



UNIVERSIDAD PERUANA  
**CAYETANO HEREDIA**  
ESCUELA DE POSGRADO

ESTUDIO COMPUTACIONAL DE  
GENOMAS PARA LA IDENTIFICACIÓN  
DE VARIANTES GENÉTICAS  
ASOCIADAS A RASGOS CLÍNICO-  
PATOLÓGICOS DEL CÁNCER DE  
PRÓSTATA.

TESIS PARA OPTAR EL GRADO DE  
MAESTRO EN INFORMÁTICA BIOMÉDICA  
EN SALUD GLOBAL CON MENCIÓN EN  
BIOINFORMÁTICA

JORGE LUIS LARA ZUÑIGA

LIMA - PERÚ

2020



**ASESORA**

Dra. Claudia Inés Gloria Machicado Rivero

**JURADO DE TESIS**

MG. ERNESTO GOZZER INFANTE

PRESIDENTE

MG. JOSE LUIS ROJAS VILCA

VOCAL

MG. LEANDRO HUAYANAY FALCONI

SECRETARIO

## **DEDICATORIA**

A mi madre por ser mi fortaleza en los momentos más difíciles.  
A mi familia, que a pesar de las adversidades siempre son el mejor apoyo para  
continuar adelante.

## **AGRADECIMIENTOS.**

En primer lugar a Dios por guiar mi camino.

A mi asesora, la Dra. Claudia Machicado Rivero, por su tiempo y dedicación en la conducción del presente trabajo. Mi más profundo agradecimiento por su paciencia y orientación.

A los profesores y amigos de la Unidad de Informática Biomédica en Salud Global de la Facultad de Salud Pública y Administración (FASPA) por su apoyo y comprensión.

Al Dr. Cárcamo por la ayuda y orientación en parte de mi trabajo de investigación.

A Daniel Condor y Paola Pflucker por la paciencia, el apoyo y el empuje.

A CONCYTEC – FONDECYT por haber financiado el programa de maestría y el proyecto de investigación.

**FUENTES DE FINANCIAMIENTO.**

El presente trabajo de investigación fue financiado por CONCYTEC -  
FONDECYT

## Tabla de contenidos

1.	INTRODUCCIÓN .....	1
2.	PLANEAMIENTO DE LA INVESTIGACIÓN.....	5
2.1.	Planteamiento del problema.....	5
2.2.	Marco Teórico.....	8
2.2.1.	Cáncer de próstata.....	8
2.2.2.	Alteraciones genéticas y cáncer .....	19
2.2.3.	Proyectos Genomas.....	25
2.3.	Pregunta de Investigación .....	26
2.4.	Justificación de la Investigación .....	26
3.	OBJETIVOS .....	28
3.1.	Objetivo General.....	28
3.2.	Objetivos específicos .....	28
4.	Materiales y Métodos.....	29
4.1.	Diseño del estudio.....	29
4.2.	Fuente y características de los datos genómicos y clínicos.....	31
4.3.	Operacionalización de variables .....	32
4.4.	Instrumentos.....	34
4.5.	Secuencia de análisis.....	37
4.6.	Técnicas y procedimientos de recolección de datos .....	38
4.6.1.	Determinación del panorama mutacional en el CP.....	39
4.6.2.	Identificación de genes con alteración en el número de copias (ANC), amplificados y deletados. ....	42
4.6.3.	Evaluación de la expresión diferencial de genes en CP, genes con subexpresión y sobreexpresión. ....	44
4.6.4.	Evaluación de la asociación de alteraciones genéticas con el fenotipo del CP .....	45
4.6.5.	Identificación de rutas enriquecidas en CP. ....	52
4.6.6.	Redes de interacción proteína-proteína (interactomas).....	54
4.6.7.	Predicción del daño en la estructura y función de las proteínas.....	55
4.6.8.	Frecuencia de mutaciones conductoras o drivers del CP. ....	59
4.6.9.	Frecuencias de polimorfismos de un solo nucleótido (SNPs) de riesgo a CP. ....	62

5.	CONSIDERACIONES ETICAS .....	64
6.	Resultados .....	65
6.1.	Población de estudio .....	65
6.2.	Paisaje de las alteraciones genómicas en el CP .....	68
6.2.1.	Genes significativamente mutados.....	68
6.2.2.	Genes con alteración en el número de copias. ....	82
6.2.3.	Genes con expresión alterada.....	92
6.3.	Asociación del genotipo con el fenotipo del CP. ....	110
6.3.1.	Correlación de las mutaciones puntuales en GSMs y fenotipo.....	110
6.3.2.	Correlación entre la mutación puntual de los GFMs y el fenotipo. ....	115
6.3.3.	Correlación entre los genes con alteración en el número de copias y el fenotipo .....	123
6.3.4.	Genes con expresión diferencial .....	127
6.3.5.	GDEs asociados a la supervivencia.....	135
6.4.	Interactomas .....	136
6.5.	Predicción del daño de estructura y función .....	140
6.6.	Determinación de la frecuencia de las mutaciones conductoras en subgrupos divididos por fenotipo.....	151
6.7.	SNPs de riesgo a CP .....	153
7.	Discusión.....	154
8.	Conclusiones .....	175
9.	Limitaciones.....	177
10.	Recomendaciones .....	178
11.	Referencias Bibliográficas.....	179
12.	Anexos	

## **RESUMEN**

El cáncer de próstata (CP) es uno de los principales problemas de salud pública que afronta la población masculina a nivel global. El perfil de las alteraciones genéticas del CP es muy variable entre un individuo y otro, lo cual influye en la diversidad de manifestaciones clínicas incluyendo el tiempo de supervivencia y la respuesta a tratamientos. Actualmente los pocos biomarcadores del CP, como el PSA o PCA3, no son completamente eficientes, no habiendo herramientas predictivas adecuadas para el manejo clínico de este tipo de tumor. Además aún no se conoce de forma integral como las alteraciones genéticas se vinculan al fenotipo de la enfermedad, y como estas pueden servir para la identificación de nuevos marcadores más útiles en el pronóstico del CP. En este estudio se determinó el panorama mutacional del CP asociando dichas características fenotípicas a rasgos fenotípicos como la supervivencia y variables clínico-patológicas. Nuestro análisis secundario comprometió la consulta de bases de datos genómicas así como el empleo de recursos web en línea y programas bioinformáticos para el tratamiento y análisis de los datos clínicos y genómicos. Los resultados obtenidos mostraron que ciertas alteraciones genéticas en CP se asocian al fenotipo, incluyendo por ejemplo la amplificación de CSC8 asociado a una menor SG y la sobreexpresión de HIF3A asociado a un peor pronóstico. Esperamos que la información generada en esta tesis contribuya al reconocimiento de nuevos biomarcadores genéticos para mejorar los instrumentos de diagnóstico y pronósticos del CP.

**PALABRAS CLAVES:** cáncer de próstata, genoma, bioinformática.

## **ABSTRACT**

Prostate cancer (PC) is one of the main public health problems in male population worldwide. Currently, the diagnostic and prognostic biomarkers for PC, like PSA or PCA3, are not efficient; the predictive tools in clinical management do not allow discerning which treatment is most suitable for patients. Bioinformatics along with molecular biology has been generating relevant information about cancer biology through cancer genome projects. In this study, the landscape for PC was explored. In addition, the association of genetic variants in PC with prognostic variables such as overall survival and disease-free time was evaluated; as well as clinical-pathological variables. Online web resources and bioinformatics programs were used for the analysis of clinical and genomic data. The results showed that certain genetic alterations in CP are associated with the phenotype, including for example CSC8 amplification was associated with a short OS and overexpression of HIF3A associated with a worse prognosis. We hope that the information generated in this thesis will contribute to the recognition of new genetic biomarkers to improve diagnostic and prognoses tools in PC.

