Al haber mutaciones de este gen se asocia el síndrome de Alport ligado al cromosoma X (hereditario) por lo que puede generar insuficiencia renal. [16]	<a href="https://www.ncbi.nlm.nih.gov/nuccore/NG_011977.2?from=4915&amp;to=262622&amp;report=genbank">https://www.ncbi.nlm.nih.gov/nuccore/NG_011977.2?from=4915&amp;to=262622&amp;report=genbank</a>	58	WindowMasker 2,619 nt  RepeatMasker 6,146 nt	Endometrio
<b>FZD9</b> (Frizzled class receptor 9)	7q11.23  <b>Inicio/Fin:</b> NC_000007.14 (73433778.. 73436120)  <b>Longitud del gen:</b> 2342	La función de la familia de genes “frizzled” es codificar las proteínas de dominio “7-transmembrane” quienes son los receptores de WNT2. Desempeñan un papel en el ensamblaje de la unión neuromuscular (NMJ) al regular negativamente la agrupación de receptores de acetilcolina (AChR) a través de la vía de señalización canónica de beta-catenina (por similitud) .	<a href="https://www.ncbi.nlm.nih.gov/nuccore/NG_012028.1?from=5000&amp;to=7342&amp;report=genbank">https://www.ncbi.nlm.nih.gov/nuccore/NG_012028.1?from=5000&amp;to=7342&amp;report=genbank</a>	1	WindowMasker 648 nt  RepeatMasker 307 nt	Cerebro

<b>WNT11</b> (Wnt family member 11)	11q13.5 <b>Inicio/Fin:</b> NC_000011.10 (76186325.. 76210842, complement) <b>Longitud del gen:</b> 24,517	Este gen forma parte de la familia WNT y está encargada de codificar las proteínas que se encuentran en procesos de desarrollo y oncogénesis. También codifica la proteína que muestra el 97%, 85% y 63% de identidad de aminoácidos con la de pollo, ratón y xenopus. Puede tener un papel en el desarrollo del esqueleto, riñón y pulmón.	<a href="https://www.ncbi.nlm.nih.gov/nucleotide/NC_000011.10?report=genbank&amp;from=76186325&amp;to=76210842&amp;strand=true">https://www.ncbi.nlm.nih.gov/nucleotide/NC_000011.10?report=genbank&amp;from=76186325&amp;to=76210842&amp;strand=true</a>	8	WindowMasker: 660 nt  RepeatMasker: 863 nt	Grasa
<b>AXIN2</b> (Axis Inhibition Protein 2)	17q24.1 <b>Inicio/Fin:</b> NC_000017.11 (65528563.. 65561648, complement) <b>Longitud del gen:</b> 33,085	Juega un papel importante en la regulación de la estabilidad de la beta-catenina en la vía de señalización de Wnt, como sus homólogos de roedores, conductina de ratón / axila de rata. Muestra una pérdida frecuente de heterocigosidad en el cáncer de mama, en el neuroblastoma hay otros tumores. Las mutaciones en este gen se han asociado con el cáncer colorrectal con reparación defectuosa de desajustes.	<a href="https://www.ncbi.nlm.nih.gov/nucleotide/NC_000017.11?report=genbank&amp;from=65528563&amp;to=65561648&amp;strand=true">https://www.ncbi.nlm.nih.gov/nucleotide/NC_000017.11?report=genbank&amp;from=65528563&amp;to=65561648&amp;strand=true</a>	15	WindowMasker: 627 nt  RepeatMasker: 467 nt	Endometrio
<b>EREG</b> (epiregulin)	4q13.3 <b>Inicio/Fin</b> NC_000004.12 (74365145.. 74388749) <b>Longitud del gen:</b> 23,604	Codifica una hormona peptídica secretada, ligada a la inflamación, cicatrización de heridas y proliferación celular. La proteína codificada puede promover la progresión de cánceres de varios tejidos humanos.	<a href="https://www.ncbi.nlm.nih.gov/nucleotide/NC_000004.12?report=genbank&amp;from=74365145&amp;to=74388749">https://www.ncbi.nlm.nih.gov/nucleotide/NC_000004.12?report=genbank&amp;from=74365145&amp;to=74388749</a>	5	WindowMasker: 6,050 nt  RepeatMasker: 6,440 nt	Médula ósea



**Tabla 5.** Comparativa de genes relevantes diferencialmente expresados para metástasis peritoneal.

Genes	Localización	Función	Expediente de la secuencia	Número de exones	Serie de nucleótidos repetidos	Tejido en el que está más expresado
<b>MUC2</b> (mucin2, oligomeric mucus/gel-forming)	11p15.5  <b>Inicio/Fin:</b> NC_000011.10 (1074875..1110508)  <b>Longitud del gen:</b> 35633nt	Este gen codifica un miembro de la familia de proteínas de mucina. Las mucinas son glicoproteínas de alto peso molecular producidas por muchos tejidos epiteliales. La proteína codificada por este gen se secreta y forma una barrera mucosa insoluble que protege la luz intestinal. La proteína se polimeriza en un gel del cual el 80% del peso está compuesto por cadenas laterales de oligosacáridos. La proteína presenta un dominio central que contiene repeticiones en tándem ricas en treonina y prolina que varía entre 50 y 115 copias en diferentes individuos. Se ha observado una regulación a la baja de este gen en pacientes con enfermedad de Crohn y colitis ulcerosa.	<a href="https://www.ncbi.nlm.nih.gov/nucleotide/NG_051929.1?from=5001&amp;to=42531&amp;report=genbank">https://www.ncbi.nlm.nih.gov/nucleotide/NG_051929.1?from=5001&amp;to=42531&amp;report=genbank</a>	52	WindowMasker: 779 nt  RepeatMasker: 802 nt.	Colon
<b>SPINK4</b> (Serine peptidase inhibitor Kazal type 4)	9p13.3  <b>Inicio/Fin:</b> NC_000009.12 (33240167..33248567)  <b>Longitud del gen:</b> 8,400nt	Este gen es un gen codificador de proteínas y un inhibidor de la serina proteasa que funciona como reguladores centrales. Se expresa abundantemente en células caliciformes humanas pero también se forma, almacena y secreta a partir de monocitos por ende funciona como un péptido gastrointestinal.  El gen SPINK4 está asociado a la escasa supervivencia en cáncer colorrectal por su baja regulación. Se encontró que mientras menor expresión se tenga de SPINK4, hay un mayor nivel del estado de diseminación del cáncer tumoral a otras partes del cuerpo (TNM) aunque todavía se encuentra en investigación su significancia clínica. [17]	<a href="https://www.ncbi.nlm.nih.gov/nucleotide/NC_000009.12?feature=any">https://www.ncbi.nlm.nih.gov/nucleotide/NC_000009.12?feature=any</a>	4	WindowMasker: 1,038 nt  RepeatMasker: 711 nt	Colon
<b>CD79B</b> (CD79b molecule)	17q23.3  <b>Inicio/Fin:</b>	Este gen codifica la proteína Ig-beta del componente antígeno de células B.	<a href="https://www.ncbi.nlm.nih.gov/nucleotide/NG_007368.1?from=5001&amp;to=">https://www.ncbi.nlm.nih.gov/nucleotide/NG_007368.1?from=5001&amp;to=</a>	6	WindowMasker: 123 nt	Ganglio linfático

	NC_000017.11 (63928740.. 63932331, complement) <b>Longitud del gen:</b> 3,591nt		<a href="#">8607&amp;report=genbank</a>		RepeatMasker: 91 nt	
<b>CD244</b> (molécula CD244)	1q23.3 <b>Inicio/Fin</b> NC_000001.11 (160830160.. 160862892, complement) <b>Longitud del gen:</b> 32,732	Codifica un receptor de la superficie celular expresado en las células asesinas naturales que median la destrucción restringida por el complejo de histocompatibilidad no mayor.	<a href="https://www.ncbi.nlm.nih.gov/nucleotide/NC_000001.1?report=genbank&amp;from=160830160&amp;to=160862892&amp;strand=true">https://www.ncbi.nlm.nih.gov/nucleotide/NC_000001.1?report=genbank&amp;from=160830160&amp;to=160862892&amp;strand=true</a>	9	WindowMasker 3,494 nt RepeatMasker 5,305 nt	Bazo
<b>CD79A</b> (molécula CD79A ó receptor de antígeno de linfocitos B)	19q13.2 <b>Inicio/Fin</b> NC_000019.10 (41877279..41881372) <b>Longitud del gen:</b> 4,093	Es un complejo multimérico que incluye el componente específico del antígeno, inmunoglobulina de superficie (Ig). Este gen codifica la proteína Ig-alfa del componente antígeno de células B. También se requiere para la expresión de la superficie de BCR y para la diferenciación eficiente de células B pro y pre-B.	<a href="https://www.ncbi.nlm.nih.gov/nucleotide/NC_000019.10?report=genbank&amp;from=41877279&amp;to=41881372">https://www.ncbi.nlm.nih.gov/nucleotide/NC_000019.10?report=genbank&amp;from=41877279&amp;to=41881372</a>	5	WindowMasker 1,449 nt  RepeatMasker 1,700 nt	Ganglio linfático
<b>FAM30A</b> (Family with sequence similarity 30 member A)	14q32.33 <b>Inicio/Fin:</b> NC_000014.9 (105917979..105932642) <b>Longitud</b> 14663	Es un gen de RNA y está afiliado a la clase lncRNA.	<a href="https://www.ncbi.nlm.nih.gov/nucleotide/NG_001019.6?from=947703&amp;to=962366&amp;report=genbank&amp;strand=true">https://www.ncbi.nlm.nih.gov/nucleotide/NG_001019.6?from=947703&amp;to=962366&amp;report=genbank&amp;strand=true</a>	6	WindowMasker 356 nt  RepeatMasker 790 nt	Ganglio linfático