**KEY WORDS:** prostate cancer, genome, bioinformatics.

## 1. INTRODUCCIÓN

El cáncer de próstata (CP) es el segundo tipo de neoplasia con mayor incidencia en hombres a nivel mundial. Según el último reporte de la Agencia Internacional de Investigación sobre el Cáncer (IARC)-Globocan 2018, el CP registro una tasa de incidencia de 29.3 por cada 100 000 hombres a nivel global. (1). En el Perú, el CP es la neoplasia con mayor incidencia (7 598 nuevos casos diagnosticados en 2018) y con la más alta tasa de mortalidad (721 muertes) (1). Dentro de los principales factores de riesgo para CP se encuentran la edad, antecedentes familiares y grupos raciales; también algunas alteraciones genéticas y la alimentación podrían estar vinculadas a su etiología (2). Conforme la edad se incrementa también crece la probabilidad de desarrollar CP, después de los 50 años el riesgo aumenta significativamente. Por otra parte, si un familiar de primer o segundo grado de consanguinidad, como padre o hermano, de un hombre tienen CP, el riesgo de que ese individuo desarrolle ese tipo de neoplasia se duplica. Se sabe que el CP presenta disparidades en cuanto a la raza del individuo, presentándose con una mayor frecuencia en hombres de raza negra que en otros grupos raciales (3). La etapa inicial del CP, en la cual el tumor se encuentra localizado, es asintomática, conforme la enfermedad va progresando hacia la etapa metastásica los síntomas pueden presentarse como problemas al miccionar, sangrado en la orina, dolor en la parte lumbar u otras regiones del cuerpo. La gran mayoría de los casos de CP son del tipo histológico adenocarcinoma, otros tipos con menor incidencia son los sarcomas, carcinomas de células pequeñas, tumores neuroendocrinos y carcinomas de células transicionales.

Una de las principales causas de las altas tasas de incidencia y mortalidad del CP es la falta de un diagnóstico oportuno, y esto debido tanto a la ausencia de políticas públicas en prevención, como a las barreras socioculturales que existen en torno a la prueba de descarte por Examen Digital del Recto (EDR) por parte de los hombres así como al acceso limitado a los sistemas de salud en países en vías de desarrollo. Sumado a lo anterior, este tipo de neoplasia es compleja, donde la heterogeneidad genética del tumor introduce una gran variabilidad de fenotipos clínicos ligados a la progresión del cáncer y a las respuestas al tratamiento (4).

Actualmente los marcadores de detección y monitorización del CP no están claramente definidos, haciendo que las herramientas de diagnóstico, pronóstico y de manejo clínico de esta enfermedad sean poco eficientes. Lo anterior se traduce en biopsias y sobretratamientos innecesarios, sumado a las resistencias hormonales que desarrollan algunos pacientes. El marcador más ampliamente usado en el diagnóstico y seguimiento del CP es el antígeno prostático específico (PSA), el cual presenta limitaciones. El PSA posee una baja especificidad (33%) y un poco valor como biomarcador de pronóstico (5). Ante esto, se vuelve de suma importancia el reconocimiento de nuevos biomarcadores que sirvan para mejorar la precisión del diagnóstico y las herramientas de monitorización actuales que se utilizan en la práctica clínica. Además, se evitaría un gran número de biopsias de próstata innecesarias y a la vez constituirían una prueba diagnóstica menos invasiva con miras hacia otros tipos de tamizajes como las biopsias líquidas.

En tanto los gobiernos afrontan este problema de salud pública en mejoras de las políticas de prevención y concientización en la población masculina, la participación de la biología molecular y la bioinformática en la investigación del cáncer ha mejorado la comprensión y caracterización de la genómica de esta enfermedad. Gracias a las tecnologías de alto rendimiento, se han podido identificar mutaciones en la línea germinal, polimorfismos de un solo nucleótido (SNP) y alteraciones en el número de copias (ANC) asociados a un incremento del riesgo para CP (6). También se han reportado algunas mutaciones hereditarias en genes de reparación del ADN que están vinculadas a un perfil más agresivo de la enfermedad (7). La búsqueda de nuevos biomarcadores para CP también ha explorado los patrones de expresión diferencial de genes como PCA3, AR, PTEN o TGFβ1 en asociación al pronóstico (8). Sin embargo, el efecto de las alteraciones genéticas sobre el grado y etapa tumoral, sobre la progresión o respuestas a tratamiento en CP aún no ha sido establecido con claridad o presentan resultados controversiales.

Las tecnologías de alto rendimiento también han proporcionado una gran cantidad de datos acerca del perfil genómico del cáncer y como parte de ello han surgido proyectos dedicados al estudio de la biología del cáncer como El Atlas del genoma del cáncer (TCGA). TCGA es un proyecto creado en el 2005 con la finalidad de estudiar la genómica del cáncer e identificar cambios en el ADN de diferentes tipos de neoplasias, y es dirigido por el Instituto Nacional del Cáncer (NCI) y el Instituto Nacional de Investigación del Genoma Humano (NHGRI), entidades que forman parte de los Institutos Nacionales de Salud (NIH) de los

Estados Unidos. La información de TCGA está disponible públicamente para análisis secundarios de datos encaminados al descubrimiento de potenciales biomarcadores, que posteriormente experimentan un proceso que va desde la identificación del marcador, la diferenciación del caso del control, evaluación de la capacidad de detectar la enfermedad preclínica, indicaciones para la aplicación y validación y finalmente estudios de control de la neoplasia (9,10) . Se han conducido diversos estudios que comprometen análisis secundarios de los datos de TCGA en diversos tipos de cáncer. Así por ejemplo se han reconocido variantes genéticas asociadas a la recurrencia de cáncer de ovario y también se han identificado blancos terapéuticos potenciales para el cáncer de pulmón de células escamosas (11,12).

En el presente trabajo de tesis se evaluó la asociación entre el genotipo y el fenotipo de individuos con CP, mediante un análisis computacional de datos de TCGA complementado con otras herramientas de bioinformática. Con este análisis se pudo vincular ciertos rasgos genómicos con el perfil clínico de los pacientes como la supervivencia global (SG), la supervivencia libre de enfermedad (SLE) y algunas variables clínicas como el estadio patológico T, N, el puntaje de Gleason y PSA. El conocimiento producido en la presente investigación podría ser el primer paso hacia la identificación de biomarcadores que asistan el pronóstico de este tipo de neoplasia.

## **2. PLANEAMIENTO DE LA INVESTIGACIÓN**

### **2.1. Planteamiento del problema**

El CP es un problema de salud pública a nivel mundial por las altas tasas de incidencia y mortalidad reportadas (1), y por el costo social y económico que acarrea esta enfermedad (13). El CP es una neoplasia muy heterogénea que muestra diferentes manifestaciones clínico patológicas interindividuales incluidas la SG, la SLE, el grado y la etapa tumoral (14). Por ejemplo, los hombres afroamericanos tienen el doble de probabilidad de morir por esta neoplasia, comparados con hombres de otras razas (15). La variación en la progresión del cáncer y la respuesta al tratamiento está vinculada a la heterogeneidad genética del tumor (16,17).

Los biomarcadores actuales para el diagnóstico y manejo del CP no son del todo eficientes, ya que no permiten hacer un diagnóstico preciso de la enfermedad o predecir si un paciente desarrollara un CP de bajo o alto riesgo. Esto conlleva, entre otras cosas, a la toma de biopsias innecesarias y a la aparición de efectos secundarios luego de recibir un tratamiento inadecuado. En este sentido, la identificación de nuevos biomarcadores se vuelve un punto crítico en la lucha contra el CP y en la mejora de calidad de vida de los individuos con esta neoplasia.

Otros cánceres hormonales, como los de mama y de ovario, hacen uso de pruebas genéticas en el manejo clínico. Una de ellas es el tamizaje de las mutaciones en

BRCA 1/2 y otros genes de susceptibilidad que asisten a la intervención personalizada de cada paciente. En cáncer colorrectal ligado a síndrome de Lynch y en cáncer de endometrio, el tamizaje de los genes MMR (involucrados en la reparación de mal apareamiento de bases del ADN) resultan ser buenos biomarcadores genéticos en la monitorización de ambas neoplasias (18). En el CP, no se conoce aún un biomarcador genético asociado con la progresión de la enfermedad ni con la resistencia medicamentosa. Solo unos pocos biomarcadores han sido aprobados por la Administración de Alimentos y Medicamentos (FDA) como el PSA (Antígeno prostático específico) en 1994, el Índice de salud prostática (phi) en 2012 y PCA3 (Antígeno de cáncer de próstata 3) en 2012 (19).