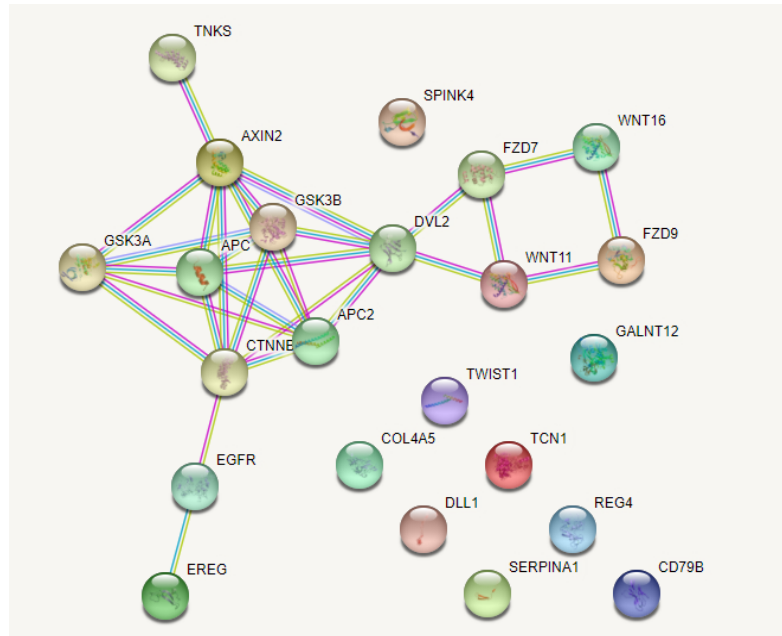
<b>AXIN2</b> (Axis Inhibition Protein 2)	17q24.1 <b>Inicio/Fin:</b> NC_000017.11 (65528563.. 65561648, complement) <b>Longitud del gen:</b> 33,085	Juega un papel importante en la regulación de la estabilidad de la beta-catenina en la vía de señalización de Wnt, como sus homólogos de roedores, conductina de ratón / axila de rata. Muestra una pérdida frecuente de heterocigosidad en el cáncer de mama, en el neuroblastoma hay otros tumores. Las mutaciones en este gen se han asociado con el cáncer colorrectal con reparación defectuosa de desajustes.	<a href="https://www.ncbi.nlm.nih.gov/nucleotide/NC_000017.1?report=genbank&amp;from=65528563&amp;to=65561648&amp;strand=true">https://www.ncbi.nlm.nih.gov/nucleotide/NC_000017.1?report=genbank&amp;from=65528563&amp;to=65561648&amp;strand=true</a>	15	WindowMasker: 627 nt  RepeatMasker: 467 nt	Endometrio
<b>MET</b> (proto-oncogene , receptor tyrosine kinase)	7q31.2 <b>Inicio/Fin:</b> NC_000007.14 (116672196.. 116798386) <b>Longitud del gen:</b> 126190	Este gen codifica un miembro de la familia de proteínas del receptor tirosina quinasa y el producto del protooncogén MET. La preproteína codificada se procesa proteolíticamente para generar subunidades alfa y beta que se unen mediante enlaces disulfuro para formar el receptor maduro. El procesamiento adicional de la subunidad beta da como resultado la formación del péptido M10, que se ha demostrado que reduce la fibrosis pulmonar. La unión de su ligando, el factor de crecimiento de los hepatocitos, induce la dimerización y activación del receptor, que desempeña un papel en la supervivencia celular, la embriogénesis y la migración e invasión celular. Las mutaciones en este gen están asociadas con carcinoma papilar de células renales, carcinoma hepatocelular y varios cánceres de cabeza y cuello. La amplificación y sobreexpresión de este gen también están asociadas con múltiples cánceres humanos.	<a href="https://www.ncbi.nlm.nih.gov/nucleotide/NG_008996.1?feature=any">https://www.ncbi.nlm.nih.gov/nucleotide/NG_008996.1?feature=any</a>	24	WindowMasker: 6003 nt RepeatMasker: 5992 nt.	Placenta
<b>F2RL1</b> (F2R like trypsin receptor 1)	5q13.3 <b>Inicio/Fin</b> NC_000005.10 (76819030.. 76835315) <b>Longitud del gen:</b> 16,285	Este gen codifica un miembro de la familia de proteínas del receptor 1 acoplado a proteína G y da como resultado un nuevo extremo amino que actúa como un ligando anclado que se une a un dominio de bucle extracelular. Esta proteína también es importante en la respuesta inflamatoria, así como en la inmunidad innata y adaptativa.	<a href="https://www.ncbi.nlm.nih.gov/nucleotide/NC_000005.1?report=genbank&amp;from=76819030&amp;to=76835315">https://www.ncbi.nlm.nih.gov/nucleotide/NC_000005.1?report=genbank&amp;from=76819030&amp;to=76835315</a>	3	WindowMasker: 398 nt RepeatMasker: 336 nt	Colon
<b>LAMC2</b> (laminin subunit gamma 2)	1q25.3	Tienen una estructura cruciforme que consta de 3 brazos cortos. Cada cadena de laminina es una proteína codificada por un gen distinto. Se expresa en varios tejidos fetales y se localiza específicamente en las células epiteliales de la piel,	<a href="https://www.ncbi.nlm.nih.gov/nucleotide/NG_007079.2">https://www.ncbi.nlm.nih.gov/nucleotide/NG_007079.2</a>	23	WindowMasker: 644 nt	Pulmón

	<b>Inicio/Fin:</b> NC_000001.11 (183186264..183245127) <b>Longitud del gen:</b> 58,863.	pulmones y riñones. Las variaciones del gen se expresan diferencialmente en tejidos embrionarios.	<a href="https://www.ncbi.nlm.nih.gov/nucleotide/NC_000001.11?from=5001&amp;to=63864&amp;report=genbank">?from=5001&amp;to=63864&amp;report=genbank</a>		RepeatMasker: 636 nt	
<b>CDH1</b>  (Cadherin 1)	16q22.1  <b>Inicio/Fin</b> NC_000016.10 (68737292..68835537) <b>Longitud de gen:</b> 98245	Codifica una cadherina clásica perteneciente a la superfamilia de cadherinas. Participa en los mecanismos que regulan las adherencias célula a célula, y en la movilidad y proliferación de células epiteliales. Sirve también como supresor invasivo, por lo que la mutación de este gen está relacionada con el desarrollo de cáncer .	<a href="https://www.ncbi.nlm.nih.gov/nucleotide/NC_000016.10?from=5001&amp;to=103250&amp;report=genbank">https://www.ncbi.nlm.nih.gov/nucleotide/NC_000016.10?from=5001&amp;to=103250&amp;report=genbank</a>	16	WindowMasker 756 nt  RepeatMasker 314 nt	Tiroides

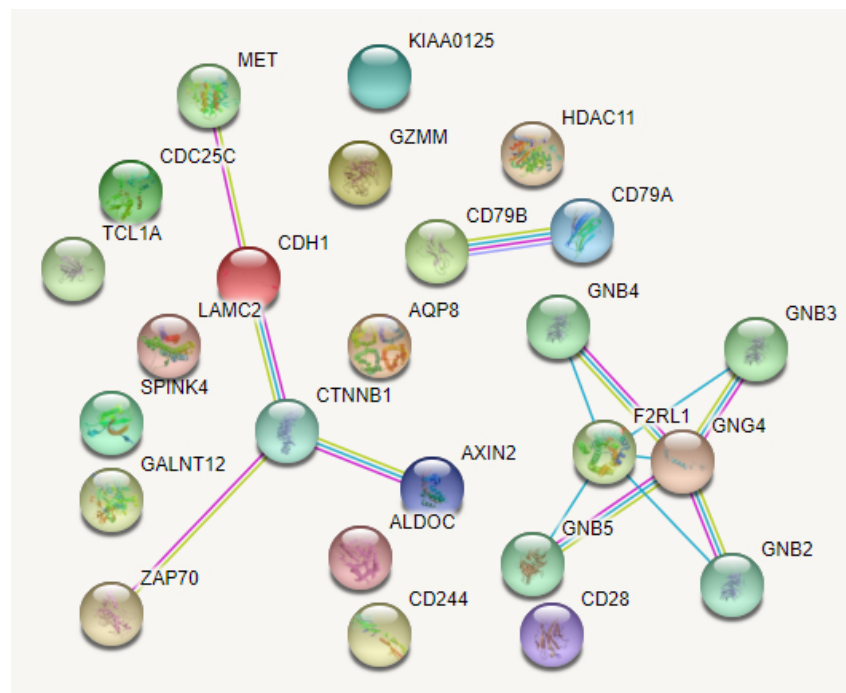
## Análisis de Resultados

### *Búsqueda de las interconexiones de los DEGs*

Con el fin de comprender la relación de funcionalidad de los genes diferencialmente expresados generamos dos redes de interconexión con los grupos de estudio:



**Figura 9.** Red de interacción proteína - proteína, para los genes diferencialmente expresados para cáncer de colon en etapa inicial.



**Figura 10.** Red de interacción proteína - proteína, para los genes diferencialmente expresados para metástasis peritoneal.

### *Selección de genes principales*

Al obtener de el programa R los genes sobreexpresados del dataset seleccionado enfatizando en el tejido de cáncer de colon en etapa inicial con presencia de mucina, sin mucina y para el tejido en cáncer de colon en metástasis peritoneal con presencia de mucina y sin mucina. Expresado en nuestros resultados tabla 1 y 2, se investigaron las funciones, localización y secuenciación de 18 genes que fueron relacionados principalmente en *string* para obtener información sobre sus mutaciones y sus frecuencias, tejidos en el que están más expresados función y zonas codificantes.

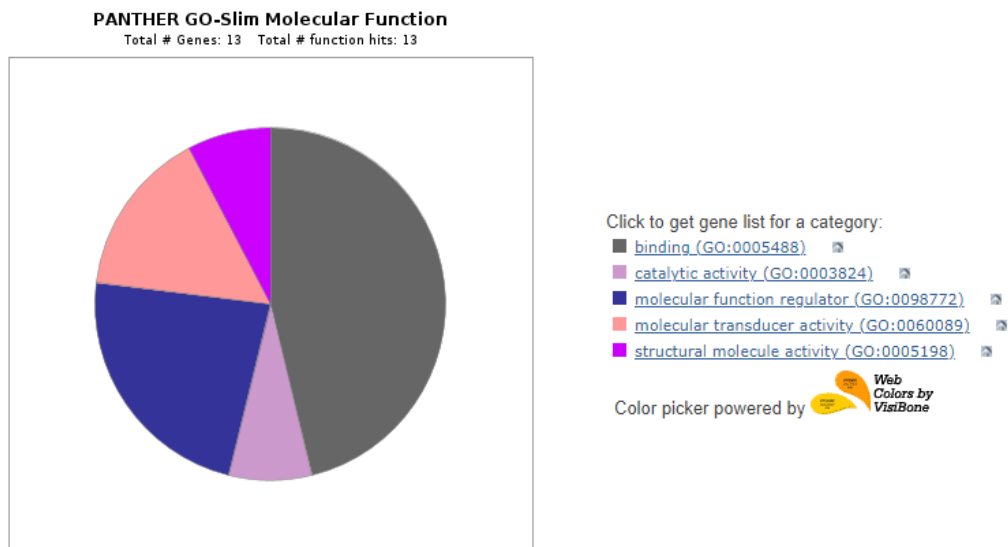
Se colocaron los 13 genes sobreexpresados en cáncer de colon en etapa inicial en *string* para la asociación funcional de las proteínas en donde los más expresados no mostraban una asociación de función proteica, por lo que se tomaron otros que sí las tenían y estaban adjuntos a varias de las mismas como se muestra en la Tabla 6 y tabla 7 más adelante.

**Gene ontology** es una herramienta bioinformática que permitió la agrupación de los genes según el funcionamiento biológico que tienen.

Se implementó el uso de esta herramienta para relacionar el funcionamiento que tienen entre la lista de genes obtenida para el estudio de cáncer de colon en etapa inicial (tabla 1) y metástasis peritoneal (tabla 2), en ambos casos para muestras en presencia de mucina y en ausencia de mucina.

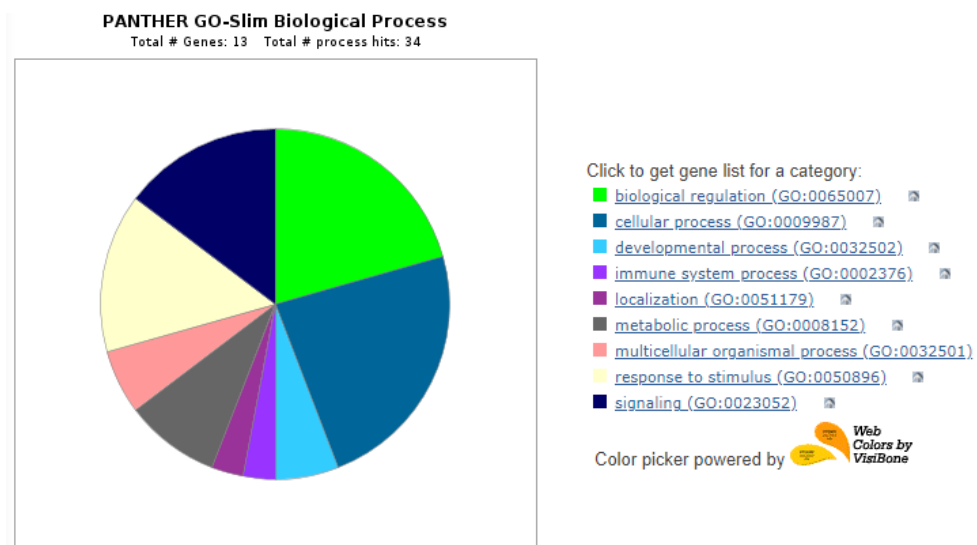
### **Cáncer de colon en etapa inicial.**

En la figura 11 se puede observar la relación entre las funciones moleculares que hay entre genes. El 46% corresponden a la categoría unión, mientras que el 23% representan aquellos genes por función regulatoria molecular.



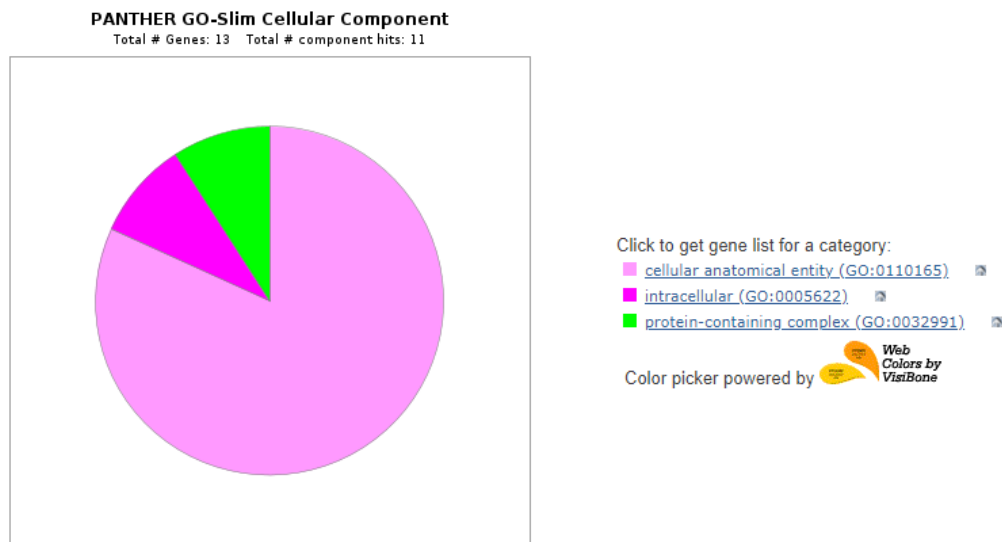
**Figura 11.** Función molecular genes en cáncer de colón etapa inicial.

En la figura 12 se observa el proceso biológico de los genes previamente seleccionados donde el 23% representan a los procesos celulares, esto seguido por el 20% que corresponden a las regulaciones biológicas.



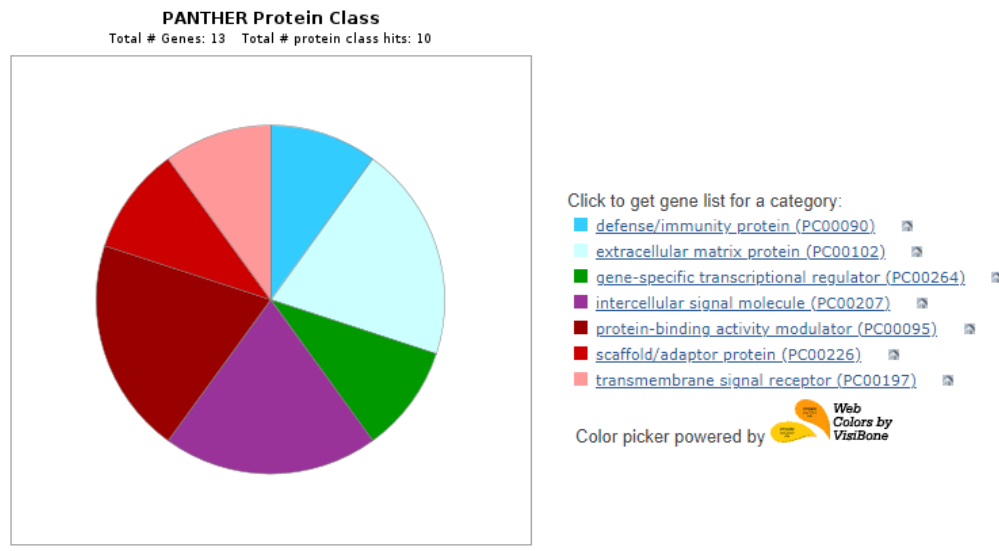
**Figura 12.** Proceso biológico de genes en cáncer de colón etapa inicial.

Posteriormente en la figura 13 se analiza el componente celular, donde el 81% pertenecen a la entidad, mientras que el 9.1% corresponde a los genes pertenecientes a un componente intracelular.



**Figura 13.** Componente celular de genes en cáncer de colón etapa inicial.

Como se puede observar en la figura 14 se muestra la gráfica de pastel correspondiente a las clases proteicas. El 20% representa a los genes pertenecientes a proteínas de la matriz extracelular, así como otro 20% corresponde al modulador de la actividad de unión de proteínas.



**Figura 14.** Clases de proteína de genes en cáncer de colón etapa inicial.

En la figura 15 se analizan y relacionan los efectos a nivel tejido, por lo antes mencionado es posible obtener información muy valiosa con relación a los genes. Viendo la relación con los primeros 4, podemos observar un papel importante en la formación y progresión tumoral por

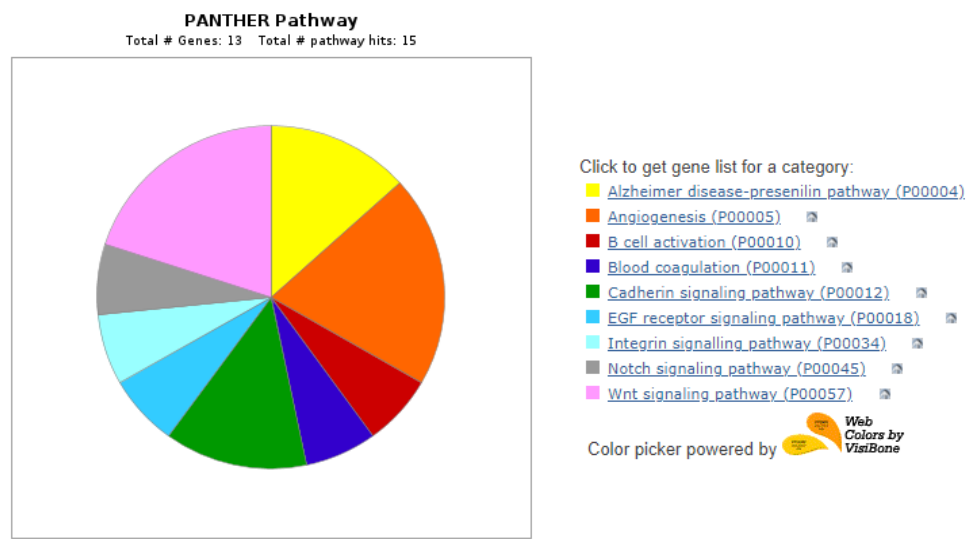


parte de la angiogénesis, la activación de células B, la coagulación sanguínea y la vía de señalización Cadherin. La angiogénesis es un proceso biológico el cual se encarga de la formación de los vasos sanguíneos nuevos a partir de los preexistentes. Para comprender su importancia en el desarrollo canceroso, los vasos sanguíneos se encargan de la morfogénesis en los órganos en la etapa de vasculogénesis y cuando la persona llega a una edad adulta estas células se encuentran en reposo pero la angiogénesis sigue activa de manera fisiológica en caso de reparación de heridas. Cuando la angiogénesis se activa en estos casos, se pasa por otro proceso fisiológico conocido como hemostasia. Su función es mantener el flujo sanguíneo y en caso de un vaso lesionado, necesita formar coágulos de sangre para reducir la pérdida. En el momento que este proceso deja de ser regulado y hay una presencia de células tumorales, inducen angiogénesis para su crecimiento creando sus propios vasos sanguíneos. Con esta información también podemos relacionar un factor de inflamación por la sobreexpresión de la proteína IL-8 la cual es sintetizada por leucocitos. [12] Esto provoca que la membrana basal se degrade, facilitando la angiogénesis del tumor. La activación de célula B se encuentra ligada al sistema inmunológico humoral adaptativo y son responsables de mediar en la producción de inmunoglobulina contra patógenos invasores, al activarse para atacar el tumor, provocando más inflamación y se vuelve un punto a favor del cáncer.[13] Finalmente podemos englobar como la disrupción de Cadherin Signaling Pathway juega un papel de suma importancia en cada uno de los procesos descritos previamente en la tumorigénesis en invasión temprana y metástasis debido a que este regula el crecimiento celular y estas proporcionan una unión entre células, manteniendo la arquitectura del tejido homeostático, al no tener esta adhesión normal, se pierde la regulación celular.[14]

Por otra parte, si se analiza la leyenda de “EGF receptor signaling pathway (P00018)” se puede relacionar que los genes pertenecientes a este grupo son de la familia receptor epidérmico humano, que a su vez se ve relacionado con 4 receptores quienes son los encargados de controlar las células del crecimiento. Esto lo podemos relacionar directamente con el cáncer de colon, debido a que si los genes responsables del crecimiento celular mandarían la señal oportuna para la destrucción de la célula, se podría evitar un crecimiento descontrolado de células cancerígenas causadas por mutaciones en el ARN. [18]

Adicionalmente al analizar el porcentaje perteneciente a “Integrin signalling pathway (P00034)”, se obtuvo una relación directa con el cáncer. Esto debido a que la señalización de la integrina es posible debido a que se ven involucrados la reorganización de la actina, activación de la MAPK y otras cascadas de señalización. Teniendo lo anterior en cuenta es que se realizó una investigación de cómo es que las cascadas de señalización pueden afectar la proliferación maligna. Obteniendo así que estudios recientes han demostrado que algunas cascadas de señalización estimulan el crecimiento celular mientras que reprimen la muerte celular programada, generando de esta manera que las células malignas sigan desarrollándose y tengan un ciclo más largo. [19]

Por último la leyenda “Wnt signaling pathway (P00057)” es de suma importancia debido a que actúa de manera similar como “EGF receptor signaling pathway (P00018)”. WNT catenina juega un papel muy importante en procesos de regulación, diferenciación, proliferación y muerte celular. Debido a lo antes mencionado es que si se llega a tener una alteración en Wnt se tendrá como consecuencia anomalías en el desarrollo, crecimiento y homeostasis en organismos animales. [20] Por lo que las alteraciones abordadas llegan a afectar a las células generando como consecuencia un crecimiento anormal de células cancerígenas.