La identificación de biomarcadores, en un principio, se da gracias a estudios bioinformáticos aplicando análisis secundarios de genomas del cáncer. Dichos estudios pretenden explicar el fenotipo de la enfermedad considerando la información genómica, y se conocen como estudios de asociación genotipo-fenotipo. Así en CP la asociación del genotipo (variantes o mutaciones) con el fenotipo (grado o etapa tumoral, tiempo de SG, entre otros) han mostrado que algunas variantes genéticas como mutaciones, alteración en el número de copias (ANC), polimorfismo de un solo nucleótido (SNPs) y variación en el patrón de expresión génica están vinculadas a variables clínico-patológicas o de supervivencia. Por ejemplo un estudio en una población finlandesa ha identificado que deleciones en los genes ERBB4, EPHA3 y CSMD1, y una duplicación gen PDZD2 tienen un rol importante en el CP (20). Un estudio en una población rumana ha reportado que algunos SNPs están asociados con el fenotipo del CP incluyendo la variante

rs1512268 en el gen NKX3-1 (Proteína Homeobox Nkx-3.1) vinculado con un tumor de bajo grado y con valores bajos de PSA, mientras que las variantes rs401681 en el gen CLPTM1L (Proteína similar a la proteína transmembrana de labio leporino y paladar hendido 1) y rs11649743 en HNF1B (Factor nuclear de hepatocitos 1-beta) estuvieron asociados con valores altos de PSA, y rs1465618 en THADA (Proteína asociada a adenoma de tiroides) , rs721048 en EHBP1 (Proteína 1 de unión al dominio EH) y rs17021918 en PDLIM5 (Proteína 5 de dominio LIM y PDZ) con tumores menos agresivos (21). Chevill et al. mostraron que la expresión del gen PCA3 se asoció con el puntaje de Gleason (22).

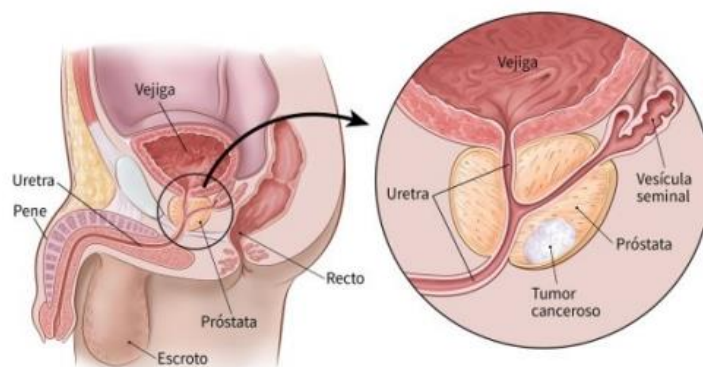
Los desafíos en el desarrollo de biomarcadores de cáncer próstata requieren una validación extensiva, el desarrollo de paneles que sean adecuados para distintos cuadros clínicos, terapias y sean accesibles (19). En dicho proceso hipotéticos marcadores para CP pierden robustez y quedan rezagados en su aplicación clínica. Se han venido desarrollando y evaluando algunas pruebas de diagnóstico y pronóstico de CP en orina, suero y para muestras de tejido tumoral. Sin embargo aún se debe trabajar en la identificación de biomarcadores para desarrollar enfoques personalizados para una detección temprana o para la estratificación del manejo clínico de la enfermedad desde la forma primaria-indolente hacia la etapa agresiva. Así mismo los nuevos potenciales biomarcadores deberían proporcionar información complementaria de las variables clínicas y patológicas del CP ya sea para mejorar la precisión en el diagnóstico, su pronóstico y/o respuesta al tratamiento.

La presente investigación pretende describir el panorama de las variantes genéticas en el CP y determinar la correlación entre el genotipo y las características clínicas-patológicas, analizando una cohorte de sujetos con CP enrolados en diferentes partes del mundo. Los resultados permitirían reconocer potenciales biomarcadores pronósticos del CP que asistan en el monitoreo de pacientes con CP.

## 2.2. Marco Teórico

### 2.2.1. Cáncer de próstata

La próstata es una glándula que forma parte del sistema reproductor de los hombres, está ubicada por encima de la base del pene, debajo de la vejiga y al frente del recto, y su función es elaborar una parte del fluido que constituye el semen que mantiene y transporta al espermatozoide. El CP es una enfermedad que se origina en los tejidos de la próstata, cuando las células de la glándula se dividen y crecen descontroladamente por cambios en los genes que regulan procesos biológicos como el de crecimiento, división celular y reparación del ADN (figura 1).



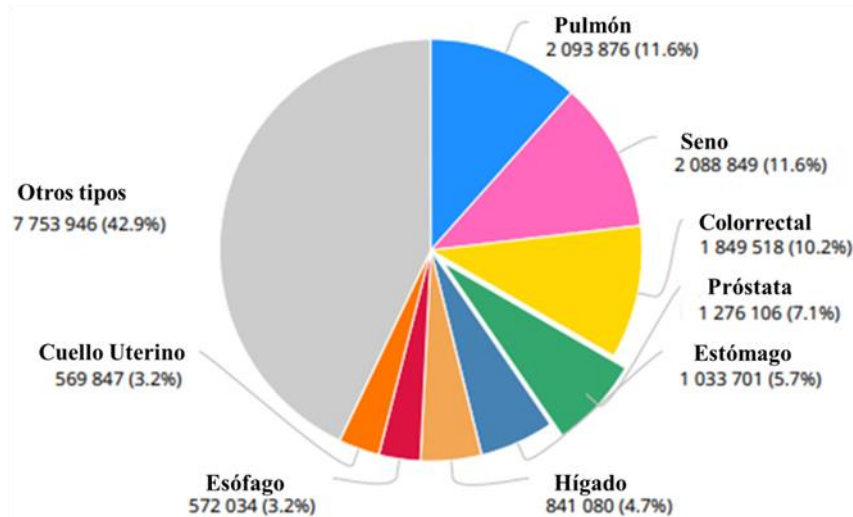
**Figura 1. Representación del crecimiento del tumor canceroso en la glándula prostática. Imagen extraída del portal web de la Sociedad American contra el Cáncer (<https://www.cancer.org>)**

En la etapa temprana del CP no se presentan síntomas claros, sin embargo; a medida que la enfermedad va progresando, los signos y síntomas pueden ser más pronunciados, dentro de los cuales se destacan:

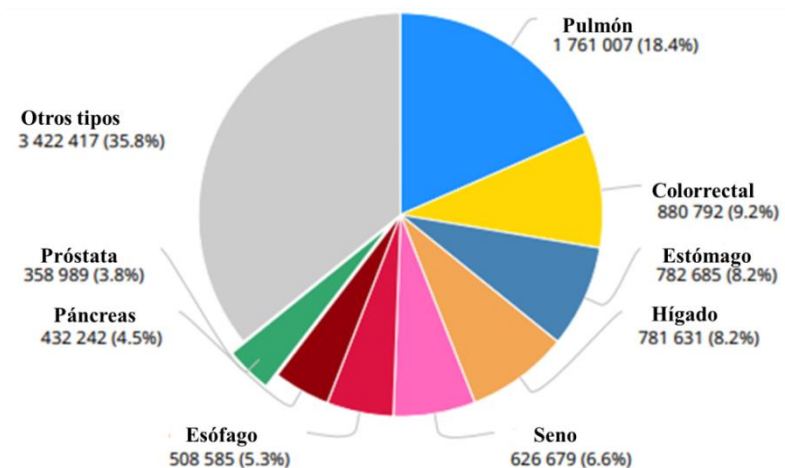
- Dificultad al miccionar.
- Rastros de sangre en fluidos biológicos (orina y/o semen).
- Alteración en la erección.
- Dolor en la espalda o Torax.

#### **2.2.1.1 Estadísticas del cáncer de próstata a nivel mundial y en el Perú.**

El cáncer de próstata es la segunda neoplasia con mayor incidencia a nivel mundial, con 1 276 106 nuevos casos registrados en el 2018 (Figura 2), y el sexto tipo de cáncer con mayor tasa de mortalidad registrando 358 989 decesos (Figura 3).