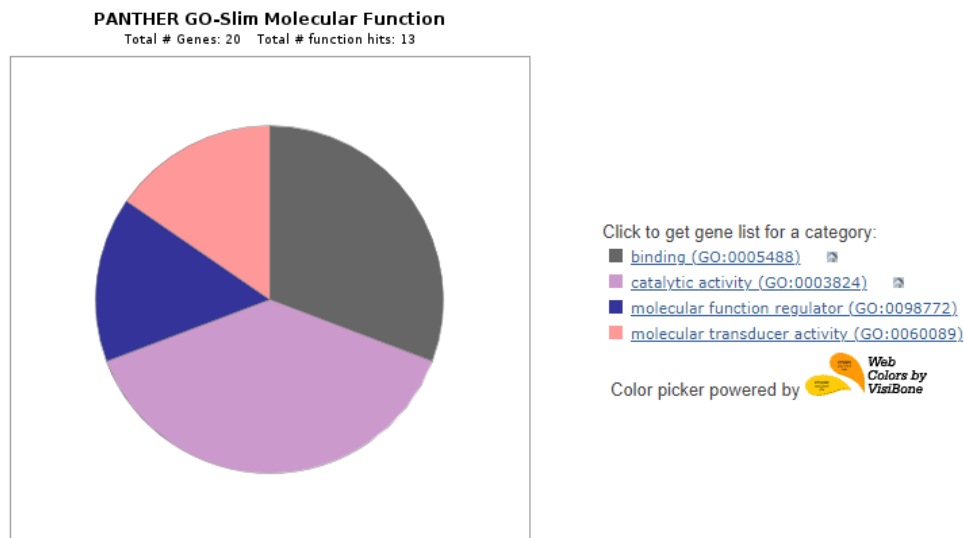


**Figura 15.** Efectos nivel tejido en cáncer de colon etapa inicial.

## Cáncer de colon en metástasis peritoneal.

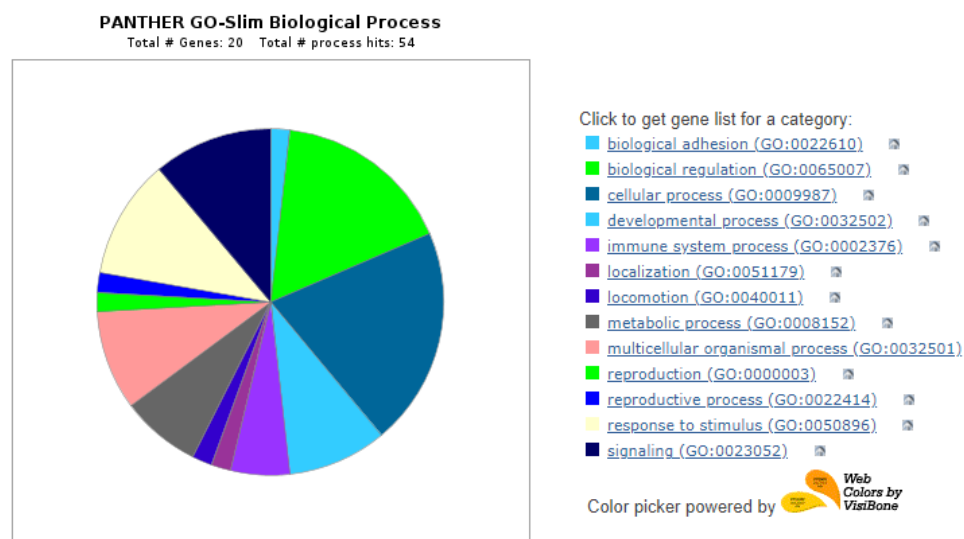
Se realizó el mismo procedimiento para metástasis peritoneal. Obteniendo de esta manera las siguientes 5 gráficas, figuras 16-20.

En la figura 16 se observa las funciones moleculares de los genes, donde el 30.8% corresponde a la categoría unión, mientras que el 38.5% representa la actividad catalítica de los genes.



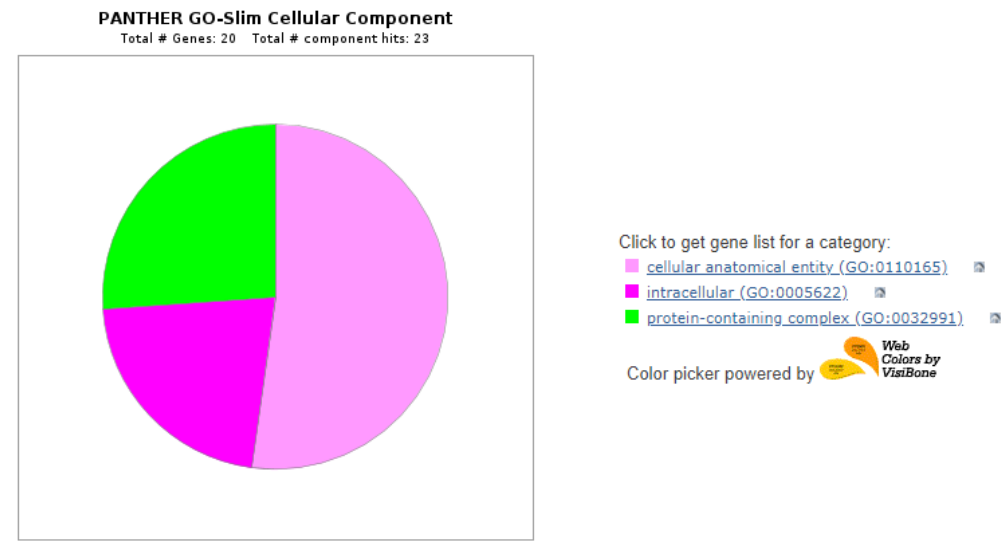
**Figura 16.** Función molecular de genes en cáncer de colon metástasis peritoneal.

En la gráfica presentada en la figura 17 se muestran los procesos biológicos de los genes, donde 20.4% corresponde a el proceso celular realizado, esto seguido por el 16.7% para los procesos regulatorios biológicos.



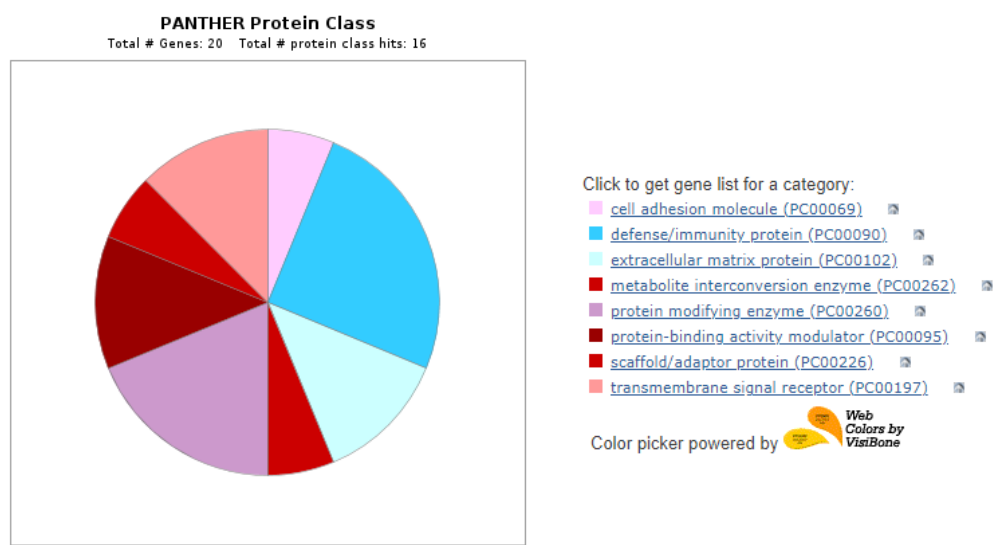
**Figura 17.** Procesos biológicos de genes en cáncer de colon metástasis peritoneal.

El comportamiento de los componentes celulares de genes en cáncer de colón en metástasis peritoneal se ven representados en la figura 18, donde el 52.3% pertenece al grupo de la entidad anatómica celular, mientras que el 21.7% corresponde a la relación intracelular.



**Figura 18.** Componentes celulares de genes en cáncer de colón metástasis peritoneal.

En la figura 19 se observan las clases de proteínas donde el 25.5% representa a la inmunidad de la proteína, por otra parte, el 18.8% corresponde a las proteínas modificadoras de las enzimas.



**Figura 19.** Clases de proteína de genes en cáncer de colón metástasis peritoneal.

Por último, se obtuvo la figura 20 donde se observan los procesos involucrados con los efectos en tejidos de colon en metástasis peritoneal. Para esta última gráfica se investigó e indagó sobre qué elementos relacionados con la problemática del cáncer de colon, obteniendo de esta manera las siguientes relaciones.