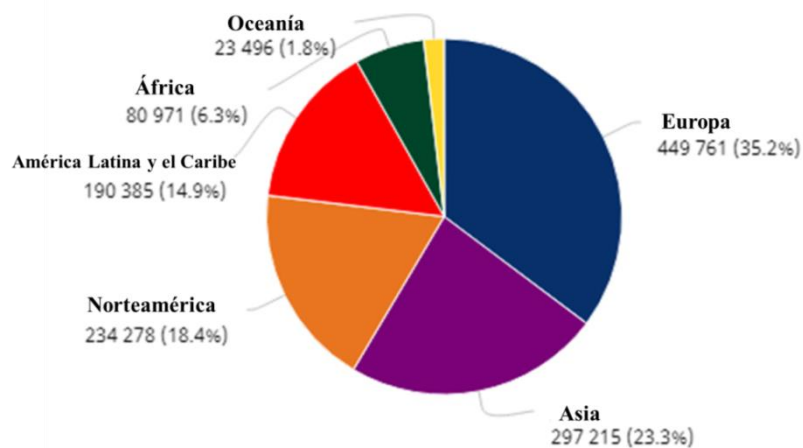


**Figura 2. Casos nuevos de los principales tipos de cáncer con mayor incidencia a nivel mundial. Información extraída del Reporte GLOBOCAN 2018, para ambos sexos y todas las edades (Total de casos 18 078 957). Representación adaptada del sitio web del Observatorio Global del Cáncer (<https://gco.iarc.fr/>).**

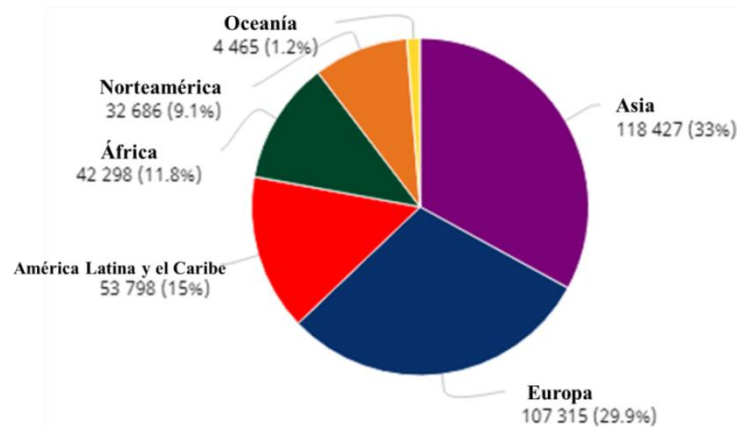


**Figura 3. Muertes causadas por los principales tipos de cáncer a nivel mundial. Información extraída del Reporte GLOBOCAN 2018, para ambos sexos y todas las edades (Total de muertes 9 555 027). Representación adaptada del sitio web del Observatorio Global del Cáncer (<https://gco.iarc.fr/>).**

El continente europeo fue la región que presentó el mayor número de casos nuevos de cáncer de próstata, con 449 761 (35.2%) hombres diagnosticados. En segundo lugar se ubicó la región asiática con el 23.3% de todos los casos reportados a nivel mundial y en tercer lugar Estados Unidos con el 18.4% (Figura 4). El número de muertes por esta neoplasia fue mayor en Asia (33%) y Europa (29.9%), en tercer lugar encontrándose la región Latinoamericana y el Caribe (15%) y en cuarto lugar el continente Africano (11.8%) (Figura 5).

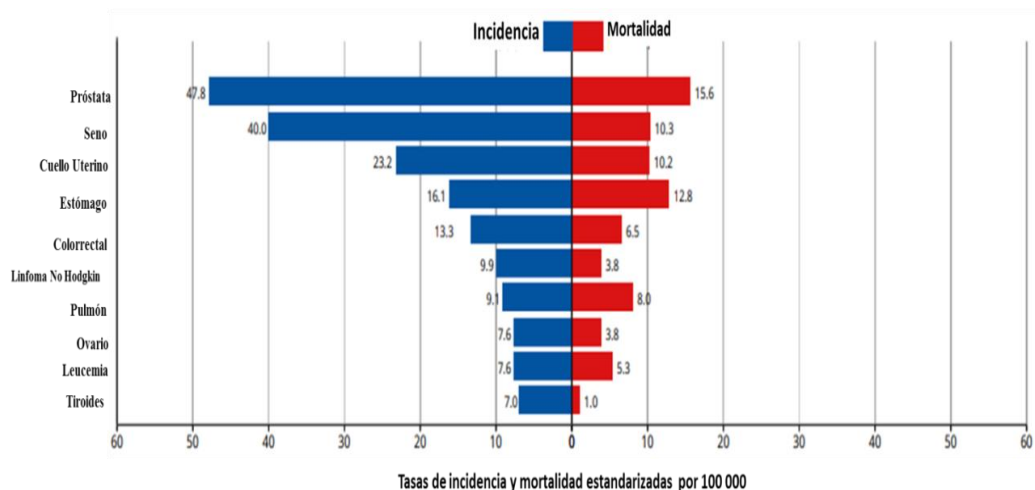


**Figura 4. Casos nuevos de cáncer de próstata estratificado por continente. Información extraída del Reporte GLOBOCAN 2018, para todas las edades (Total de casos nuevos 1 276 106). Representación adaptada del sitio web del Observatorio Global del Cáncer (<https://gco.iarc.fr/>).**



**Figura 5. Muertes causadas por cáncer de próstata estratificado por continente. Información extraída del Reporte GLOBOCAN 2018, para todas las edades (Número total de muertes 358 989). Representación adaptada del sitio web del Observatorio Global del Cáncer (<https://gco.iarc.fr/>).**

En Perú, el CP es la neoplasia con mayor incidencia, registrándose en el 2018 una tasa de 47.8 por cada 100 000 personas. Así también es la neoplasia que más muertes generó, con una tasa de mortalidad de 15.6 por cada 100 000 personas, ubicándose por encima del cáncer de mama, útero, estómago y colorrectal (Figura 6). Los casos nuevos de cáncer de próstata sumaron 7598, mientras que los decesos causados por esta enfermedad alcanzaron la cifra de 2721.



**Figura 6. Tasas de incidencia y mortalidad de los 10 principales tipos de cáncer en el Perú. Información extraída del Reporte GLOBOCAN 2018. Representación adaptada del sitio web del Observatorio Global del Cáncer (<https://gco.iarc.fr/>).**

### **2.2.1.2 Factores de riesgo para cáncer de próstata**

Un factor de riesgo es una característica, circunstancia o exposición que aumenta la probabilidad de que una persona desarrolle alguna enfermedad como el cáncer. Para el CP se han identificado algunos factores que afectan el riesgo de que un hombre desarrolle esta neoplasia, los cuales se describen a continuación:

#### **a) Edad**

La edad es un factor de riesgo del CP tal que la probabilidad de padecer esta neoplasia aumenta rápidamente en hombres a partir de los 50 años en adelante.

#### **b) Raza**

El CP tiene una alta tasa de incidencia en hombres de raza negra y en hombres del Caribe con ascendencia africana, comparado con hombres de otras razas. El CP tiene una tasa de incidencia menor en los hombres asiático-americanos y en los hispanos-latinos que en los hombres blancos.

#### **c) Geografía**

El CP es más frecuente en algunas partes del mundo como en los Estados Unidos y en la región noroeste de Europa, Australia, y en las islas del Caribe en comparación con otras regiones del planeta. El uso más frecuente de pruebas de tamizaje en algunos países desarrollados y el estilo de vida influyen en estas diferencias.

#### **d) Antecedentes familiares**

La mayoría de los casos de CP ocurre en hombres que no tienen antecedentes familiares, es decir, es somático. Alrededor del 20% de los hombres diagnosticados con CP tienen un pariente de primer grado (es decir, hermano,

padre) con este cáncer (23) . En el CP hereditario, si el padre o el hermano de un hombre tienen CP, se duplica el riesgo de que este hombre desarrolle la enfermedad.

#### **e) Factores genéticos hereditarios**

Algunos cambios genéticos heredados pueden incrementar el riesgo de padecer CP, pero sólo representan un 5% de todos los casos (23). Por ejemplo:

-Ciertas mutaciones heredadas de los genes BRCA1 o BRCA2 pueden aumentar el riesgo de padecer CP en algunos hombres.

-Los hombres con el síndrome de Lynch, una condición hereditaria que predispone a desarrollar cáncer colorrectal hereditario sin poliposis, también tienen un mayor riesgo de padecer CP.

#### **2.2.1.3 Sistemas de estadificación del cáncer**

La estadificación de un cáncer es la determinación del grado de diseminación del tumor en el cuerpo de una persona. Esta información clínica es importante porque permite conocer la gravedad de la enfermedad y determinar el mejor tratamiento para el paciente (24). Los tipos de determinación del estadio del cáncer de próstata son 2:

**a) Estadio clínico:** Se basa en los resultados del examen digital del recto (EDR), el análisis del PSA y la puntuación de Gleason. Los resultados de estas pruebas ayudan a determinar si otros exámenes de laboratorio como radiografías, gammagrafías óseas, tomografías computarizadas también son necesarios para una mejor evaluación del tumor. Cabe señalar que el EDR es una prueba en la que el

médico evalúa la próstata mediante el tacto, para lo cual introduce un dedo enguantado y toca la superficie de la glándula para identificar alguna alteración.

**b) Estadio patológico:** Es un parámetro clínico que permite medir el grado de diseminación del tumor en base en la información obtenida después de haber extraído una parte o toda la glándula prostática. La extracción de ganglios linfáticos podría ayudar a medir con mejor precisión en el estadio patológico del tumor.

#### **2.2.1.4 Sistema TNM de determinación del estadio**

El sistema TNM es una herramienta empleada para la determinar el estadio del cáncer. Los profesionales de la salud utilizan los resultados de las pruebas de diagnóstico y de exploración para determinar los siguientes criterios:

##### **a) Tumor “T”**

En primer lugar se debe identificar la dimensión y ubicación del tumor, para ello se emplea una escala utilizando la letra “T” junto a un número que va desde 0 a 4. El estadio patológico a menudo se expresa como “pT”, mientras si se desea expresar el estadio clínico se indica como “cT”. A continuación se detallan las subcategorías para la letra T:

##### **Categoría T clínico “cT”**

- Subcategoría TX: No es posible identificar un tumor primario.
- Subcategoría T0: No se detecta la presencia de algún tumor.

-Subcategoría T1: No es posible identificar el tumor mediante la prueba EDR o exámenes por imágenes, sin embargo la presencia del cáncer podría detectarse por alguna intervención quirúrgica en otra circunstancia.

-Subcategoría T2: El tumor se puede identificar por la prueba EDR y se encuentra localizado solo en la próstata.

-Subcategoría T3: El tumor se ha desarrollado a través de la parte externa glándula prostática.

-Subcategoría T4: El cáncer podría encontrarse fijado en la próstata o estar desarrollándose hacia las estructuras contiguas como la vejiga, pared pélvica, recto, etc.

#### **Categoría T patológico “pT”**

-Subcategoría No se describen las categorías TX, T0 o T1 para el estadio patológico del CP

-Subcategoría T2: El tumor se ubica únicamente en la glándula prostática.

-Subcategoría T3: El cáncer se ha desarrollado a través de la próstata, puede afectar uno o ambos lados de la glándula.

-Subcategoría T4: La tumoración puede estar localizada en la próstata o estar desarrollándose hacia zonas contiguas como recto, vejiga, pared pélvica, etc.

#### **b) Categoría Ganglio “N”**

En la estadificación de un cáncer es importante determinar si el tumor se ha extendido o no a otros órganos o ganglios linfáticos. Para ello se emplea la categoría “N” que permite identificar si los ganglios linfáticos se han visto comprometidos en la extensión del tumor. Los ganglios linfáticos próximos a la

próstata se conocen como ganglios linfáticos regionales y aquellos ubicados en otras partes del cuerpo son los ganglios linfáticos distantes. Las subcategorías para la letra “N” son las siguientes:

-Subcategoría NX: No es posible la evaluación de los ganglios linfáticos regionales.

-Subcategoría N0: EL tumor no ha alcanzado a los ganglios linfáticos regionales.

-Subcategoría N1: El tumor se extendió a uno o varios de los ganglios linfáticos regionales.

### **c) Categoría Metástasis (M)**

Para evaluar si el cáncer ha invadido otros órganos se emplea la categoría “M”, que permite catalogar la metástasis de un tumor. Las subcategorías que podría abarcar la metástasis a distancia son las siguientes:

-Subcategoría MX: No es posible estimar la metástasis del tumor.

-Subcategoría M0: El cáncer no ha realizado metástasis.

-Subcategoría M1: El tumor ha hecho metástasis.

### **2.2.1.5 Escala de Gleason**

La escala de Gleason es un sistema de clasificación de la agresividad del CP, que permite, en base a las características celulares de una muestra del tejido, describir el avance de la enfermedad con la ayuda de una escala numérica. El patólogo evalúa en dos zonas los rasgos histológicos de la biopsia y asigna un número, que va desde 1 al 5, a cada región. Un menor valor número se emplea para células que presentan una buena diferenciación, es decir aquellas que muestran un patrón

de crecimiento similar al de un tejido sano. De caso contrario, si las células muestran poca diferenciación o estas muestran un crecimiento agresivo, se les otorga un valor numérico alto. Se suman los valores dados a las dos regiones y se obtiene el puntaje de Gleason para la muestra, y se interpreta según lo indicado en la tabla 1.

Los valores de la escala de Gleason también se emplean para definir determinadas categorías denominadas “Grupo de grado” que son empleadas en la práctica clínica para clasificar la agresividad del tumor. A continuación se detalla los Grupos de grado en relación a las puntuaciones de Gleason:

- Grupo de grado 1 corresponde a un valor de Gleason de 6 o menos.
- Grupo de grado 2 corresponde a un valor de Gleason  $(3 + 4) = 7$
- Grupo de grado 3 corresponde a un valor de Gleason  $(4 + 3) = 7$
- Grupo Gleason 4 corresponde a un valor de Gleason de 8
- Grupo Gleason 5 corresponde a un valor de Gleason de 9 o 10

**Tabla 1. Interpretación de los diferentes puntajes en la Escala de Gleason**

Escala de Gleason	Interpretación
Gleason X	No se puede establecer un valor de Gleason para la muestra analizada.
Gleason 6 o menor	El patrón de crecimiento celular es regular y las células se encuentran bien diferenciadas. Indicio de un cáncer poco agresivo o de bajo grado
Gleason 7	Las células presentan un intermedio grado de diferenciación, teniendo características similares a células normales. El tumor puede presentar una agresividad media.
Gleason 8, 9 o 10	En la muestra se visualiza células escasamente o nada diferenciadas, siendo indicio de un tumor agresivo o de alto grado.

### **2.2.1.6 PSA**

La prueba de PSA es un test que mide en sangre la concentración del antígeno prostático específico, una proteína producida por las células de la glándula prostática, que sirve para evaluar la presencia y/o monitorización de un CP. Cuando el valor de PSA es alto, mayor a 4,0 ng/ml, el médico tratante puede solicitar exámenes auxiliares para diagnosticar con mayor certeza un CP. Sin embargo, el nivel de PSA en la sangre puede incrementarse a causa de otras condiciones benignas como una prostatitis o una hiperplasia benigna de la próstata (BPH). La mayoría de los médicos considera un valor de 4,0 ng/ml o menos como valor referencial normal del PSA. Aunque la prueba de PSA sigue siendo el test más ampliamente usado para la detección y monitorización del CP, no es aún una herramienta médica eficiente ya que su especificidad es baja, donde un paciente que presente un valor bajo de PSA puede tener CP mientras otro que tenga un valor alto no necesariamente tenga un tumor en desarrollo.

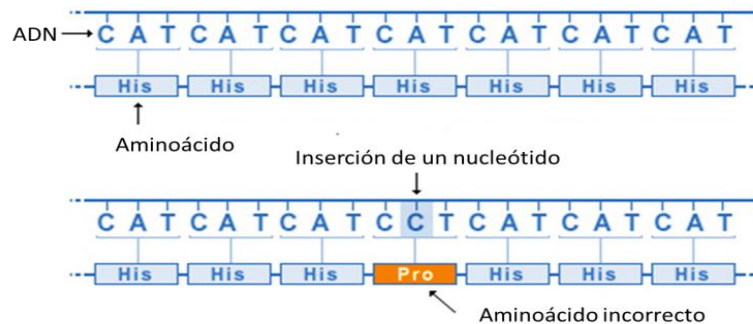
### **2.2.2. Alteraciones genéticas y cáncer**

El cáncer es un padecimiento que tiene un origen genético, donde alteraciones en los genes encargados de la regulación de los procesos celulares hacen que las células alteren su ciclo celular y se dividan y crezcan de forma no controlada.