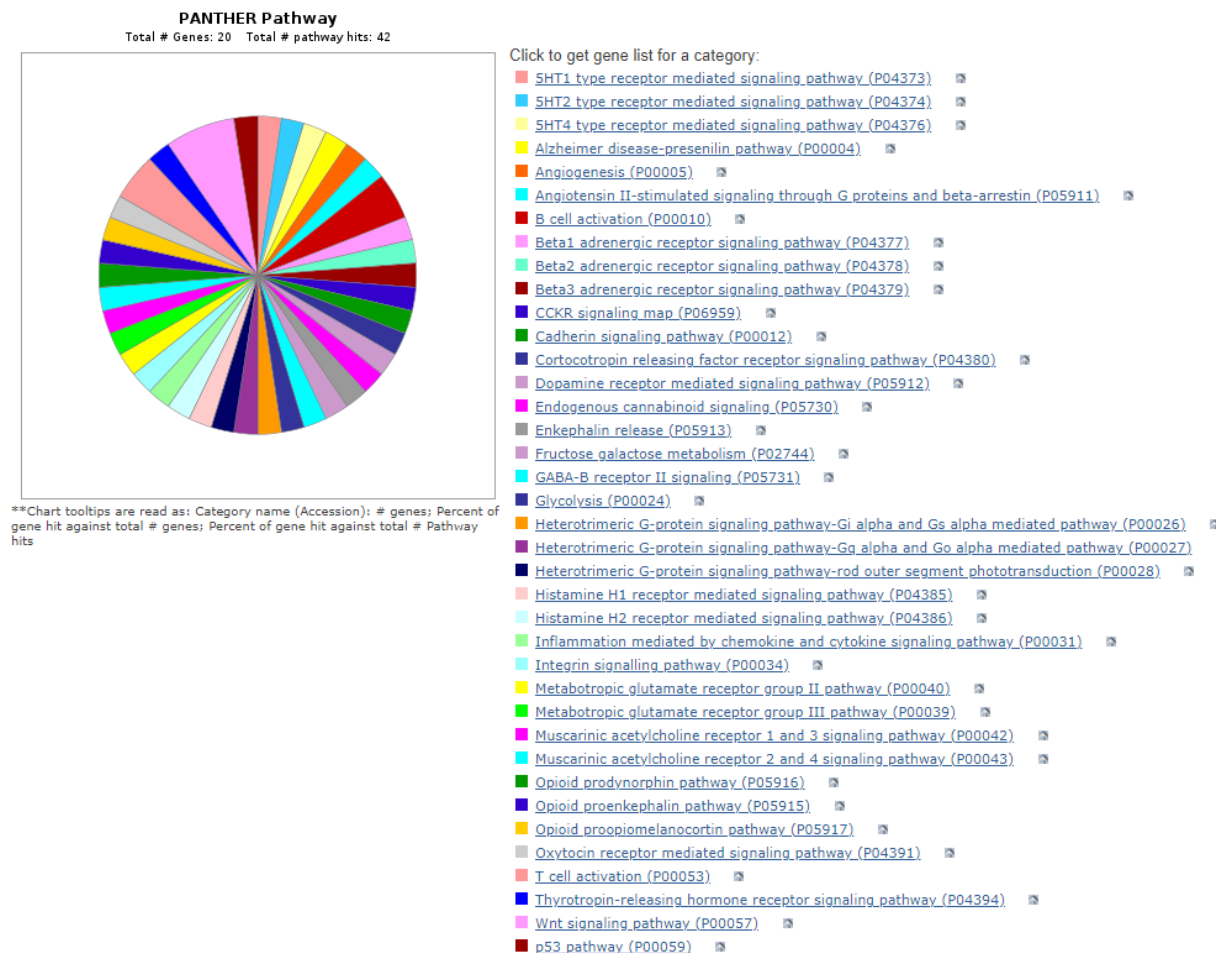
Al analizar e investigar la leyenda “CCKR signaling map (P06959)” se obtuvo que en este proceso se ve involucrada la hormona de la gastrina. Dicha hormona tiene como función principal estimular la secreción de ácido clorhídrico por parte de la mucosa gástrica formando así gastrina. El gen asociado a la gastrina está establecido en varios mecanismos del funcionamiento del cáncer, de igual manera, estos influyen en la expresión de genes que afectan directamente a la supervivencia celular, angiogénesis y la invasión de la misma. [21]

Examinando las categorías de la gráfica en pathway de genes, la gran mayoría están relacionadas con el sistema inmune y su reacción en presencia de células cancerígenas. Hay un gran porcentaje de receptores histamina lo que conlleva a la causa de inflamación por la liberación de citocina y quimiocinas que forman parte de la señalización para la activación de mecanismos celulares. También la proteína G y el receptor de metabotrópico que está unido los receptores de la membrana la cual inicia una serie de pasos metabólicos para la actividad celular.

Al examinar el contenido de la leyenda “T cell activation (P00053)”, se obtuvo que el cáncer de colon se desarrolla frente al sistema inmunitario asociado al tejido de la mucosa intestinal. Con respecto a lo anterior se ha observado que la respuesta del sistema inmune contra el tumor afecta directamente al pronóstico del cáncer de colon, por lo que en recientes estudios las células de activación “T” toman un lugar de suma importancia. Tomando en cuenta lo antes mencionado, es que las células “T” se modificaron en función de su estado oncogénico donde se pudo observar que si se contaba con un mayor número de células “T” contaban con un tumor inestable. [22]

Por último la leyenda “Wnt signaling pathway (P00057)” es de suma importancia. Esto debido a que WNT catenina juega un papel muy importante en procesos de regulación, diferenciación, proliferación y muerte celular. Debido a lo antes mencionado es que si se llega a tener una alteración en Wnt se tendrá como consecuencia anomalías en el

desarrollo, crecimiento y homeostasis en organismos animales. [20] Por lo que las alteraciones abordadas llegan a afectar a las células generando como consecuencia un crecimiento anormal de células cancerígenas.



**Figura 20.** Efectos en tejidos de cáncer de colón metástasis peritoneal.

Análisis de los genes mutados

En seguida se presentan los genes “relacionados” obtenidos de STRING, con mayor susceptibilidad a mutar:

**Tabla 6.** Comparativa de mutación en genes relevantes en string expresados para cáncer de colon en etapa inicial.

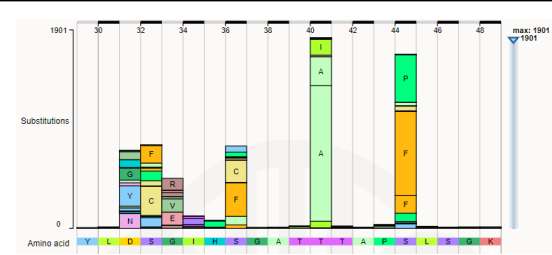
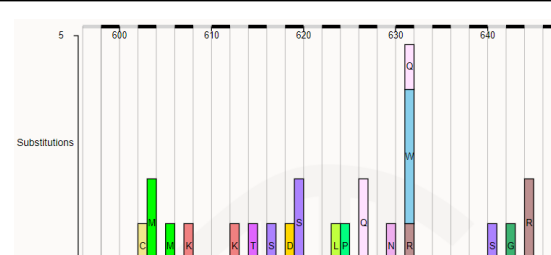
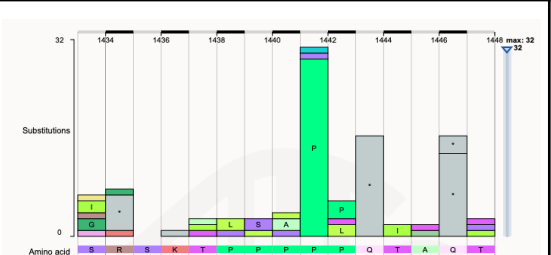
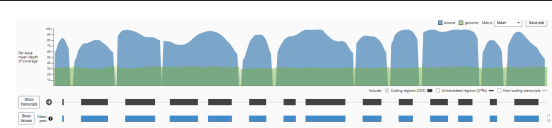
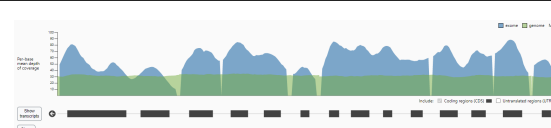
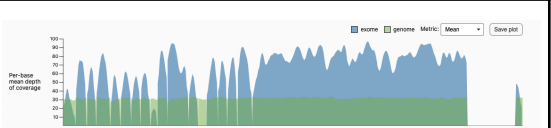
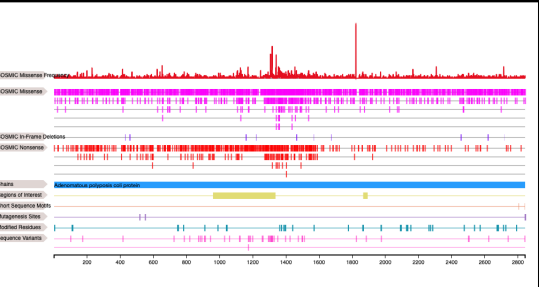
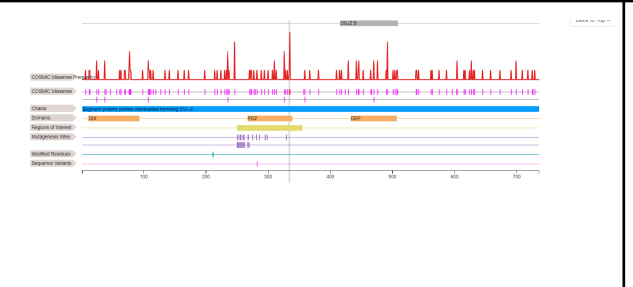
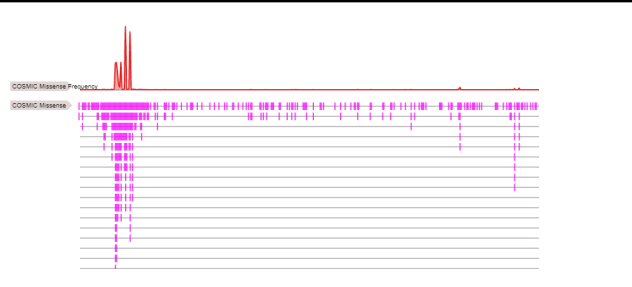
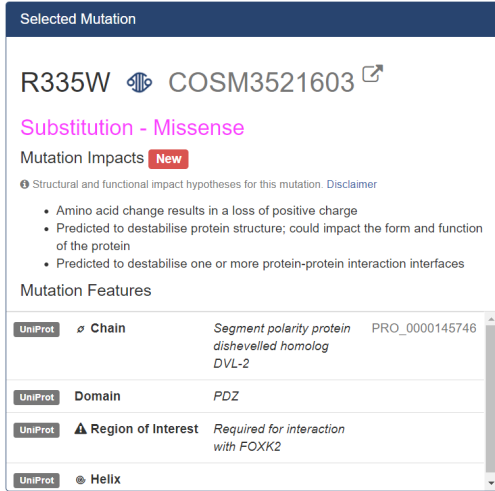
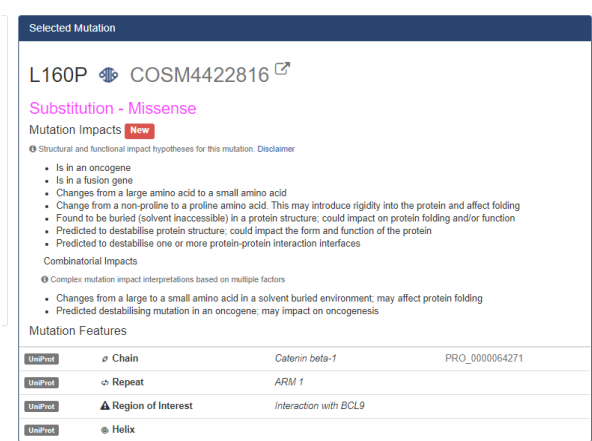
Genes	CTNNB1 (catenin beta 1)	DVL2 (dishevelled segment polarity protein 2)	APC (APC regulator of WNT signaling pathway)
Función	Esta proteína se une al producto del gen APC, que está mutado en la poliposis adenomatosa del colon. Las mutaciones en este gen son una causa de cáncer colorrectal.	Este gen codifica una proteína de 90 kD que se somete a fosforilación postraducciona para formar una proteína citoplasmática de 95 kD, que puede desempeñar un papel en la vía de transducción de señales mediada por múltiples proteínas Wnt. Es probable que los mecanismos de la función desordenada en la señalización de Wnt se conserven entre los metazoos.	Este gen codifica una proteína supresora de tumores que actúa como antagonista de la vía de señalización Wnt. Está en los procesos que incluyen la migración y adhesión celular, la activación transcripcional y la apoptosis. Se ha descubierto que en la mayoría de los cánceres colorrectales se producen mutaciones en el gen APC.
Localización	3:41236328-41301587	17:7128660-7137864	5:112043195-112181936
Tejido con mayor expresión	Placenta	Ovario	cerebro
Cambio de base (SNP)			
Zonas codificantes			

Gráfico de frecuencia de mutaciones



3D Mutación

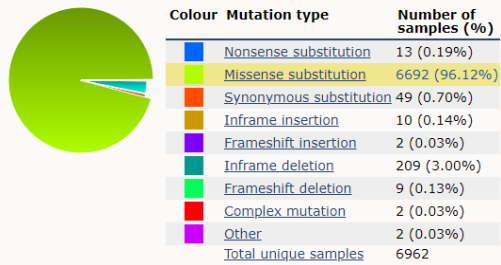




Cantidad de mutaciones:

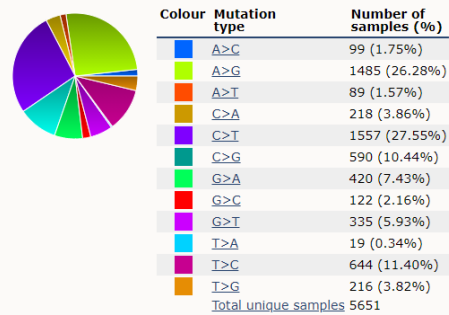
Summary

An overview of the types of mutation observed.



Substitutions

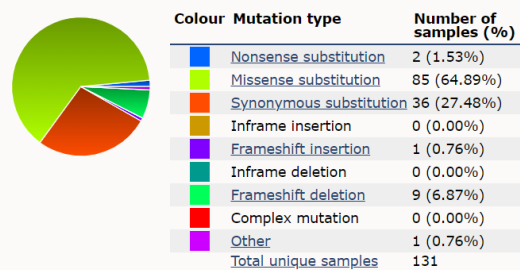
A breakdown of the observed substitution mutations.



There are 1100 substitutions where the nucleotide change is unknown.

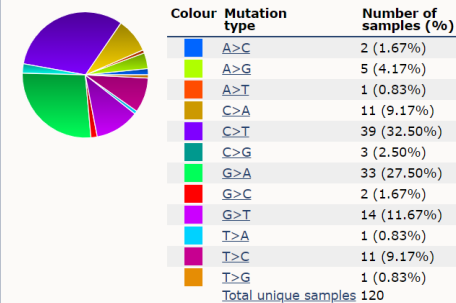
Summary

An overview of the types of mutation observed.



Substitutions

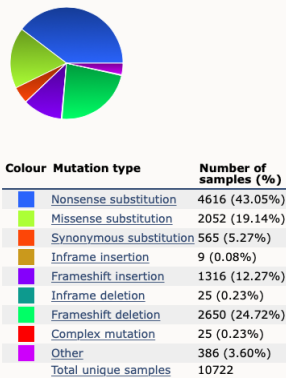
A breakdown of the observed substitution mutations.



There are 1 substitutions where the nucleotide change is unknown.

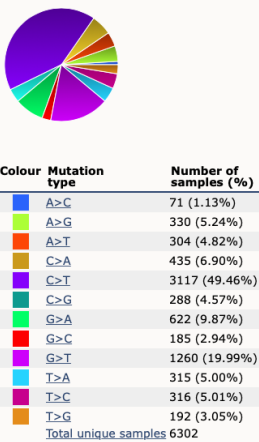
Summary

An overview of the types of mutation observed.



Substitutions

A breakdown of the observed substitution mutations.



**Tabla 7.** Comparativa de mutación en genes relevantes en string para metástasis peritoneal

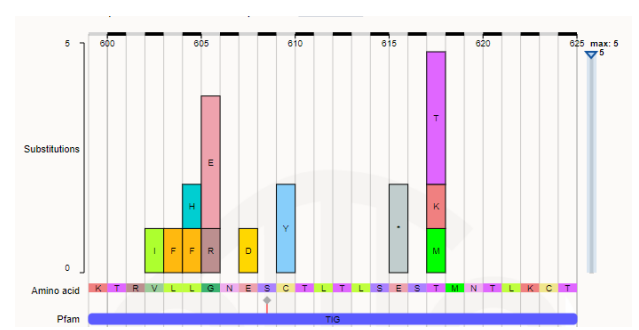
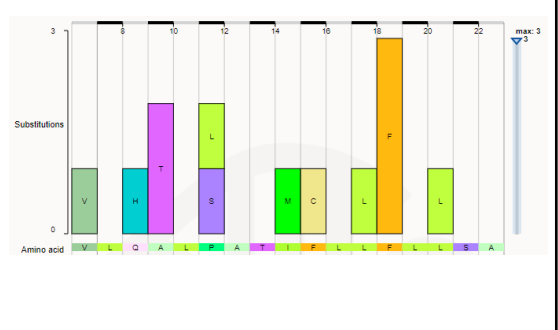
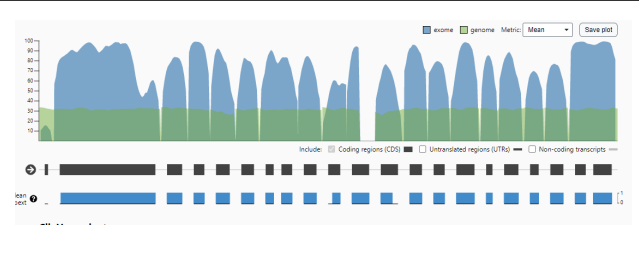
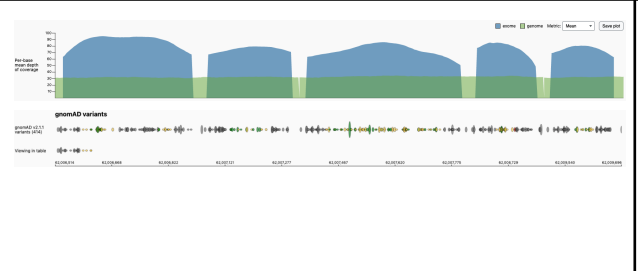
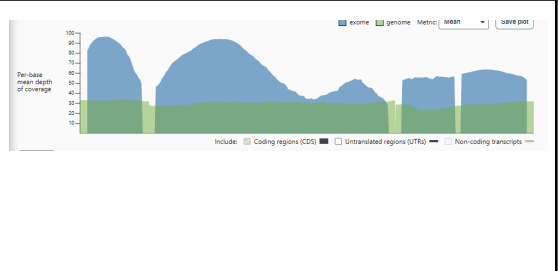
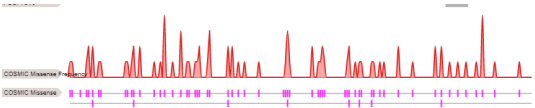
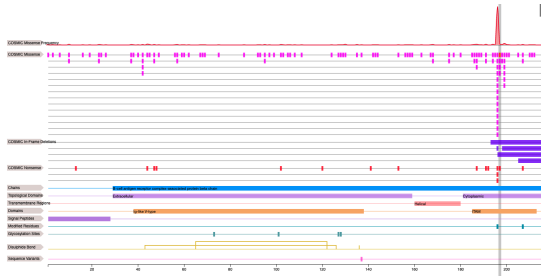
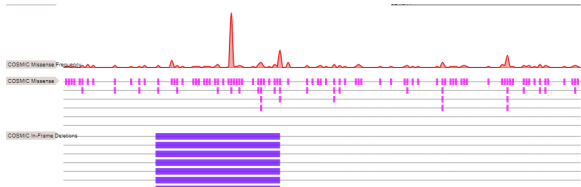
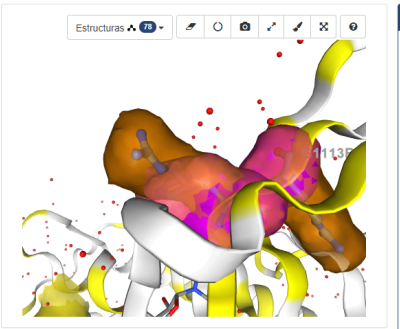
Genes	<b>MET</b> (proto-oncogene, receptor tyrosine kinase)	<b>CD79B</b> (molécula CD79B)	<b>CD79A</b> (molécula CD79A)
Función	La proteína se polimeriza en un gel del cual el 80% del peso está compuesto por cadenas laterales de oligosacáridos, además presenta un dominio central que contiene repeticiones en tándem ricas en treonina y prolina que varía entre 50 y 115 copias en diferentes individuos	Este gen codifica la proteína Ig-beta del componente antígeno de células B. Alternativamente, se han descrito variantes de transcripciones empalmadas que codifican diferentes isoformas.	El receptor de antígeno de linfocitos B es un complejo multimérico que incluye el componente específico del antígeno, inmunoglobulina de superficie (Ig).Surface Ig se asocia de forma no covalente con otras dos proteínas, Ig-alfa e Ig-beta, que son necesarias para la expresión y función del receptor de antígeno de células B. Este gen codifica la proteína Ig-alfa del componente antígeno de células B
Localización	7:116312444-116438440	17:62006100 - 62009714	19q13.2
Tejido con mayor expresión	Colon	Ganglio linfático	Ganglio linfático
Cambio de base(SNP)			
Zonas codificantes			

Gráfico de frecuencia de mutaciones



3D Mutación



Información

**C385Y** **COSM96939**

**Sustitución - Missense**

Impactos de mutación **Nuevo**

Hipótesis de impacto estructural y funcional para esta mutación. Descarga de responsabilidad

- Está en un oncogén
- Cambios de un aminoácido hidrofóbico a un aminoácido polar
- En un sitio de unión de moléculas pequeñas; podría afectar la unión de moléculas pequeñas
- Previsto para desestabilizar la estructura de la proteína; podría afectar la forma y función de la proteína