#### **2.2.2.1 Mutaciones genéticas**

Se entiende por mutación genética a las modificaciones que producen cambios en la secuencia nucleotídica del ADN (Figura 7). Las mutaciones en el ADN pueden, en algunos casos, desencadenar en la sustitución de aminoácidos en las proteínas traducidas, las cuales pueden o no repercutir en su función. A diferencia de una

mutación génica, que ocurre dentro de un gen, la mutación genética puede ocurrir en cualquier región del ADN(25).



**Figura 7. Mutaciones genéticas. Imagen adaptada de la página web del NHI (Institutos Nacionales de salud, <https://ghr.nlm.nih.gov/primer/mutationsanddisorders/possiblemutations>)**

Algunos tipos de mutaciones genéticas:

**-Mutación silenciosa o sinónima:** Ocurre cuando no hay un cambio en la secuencia de aminoácidos de la cadena polipeptídica. Este tipo de mutaciones no generan modificaciones resaltantes en la estructura y/o función de la proteína resultante.

**-Mutación no sinónima:** Es aquella en la que los cambios en los nucleótidos provocan sustitución de un aminoácido por otro, pudiendo ocasionar modificaciones en la estructura o función en la proteína.

**-Mutaciones de corrimiento del marco de lectura:** En este tipo de mutaciones se agregan o se retiran pares de nucleótidos alterándose la longitud de la cadena polipeptídica.

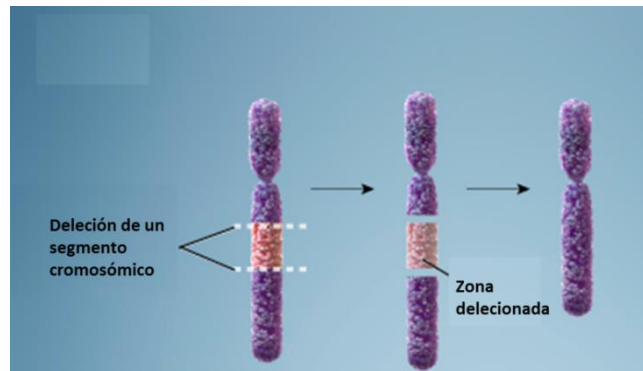
**-Mutación trunca:** Es un cambio en el ADN que genera un codón de STOP o de finalización de la traducción, es decir; hace que una proteína termine su traducción antes de lo esperado.

### **2.2.2.2 Alteración del número de copias y cáncer**

Las alteraciones en el número de copias (ANC) o VNC (variación en el número de copias) son un tipo de variantes estructurales que aparecen en regiones de ADN con tamaños que varían desde unos pocos cientos de nucleótidos a cromosomas completos y que implican un cambio en el número de copias respecto a un genoma de referencia (Figura 8). Alteraciones como inserciones, deleciones, duplicaciones y variantes complejas de múltiples sitios se han identificado tanto en humanos como otros mamíferos (26,27). La inestabilidad genómica en células que proliferan descontroladamente puede ser consecuencia de cambios en la estructura, composición y/o número de cromosomas. Así, una ANC puede consistir desde una variación simple como la duplicación en tándem hasta constituir ganancias o pérdidas complejas de secuencias homólogas en varios lugares en el genoma (28).

Las ANCs pueden afectar a regiones importantes del genoma, por ello resulta importante comprender sus mecanismos de acción para conocer cómo contribuyen al desarrollo de enfermedades como el cáncer. Las alteraciones del número de copias somáticas están casi siempre presentes en tumores malignos y alteran gran parte del genoma más que cualquier otro tipo de alteración genética somática (29,30). El CP tiene tasas relativamente altas de ANC en comparación con mutaciones puntuales (31). Los patrones o agrupamientos de la ANC se han asociado con un alto puntaje en la escala de Gleason (mayores a 8) y a la recurrencia de la enfermedad (32–34). También se ha determinado, analizando las cohortes de TCGA, que las ANCs se asocian significativamente con la sobrevida

libre de enfermedad y la supervivencia general en neoplasias como el cáncer primario de mama, endometrio, células claras renales, tiroides y colorrectal (31).

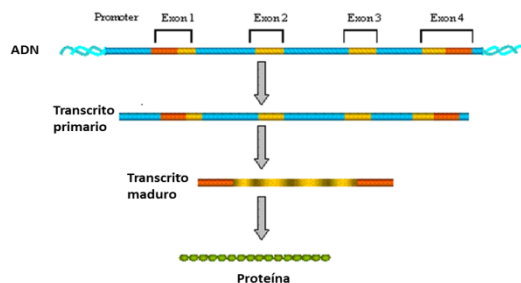


**Figura 8. Deleción de un segmento de ADN. Imagen adaptada de la página web del NHI (Institutos Nacionales de salud, <https://ghr.nlm.nih.gov/primer/mutationsanddisorders/structuralcha>**

### 2.2.2.3 Variación en la expresión génica y cáncer

La expresión génica es el mecanismo por medio del cual los seres vivos decodifican la información contenida en los ácidos nucleicos y la convierten en proteínas, las cuales son necesarias para realizar diversas funciones (Figura 9) (25). A pesar de que un individuo posee células que contienen los mismos genes, estos no se expresan de la misma forma en todas las células, incluso algunos genes pueden no expresarse. La caracterización de los mecanismos detrás de estas variaciones, a nivel de la expresión génica, que intervienen en el desarrollo y patologías, podría abrir un camino hacia el desarrollo de nuevos enfoques terapéuticos. En el ámbito clínico, el uso de los perfiles de expresión génica aplicados al estudio del cáncer ha servido para identificar diversos subtipos moleculares y para predecir supervivencia y recurrencia de la enfermedad (35,36). Para el CP, se han estudiado las variaciones

en la expresión de genes candidatos a marcadores como PCA3, AR, PTEN, TGFβ1, AMACR, cromogranina A, IL-6, IGFBP, PSCA así como de genes reguladores del ciclo celular, apoptosis, adhesión celular y angiogénesis (37,38). El análisis de PCA3, basado en la medición del ARNm en orina, ha relacionado la expresión de este gen tanto con la agresividad del CP como con los estadios clínicos, aunque con resultados controvertidos. En un trabajo conducido por Chevill et al. se describió que la expresión de PCA3 estuvo significativamente asociado con el puntaje de Gleason (22). En tanto que en un estudio experimental realizado por Durand et al. se encontró que la expresión de PCA3 fue marcadamente diferente tanto entre los pacientes con puntuación de Gleason 6 comparado a los que tenían Gleason  $\geq 7$  como entre los grupos de estadios patológicos pT0/T2 comparado con pT3/T4 (39).



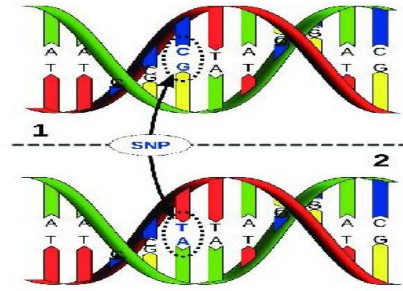
**Figura 9. Flujo de la expresión génica del ADN. Imagen adaptada del portal web NCBI (<https://www.ncbi.nlm.nih.gov/probe/docs/applexpression/>).**

#### **2.2.2.4 Polimorfismo de nucleótido único (SNP) y el cáncer**

Un polimorfismo de un solo nucleótido o SNP (Single Nucleotide Polymorphism) es una variante genética común que compromete el cambio una sola base nucleotídica por otra en la secuencia del ADN (Figura 10) (40). Estas

variantes genéticas ocurren al menos en un 1% de la población a diferencia de las mutaciones puntuales que ocurren en menos del 1% de la población (20). Los SNPs son uno de los tipos más comunes de variaciones genéticas en el genoma humano y pueden encontrarse en diversos tipos de genes como aquellos que regulan la reparación del ADN, el ciclo celular, el metabolismo y la inmunidad. Algunos SNPs pueden estar asociados con un rasgo observable de algún tipo determinado de cáncer y que varía dependiendo de la ubicación del SNP. Estos polimorfismos pueden ubicarse dentro de una región promotora alterando significativamente su función, así como la capacidad de uniones de factores de transcripción al ADN. Otros efectos pueden incurrir en cambios epigenéticos como la modificación de histonas y metilaciones en las cadenas nucleotídicas (41–43).