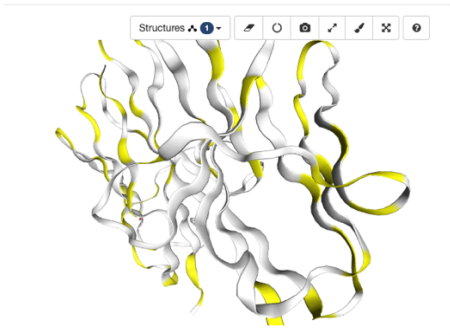
Impactos combinatorios

Interpretaciones complejas del impacto de las mutaciones basadas en múltiples factores

- Mutación desestabilizadora prevista en un oncogén; puede afectar la oncogénesis

Características de mutación

UniProt	Chain	Receptor del factor de crecimiento de hepatocitos	PRO_0000024440
UniProt	Topological Domain	Citoplasmático	



Selected Mutation

**E197G** **COSM3718288**

**Substitution - Missense**

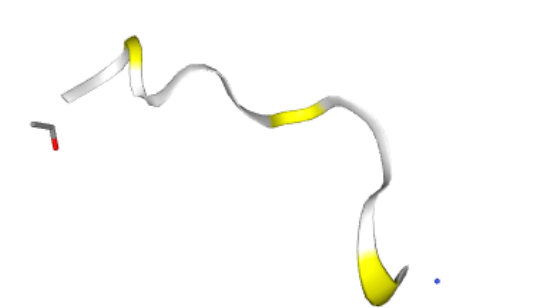
Mutation Impacts **Nuevo**

Structural and functional impact hypotheses for this mutation. Disclaimer

- Is in an oncogene
- Changes from a large amino acid to a small amino acid
- Change from a non-glycine amino acid to a glycine. This may introduce flexibility into the protein and affect folding
- Amino acid change results in a loss of negative charge
- Changes from a polar amino acid to a hydrophobic amino acid

Mutation Features

UniProt	Chain	B-cell antigen receptor complex-associated protein ...	PRO_0000014560
UniProt	Topological Domain	Cytoplasmic	
UniProt	Domain	ITAM	



Selected Mutation

**G47R** **COSM3286484**

**Substitution - Missense**

Mutation Impacts **Nuevo**

Structural and functional impact hypotheses for this mutation. Disclaimer

- Is in an oncogene
- Changes from a small amino acid to a large amino acid
- Change from a glycine to a non-glycine amino acid. This may introduce rigidity into the protein and affect folding
- Amino acid change results in a gain of positive charge
- Changes from a hydrophobic amino acid to a polar amino acid

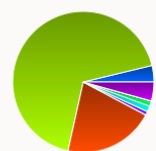
Mutation Features

UniProt	Chain	B-cell antigen receptor complex-associated protein ...	PRO_0000014558
UniProt	Topological Domain	Extracellular	
UniProt	Domain	Ig-like C2-type	
UniProt	Sequence Conflict	In Ref. 10; BAD97091.	

Cantidad de mutaciones:

Summary

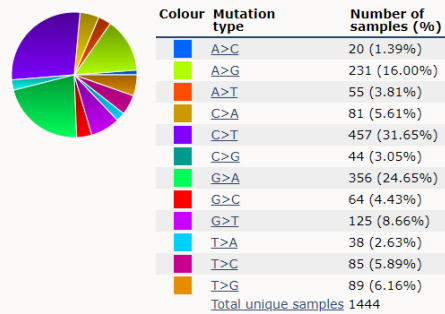
An overview of the types of mutation observed.



Colour	Mutation type	Number of samples (%)
<span style="color:blue">■</span>	<a href="#">Nonsense substitution</a>	67 (2.73%)
<span style="color:green">■</span>	<a href="#">Missense substitution</a>	1206 (49.16%)
<span style="color:orange">■</span>	<a href="#">Synonymous substitution</a>	369 (15.04%)
<span style="color:yellow">■</span>	<a href="#">Inframe insertion</a>	1 (0.04%)
<span style="color:purple">■</span>	<a href="#">Frameshift insertion</a>	14 (0.57%)
<span style="color:teal">■</span>	<a href="#">Inframe deletion</a>	24 (0.98%)
<span style="color:lightgreen">■</span>	<a href="#">Frameshift deletion</a>	22 (0.90%)
<span style="color:red">■</span>	<a href="#">Complex mutation</a>	0 (0.00%)
<span style="color:magenta">■</span>	<a href="#">Other</a>	77 (3.14%)
	<a href="#">Total unique samples</a>	2453

Substitutions

A breakdown of the observed substitution mutations.



Summary

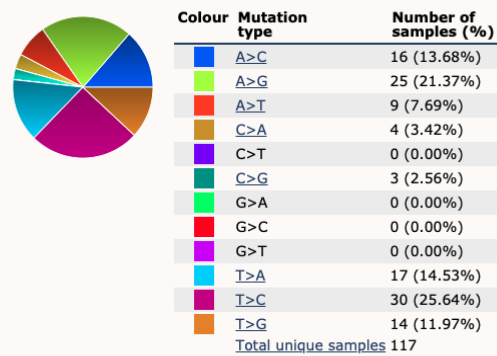
An overview of the types of mutation observed.



Colour	Mutation type	Number of samples (%)
<span style="color:blue">■</span>	<a href="#">Nonsense substitution</a>	4 (3.39%)
<span style="color:green">■</span>	<a href="#">Missense substitution</a>	114 (96.61%)
<span style="color:red">■</span>	<a href="#">Synonymous substitution</a>	0 (0.00%)
<span style="color:yellow">■</span>	<a href="#">Inframe insertion</a>	0 (0.00%)
<span style="color:purple">■</span>	<a href="#">Frameshift insertion</a>	0 (0.00%)
<span style="color:teal">■</span>	<a href="#">Inframe deletion</a>	0 (0.00%)
<span style="color:lightgreen">■</span>	<a href="#">Frameshift deletion</a>	0 (0.00%)
<span style="color:red">■</span>	<a href="#">Complex mutation</a>	0 (0.00%)
<span style="color:magenta">■</span>	<a href="#">Other</a>	0 (0.00%)
	<a href="#">Total unique samples</a>	118

Substitutions

A breakdown of the observed substitution mutations.



There are [1 substitutions](#) where the nucleotide change is unknown.

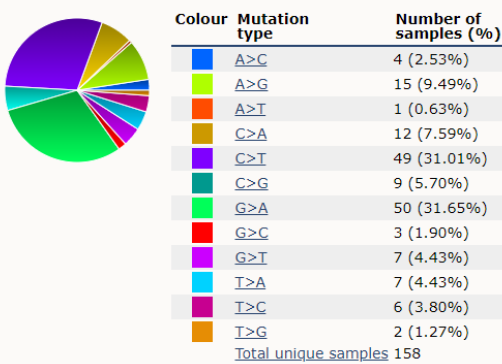
An overview of the types of mutation observed.



Colour	Mutation type	Number of samples (%)
<span style="color:blue">■</span>	<a href="#">Nonsense substitution</a>	0 (0.00%)
<span style="color:green">■</span>	<a href="#">Missense substitution</a>	7 (50.00%)
<span style="color:orange">■</span>	<a href="#">Synonymous substitution</a>	6 (42.86%)
<span style="color:yellow">■</span>	<a href="#">Inframe insertion</a>	0 (0.00%)
<span style="color:purple">■</span>	<a href="#">Frameshift insertion</a>	0 (0.00%)
<span style="color:teal">■</span>	<a href="#">Inframe deletion</a>	0 (0.00%)
<span style="color:lightgreen">■</span>	<a href="#">Frameshift deletion</a>	1 (7.14%)
<span style="color:red">■</span>	<a href="#">Complex mutation</a>	0 (0.00%)
<span style="color:magenta">■</span>	<a href="#">Other</a>	0 (0.00%)
	<a href="#">Total unique samples</a>	14

Substitutions

A breakdown of the observed substitution mutations.



There are [5 substitutions](#) where the nucleotide change is unknown.

### ***Análisis de mutaciones de los genes principales***

Para la selección de genes a analizar, se usó en el mapa de calor (Fig 8) obtenida con R. Se introdujeron los genes diferencialmente expresados para cáncer de colon en etapa inicial en un STRING y los genes diferencialmente expresados para metástasis peritoneal en otro para saber cómo se relacionaban entre ellos, ya que algunos genes como DLL1, CD244 no tenían ninguna interacción física, por lo que se tuvieron que descartar como objetos de investigación de genes de mutación.

Esta investigación se hizo únicamente sobre 6 genes que tenían gráfica 3D en la página “COSMIC” de donde fue consultada la información junto con la página “GenomAD”. A continuación se dará una explicación más precisa sobre cada gen.

#### **Análisis de Tabla 6**

- **CTNNB1** (catenin beta 1)

Este gen es uno de los pocos que están directamente relacionados al cáncer de colon ya que sus mutaciones generan un cambio de aminoácidos hidrofóbicos a aminoácidos polares que pueden afectar la unión de moléculas pequeñas y la oncogénesis. La proteína codificada de este mismo puede ser responsable de transmitir una señal de inhibición por contacto para generar que las células dejen de dividirse una vez completada la capa epitelial. Las regiones codificantes de este gen son aproximadamente el doble de grandes en la regulación del funcionamiento de este gen.

- **DVL2** (dishevelled segment polarity protein 2)

Este gen codifica una proteína que se somete a fosforilación para formar otra proteína que puede formar parte de la transducción de señales mediadas por múltiples proteínas. Su mutación es “R335W” y genera un impacto en la estructura de la proteína que la desestabiliza debido a que sus aminoácidos cambian los resultados de cargas positivas y en la interacción interfaccética de una o más proteínas. El mayor número de sustituciones se generan en el aminoácido Y, sin embargo, esta mutación es escasa o poco frecuente en este gen.

- **APC** (APC regulator of WNT signaling pathway)

Este gen es otro que está altamente relacionado con los cánceres colorrectales. Este codifica a una proteína que actúa en contra de la vía de señalización Wnt y forma parte de los procesos que incluyen la migración y adhesión celular y se encuentra mayormente en el tejido del

cerebro. Sus mutaciones se encuentran en el tumor supresor del gen, este muta debido a que sus aminoácidos generan un cambio en la ganancia de cargas negativas y cambia de aminoácido hidrofóbico a polar. Con el 43%, el mayor tipo de sus mutaciones es de sustitución sin sentido.

### **Análisis de la Tabla 7**

- **MET** (proto-oncogene, receptor tyrosine kinase)

El gen MET participa en el envío de señales dentro de las células y en la multiplicación y supervivencia celulares, sus formas mutadas estimulan la multiplicación y la diseminación de las células anormales. Este gen es el único de la lista que se encuentra en cáncer de colon como el tejido mayor expresado y también se encuentra relacionado en la red string.