Por otro lado, los SNPs que se encuentran en intrones pueden alterar el rol de los lncRNAs (44). Se han identificado algunos genes como RNaseL (ubicado en el locus HPC1, cromosoma 1), que codifica para una enzima que participa en la respuesta inmune a infecciones virales, en la inducción de apoptosis, en la regulación del ciclo celular y la diferenciación celular, conteniendo SNPs que podrían ser candidatos a marcadores de riesgo para CP; así como el gen ELAC2 (cromosoma 17) (45). Profundizar en la caracterización de SNPs de riesgo a CP podría contribuir a la estratificación de los pacientes y a la evaluación de las respuestas a tratamientos individuales.



**Figura 10. Polimorfismo de un solo nucleótido. Se representa un SNP en la secuencia del ADN. Imagen adaptada de Ayday et al, 2013 (40).**

### 2.2.3. Proyectos Genomas

En 2005, El Atlas del Genoma del Cáncer (TCGA), creado por El Instituto Nacional de Salud de los Estados Unidos (NIH) fue presentado como uno de los proyectos más importantes para el estudio integral de la genómica del cáncer (46). El proyecto TCGA ha generado un atlas completo de perfiles genómicos de 33 tipos de cáncer, incluido el CP. La información de TCGA es pública, su procesamiento y análisis mediante herramientas bioinformáticas hoy en día ha hecho posible la investigación de biomarcadores potenciales del cáncer que ayuden a mejorar los métodos de diagnóstico y estrategias de manejo clínico y prevención (9,10). Un grupo de investigación del Instituto Johns Hopkins al analizar los datos TCGA para cáncer de ovario, demostraron la correlación de la amplificación de *NACC1* con la recurrencia del tumor (12).

El proyecto TCGA también ha permitido de identificar dianas para el tratamiento del cáncer de pulmón de células escamosas basado en perfiles genómicos y epigenómicos de 180 casos de este tipo de neoplasia (11). A través de TCGA también se ha podido explorar el paisaje de las alteraciones genómicas en el CP y otras neoplasias genitourinarias como mutaciones somáticas y variación en el número de copias (17). También ha contribuido a evaluar la expresión de genes

vinculados en la oncogénesis del CP, por ejemplo se ha encontrado que la expresión de KIF18A está asociada con un aumento del estadio tumoral y la proliferación celular en pacientes de la cohorte TCGA (47).

### **2.3. Pregunta de Investigación**

Habiendo reconocido la carencia de información en torno a la asociación entre el genotipo del CP y el fenotipo manifestado por pacientes con cáncer de próstata, se ha planteado la siguiente pregunta de investigación: ¿Cómo es el panorama de las alteraciones genéticas del CP y si dichas variaciones están vinculadas a ciertas características clínico-patológicas de la enfermedad?

### **2.4. Justificación de la Investigación**

El CP es un problema de salud pública a nivel mundial y en el Perú es la neoplasia con la mayor tasa de incidencia y la principal causante de mortalidad por cáncer en la población masculina (1). La acumulación de alteraciones genéticas cumple un rol importante en el origen y progresión del cáncer y las mismas pueden conocerse mediante un análisis genómico (48). La heterogeneidad genética del CP, manifestada a través del panorama mutacional, conduce a variabilidad en el desarrollo de la enfermedad y las respuestas a tratamientos. Actualmente los biomarcadores de seguimiento del CP son insuficientes para estimar el pronóstico de la enfermedad.

La bioinformática junto con las técnicas ómicas como la genómica han proporcionado información significativa sobre la biología del cáncer a través de proyectos de genomas del cáncer como TCGA. La información genómica y

transcriptómica de los individuos participantes de TCGA son de dominio público y contienen valiosa información sobre el componente molecular del cáncer. Así mismo, los datos del fenotipo de los individuos participantes están disponibles de manera anonimizada, incluyendo el pronóstico de la enfermedad, edad, raza, peso, entre otras variables.

La cuantiosa información genética y fenotípica revelada por proyectos como TCGA hace posible establecer un enfoque dirigido a estudiar el panorama mutacional e identificar las alteraciones genómicas que influyan en las diversas manifestaciones patológicas del CP. En ese sentido, la importancia del presente trabajo se centra en el estudio de la información genómica y clínica disponible para identificar variantes genéticas en el cáncer de próstata que puedan perfilarse como potenciales marcadores pronósticos de la enfermedad.

### **3. OBJETIVOS**

#### **3.1. Objetivo General**

Identificar las variantes genéticas asociadas a características clínico-patológicas del CP con el uso herramientas bioinformáticas para el minado de bases de datos genómicos y clínicos de estudios albergados en el portal de acceso público cBioPortal, a través de un estudio computacional del perfil molecular de la neoplasia.

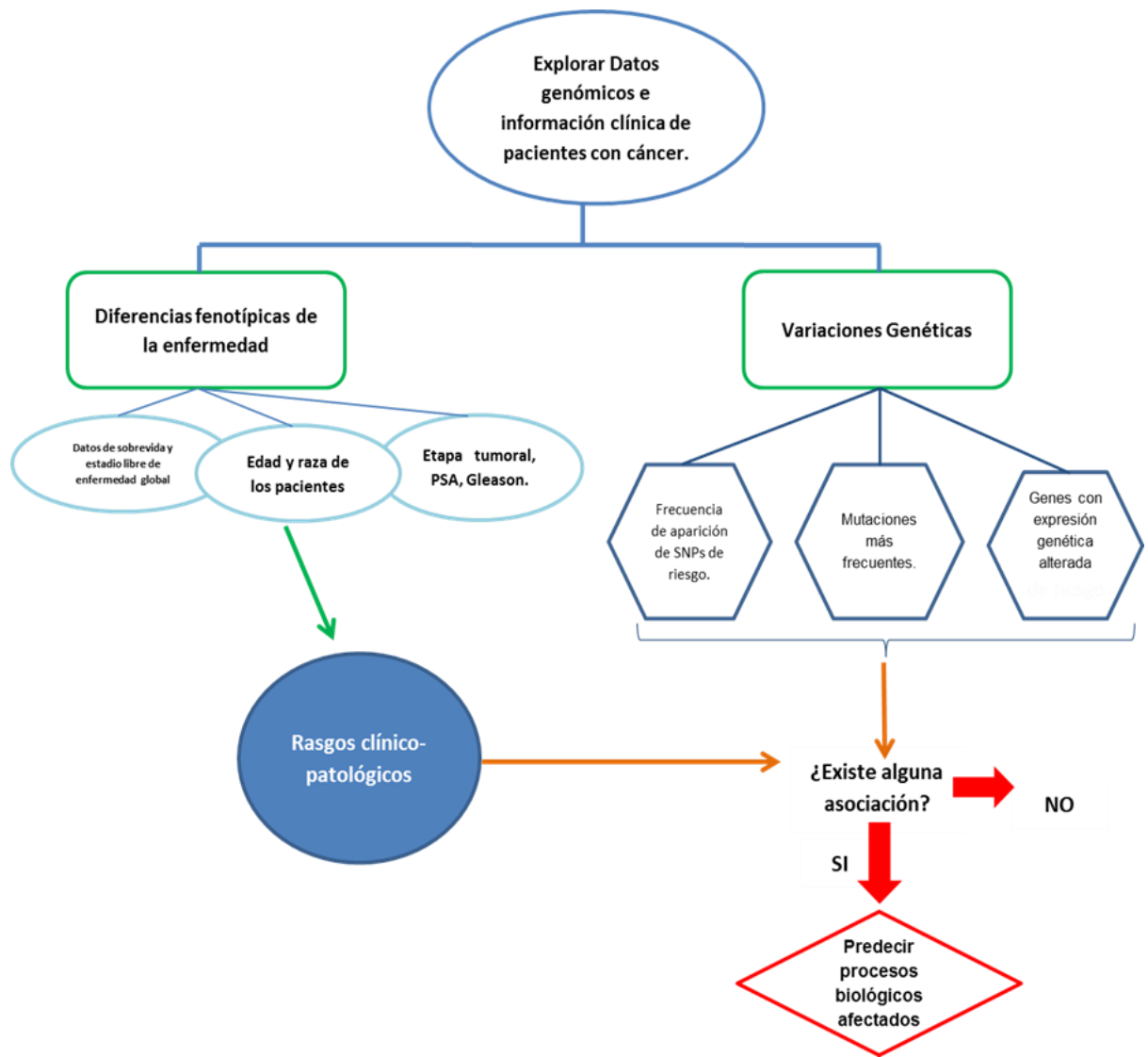
#### **3.2. Objetivos específicos**

1. Definir el panorama de genes significativamente mutados en el CP.
2. Identificar los genes con alteración en el número de copias (amplificados y deletados) en CP.
3. Reconocer genes con expresión diferencial (sobreexpresados y subexpresados) en CP.
4. Evaluar la asociación de alteraciones genéticas con caracteres fenotípicos de pacientes con CP
5. Estimar el daño en la estructura y función de las proteínas de los principales genes significativamente mutados en CP y la afectación de los procesos biológicos involucrados.
6. Determinar las frecuencias de mutaciones conductoras del CP.
7. Revisar las frecuencias de polimorfismos de un solo nucleótido (SNPs) de riesgo a CP.