- **CD79B** (molécula CD79B)

El gen CD79B es un codificador de proteínas, su mutación es el residuo fosforilado cambió a residuo no fosforilable, tiene un probable impacto funcional. Su tejido más expresado no es el cáncer del colon, sin embargo se encuentran relacionados en la red string y contiene su gráfica 3D, por lo que se tomó como importante en la parte de la investigación de mutaciones.

- **CD79A** (molécula CD79A)

El gen CD79A codifica la proteína Ig-alfa del componente antígeno de células B. De igual manera que el gen CD79B su tejido más expresado no es el cáncer de colon, pero si se encuentra relacionada en la red string con los demás genes. En comparación con el gen CD79B, el gen CD79A tiene un poco más relación con el cáncer de colon.

## **Conclusión**

Retomando la hipótesis planteada se determinó que la presencia o ausencia de mucina, es un factor que contribuye al desarrollo del cáncer de colon, sin embargo, se encontró que hay otros factores que trabajan en conjunto con esta sustancia contribuyendo al desenvolvimiento de la etapa inicial y metástasis peritoneal.

Gracias a la investigación y el análisis de los resultados podemos respaldar que el desarrollo de cáncer de colon en etapa inicial de la cavidad peritoneal depende de la activación y reacción del sistema inmunológico y nervioso. Aquellos genes sobreexpresados y que a su vez se encuentran relacionados con otras proteínas encargadas de procesos biológicos de reparación y adaptación nos demostraron cómo es que sus funciones afectan directamente la propagación tumoral.

Adicionalmente se logró extraer del análisis de datos sobre cómo es que los genes se comportan y qué consecuencias esto puede traer. Teniendo en cuenta lo antes mencionado, es que se encontró relación directa con los genes al crecimiento descontrolado de las células cancerígenas, esto debido a que las células embaucan a los genes obteniendo así que la detección sea más tardía.

Por ello se genera inflamación local como reacción de los receptores gracias a la liberación de moléculas orgánicas ante la presencia de células tumorales y del sistema inmunológico humoral adaptativo. Debido a la inflamación, la membrana basal se degrada, facilitando de esta manera una invasión temprana del tumor al inducir angiogénesis para una mejor adaptación a la cavidad peritoneal desencadenando una mayor pérdida de control, diferenciación y proliferación en el crecimiento celular. Además, se logró observar que las mutaciones que se generan por un cambio de aminoácidos hidrofóbicos a aminoácidos polares, influyen en la detonación del cáncer colorrectal.

## **Reflexión Personal**

A lo largo de la historia hemos observado cómo las ciencias biológicas llevan a cabo múltiples análisis de datos para la obtención de información que ayude al mejor entendimiento de los sistemas biológicos, así como comprender la relación que hay entre esta información, sin embargo con el paso del tiempo esta cantidad de datos ha ido en aumento, volviéndose cada vez más difícil su manejo y comprensión; y convirtiéndose en todo un reto para los científicos e investigadores la capacidad de poder procesar toda esta gran cantidad de datos .

Gracias a los grandes avances tecnológicos que se han desarrollado hasta el momento es posible que podamos manejar grandes volúmenes de información para la realización de estos estudios biológicos haciendo uso de la ciencia de datos. Este campo de estudio, que combina la ciencia estadística, métodos científicos y análisis de datos, ha contribuido en la generación de beneficios para los diferentes sectores, dentro de ellos se encuentra el área de la salud pública. Al hacer uso de la ciencia de datos dentro de esta área es posible obtener una serie de tendencias y comportamientos de la información estudiada contribuyendo en la mejora de la predicción de enfermedades, prevención de riesgos a la salud, mejora de tratamientos, haciendo posible la comparación de diferentes muestras de pacientes a nivel mundial.

Como se vio a lo largo de la resolución de toda la situación problema existen diversas herramientas bioinformáticas que brindan acceso a todos estos conjuntos de datos o bien de muestras realizadas en una determinada población y que gracias a plataformas como lo es R-Studio es posible hacer un análisis estadístico basado en el método científico para el procesamiento de los datos. Antes de comenzar el procesamiento de los datos en R, se debe tener un planteamiento adecuado de los métodos estadísticos que se van a emplear sobre los datos, esto permitirá que al ir analizando la información se tenga un conocimiento previo sobre el tipo de comportamiento que se estará observando y de esta forma poder determinar criterios y parámetros adecuados que sean útiles para la obtención de resultados, en el caso del análisis de genes como lo fue en la situación problema, uno de los criterios para la toma de decisiones fue establecer la proporción de las muestras estudiadas y dentro de los parámetros establecidos para el análisis fue la selección adecuada del nivel de significancia observado (p-value) y el tamaño del efecto grande es decir cuantas veces más se expresaban los genes en los grupos estudiados.



A partir de una buena selección de estos parámetros y criterios es posible obtener una lista de genes diferencialmente expresados, cuya expresión mayor estará relacionada a la proporción establecida en R, tendremos un grupo de genes que se encuentran con mayor expresión en cada uno de los grupos establecidos y a partir de estos genes hacer la construcción de gráficas que permitan darle una mejor interpretación a los datos obtenidos.

Es así que se puede observar cómo la ciencia de datos tiene una gran aplicación para la determinación de genes asociados a una enfermedad en específico y a través de una buena interpretación biológica se pueden brindar mejores herramientas para su tratamiento y prevención.

## Referencias

- [1] World Health Organization. (2014, 5 diciembre). *OMS | Datos y cifras sobre el cáncer*. Organización Mundial de la Salud. <https://www.who.int/cancer/about/facts/es/>
- [2] Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A (2015) Global cancer statistics, 2012. *CA Cancer J Clin* 65:87–108
- [3] Kuipers, E. J., Grady, W. M., Lieberman, D., Seufferlein, T., Sung, J. J., Boelens, P. G., van de Velde, C. J. H., & Watanabe, T. (2015). Colorectal cancer. *Nature Reviews Disease Primers*, 1(1), 1. <https://doi.org/10.1038/nrdp.2015.65>
- [4] O'Connell, J. B., Maggard, M. A., & Ko, C. Y. (2004). Colon Cancer Survival Rates With the New American Joint Committee on Cancer Sixth Edition Staging. *JNCI Journal of the National Cancer Institute*, 96(19), 1420–1425. <https://doi.org/10.1093/jnci/djh275>
- [5] Jayne D. Molecular biology of peritoneal carcinomatosis. *Cancer Treat Res* 2007; 134:21–33.
- [6] Barriuso, J., Nagaraju, R. T., Belgamwar, S., Chakrabarty, B., Burghel, G. J., Schlecht, H., Foster, L., Kilgour, E., Wallace, A. J., Braun, M., Dive, C., Evans, D. G., Bristow, R. G., Saunders, M. P., O'Dwyer, S. T., & Aziz, O. (2020). Early Adaptation of Colorectal Cancer Cells to the Peritoneal Cavity Is Associated with Activation of “Stemness” Programs and Local Inflammation. *Clinical Cancer Research*, 27(4), 1119–1130. <https://doi.org/10.1158/1078-0432.ccr-20-3320>
- [7] *RStudio* (1.4.1106). (2009-2021). [Software].
- [8] *Home - Nucleotide - NCBI*. National Center for Biotechnology Information. <https://www.ncbi.nlm.nih.gov/>
- [9] *GenBank Overview*. National Center for Biotechnology Information. <https://www.ncbi.nlm.nih.gov/genbank/>
- [10] *STRING: functional protein association networks*. STRING. [https://string-db.org/cgi/input?sessionId=bFFiQ4g73yJw&input\\_page\\_active\\_form=multiple\\_identifiers](https://string-db.org/cgi/input?sessionId=bFFiQ4g73yJw&input_page_active_form=multiple_identifiers)
- [11] *Gene Ontology*. (1999). Gene Ontology. <http://geneontology.org/>
- [12] Letelier, R. (2007) *La angiogénesis y el cáncer*. Medwave. <https://www.medwave.cl/link.cgi/Medwave/PuestaDia/Cursos/3546>
- [13] Ali Roghanian (2021) *B Cells*. British Society of immunology. Recuperado de <https://www.immunology.org/public-information/bitesized-immunology/cells/b-cells>

- [14] Yang, L. (2019) Cadherin Signaling in Cancer: Its Functions and Role as a Therapeutic Target Recuperado de <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6788064/>
- [15] C. (2020, 27 agosto). *COSMIC - Catalogue of Somatic Mutations in Cancer*. COSMIC. <https://cancer.sanger.ac.uk/cosmic>
- [16] *gnomAD*. (000). Genome Aggregation Database. <https://gnomad.broadinstitute.org/>
- [17] NCBI (2021) *Nucleotides*. Recuperado de <https://www.ncbi.nlm.nih.gov/gene/28514>
- [18] Musharraf J. (2019) *Delta Like-1 Gene Mutation: A Novel Cause of Congenital Vertebral Malformation*. Recuperado de <https://www.frontiersin.org/articles/10.3389/fgene.2019.00534/full#:~:text=The%20Delta%20Like%20D1%20>
- [19] Genet, J. (1996) Spectrum of mutations in the COL4A5 collagen gene in X-linked Alport syndrome. Recuperado de <https://pubmed.ncbi.nlm.nih.gov/8940267/>
- [20] BMC Cancer (2019) *Downregulated SPINK4 is associated with poor survival in colorectal cancer*. Recuperado de <https://bmccancer.biomedcentral.com/articles/10.1186/s12885-019-6484-5>
- [21] American Cancer Society. (2020). What Causes Colorectal Cancer?. Recuperado de: <https://www.cancer.org/cancer/colon-rectal-cancer/causes-risks-prevention/what-causes.html>
- [22] Mondragón, P. López, B. Gutiérrez, J. Suárez, A. Luna, R. Valle, a. (2015). Mecanismos de señalización intracelular en cáncer de tiroides. Recuperado de: <https://www.elsevier.es/es-revista-cirugia-cirujanos-139-articulo-mecanismos-senalizacion-in-tracelular-cancer-tiroides-S0009741116300512#:~:text=Los%20estudios%20actuales%20han%20puesto,natural%20de%20la%20proliferaci%C3%B3n%20celular>.
- [23] Ochoa, A. Juárez, I. Rosales, M. Barros, P. (agosto, 2012). Cirugía y cirujanos, academia mexicana de cirugía. *La vía de señalización Wnt- $\beta$ -catenina y su relación con cáncer*. Recuperado de: <https://www.redalyc.org/pdf/662/66224459015.pdf>
- [24] Panther. (2021). CCKR SIGNALING MAP\_ pathway description. Recuperado de: <http://www.pantherdb.org/pathway/pathDetail.do?clsAccession=P06959>.
- [25] Chirica, M. Bourhis, L. Lehmann, J. Chardiny, V. Bouhidel, F. Foulbouef, L. Gornet, J. Lourenco, N. Dulphy, N. Toubert, A. Allez, M. (2015). Phenotypic analysis of T cells infiltrating colon cancers: Correlations with oncogenetic status. Recuperado de: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4570110/>.