## **4. Materiales y Métodos**

### **4.1. Diseño del estudio**

El diseño del presente estudio fue del tipo descriptivo, observacional retrospectivo, donde se utilizaron recursos e instrumentos bioinformáticos para la exploración y análisis de datos genómicos y fenotípicos de pacientes con CP (Figura 11). La información estuvo disponible en portales de libre acceso provenientes de proyectos de cáncer a gran escala. A través de un análisis computacional, se exploró el perfil genómico del CP y se evaluaron las asociaciones de variantes genéticas con atributos clínicos de la enfermedad.



**Figura 11. Flujograma del trabajo de investigación**

#### **4.2. Fuente y características de los datos genómicos y clínicos**

**-Fuente de base de datos :** Información de hombres mayores de 18 años diagnosticados con CP cuyos datos clínicos y genómicos están almacenados y son de acceso libre en portales de genomas de cáncer tales como cBioPortal, Firebrowse y Xena. Estos recursos permiten la exploración de datos de estudios en cáncer de gran envergadura. La información es exportada desde “The Cancer Genome Atlas” (TCGA), proyecto del esfuerzo conjunto del Instituto Nacional del Cáncer y el Instituto Nacional de Investigación del Genoma Humano; que tiene como objetivo caracterizar el perfil genómico de diferentes tipos de neoplasias. Estos portales también seleccionan conjuntos de datos de la literatura, además de solicitar directamente a los investigadores data adicional como atributos clínicos cuando se dispongan de ellos. Además, estos recursos emplean algoritmos para homogenizar la determinación de variantes genéticas como las mutaciones a fin de identificarlas de forma consistente en todos los estudios.

La muestra fue de 3312 que representa la sumatoria de pacientes enrolados en los proyectos sobre CP albergados en cBioPortal. Cabe señalar que para efectos de la identificación de GSM, genes con expresión alterada y estudio de asociación de genotipo-fenotipo se empleó solamente la cohorte TCGA debido a que esta población presento la información completa de análisis mutacional y expresión genética, además de contar con datos consistentes de la características clínico-patológicas de los individuos enrolados. Se debe señalar que el proyecto TCGA proporciona datos biológicos a través de herramientas de análisis genómicos como

la secuencia del genoma a gran escala, y hasta el momento ha evaluado 27 tipos de tumores diferentes.

**-Criterios de inclusión:**

- a) Información de hombres diagnosticados con CP, en cualquier estadio de la enfermedad.
- b) Información de participantes cuyos datos genómicos e información fenotípica (demográfica y clínica) esté disponible.

**-Criterios de exclusión:**

- a) Información de hombres que no tengan Cpp.
- b) Información de hombres que presenten CP neuroendocrino.

### **4.3. Operacionalización de variables**

En el presente estudio se evaluaron variables genómicas y clínicas en pacientes con CP (tabla 2). Dentro de las variables genómicas se estudiaron las mutaciones génicas, alteración en el número de copias y expresión génica. Las variables demográficas evaluadas fueron edad, raza y etnia, mientras que las variables clínicas consistieron de PSA, escala de Gleason (EG), estadio T y N, sobrevida global (SG) y sobrevida libre de enfermedad (SLE). La operacionalización de las variables se muestra en la tabla 2, donde se detalla la definición de las mismas, su naturaleza y la medida empleada en el estudio.

**Tabla 2. Operacionalización de las variables de estudio.**

Variable	Definición	Tipo de variable según su naturaleza	Tipo de variable según su relación	Escala de medición	Indicador	Definición operacional
Presencia de mutaciones puntuales	La existencia de una variación de nucleótido en un gen.	Cualitativa	Independiente	Nominal	0,1	0:Ausente 1: Presente
Alteración en el número de copias	Ganancia o pérdida en el número de copias de un gen determinado.	Cualitativa	Independiente	Nominal	0,1,2	0:Sin alteración 1:Amplificado 2:Delecionado
Alteraciones en la expresión génica	Variación en la expresión de un gen determinado	Cualitativa	Independiente	Nominal	0,1	0: Subexpresión 1: Sobreexpresión
Frecuencia de aparición de SNPs de riesgo a CP	Presencia de un SNP en la secuencia del ADN	Cuantitativa	Independiente	Razón	Número de SNP	Número
Presencia de CP	Condición en la cual un paciente es detectado con CP	Cualitativa	Dependiente	Nominal	0,1	0: Ausente 1: Presente
Características clínico-patológicas del cáncer	Estadaje del tumor: etapa en la cual se encuentra el desarrollo del tumor.	Cualitativa	Dependiente	Ordinal	Sistema TNM	T: Extensión del tumor principal N: el cáncer se propagó a los ganglios linfáticos cercanos M: el cáncer se ha propagado a otras partes del cuerpo
	Escala Gleason: grado de agresividad de un CP, teniendo en cuenta las características que presentan las células en una biopsia.	Cuantitativa	Dependiente	Numeral	Escala de Gleason	2 a 6: Cáncer con escasa agresividad, crecimiento lento y por lo tanto de mejor pronóstico. 7: Cáncer con agresividad intermedia. 8: Cáncer de alta agresividad, y peor pronóstico.
Sobrevida global	Tiempo en el cual el paciente sigue vivo desde que fue diagnosticado con la enfermedad o inició su tratamiento.	Cuantitativa	Dependiente	Razón	Tiempo	Meses

Sobrevivida libre de enfermedad	Tiempo que se mide después del tratamiento, en el que no se detecta algún signo del cáncer.	Cuantitativa	Dependiente	Razón	Número	Meses
Nivel de PSA	Nivel del antígeno prostático específico (PSA) que se obtiene de un paciente	Cuantitativa	Co-variable	Razón	Número	ng/ml
Raza	La etnia a la que pertenece el paciente.	Cualitativa	Co-variable	Nominal	1,2,3,4	1:Blanco 2:Asiático 3: Negro o africano americano 4: Latinoamericano
Edad al momento del diagnóstico	Años de vida que presenta el paciente al momento del diagnóstico	Cuantitativa	Co-variable	Razón	Edad	Años

#### 4.4. Instrumentos

En el presente trabajo de investigación se utilizaron los recursos bioinformáticos y bases de datos descritos en la Tabla 3 para la recolección y análisis de la información.

**Tabla 3. Recursos web empleados en la recolección de información del estudio.**

Recurso	Descripción y utilidad	Sitio	Fecha de consulta
cBIOPORTAL	Es un recurso informático de acceso abierto que permite visualizar, analizar y descargar conjuntos de datos de genómica del cáncer a gran escala.	<a href="http://www.cBioPortal.org">http://www.cBioPortal.org</a>	Abril- Agosto 2018
GWAS	Es una base de datos de asociaciones de SNPs con rasgos fenotípicos, que puede integrarse fácilmente con otros recursos.	<a href="https://www.ebi.ac.uk/gwas/">https://www.ebi.ac.uk/gwas/</a>	Junio 2018
INTOGEN	Es una herramienta que permite obtener datos de mutación tumoral, métodos	<a href="https://www.intogen.org/search">https://www.intogen.org/search</a>	Mayo- Junio 2018

---

	de identificación de genes/mutaciones conductoras, mapeo de mutaciones entre otras cosas.		
DriverDB	Es una base de datos que proporciona datos genómicos relacionados con cáncer, además de algoritmos bioinformáticos dedicados a la identificación de genes y mutaciones conductoras.	Disponible: <a href="http://driverdb.tms.cmu.edu.tw/driverdbv2/">http://driverdb.tms.cmu.edu.tw/driverdbv2/</a>	Junio 2018
COSMIC	Es un catálogo de mutaciones somáticas que permite estudiar el impacto de estas en el cáncer humano.	<a href="http://cancer.sanger.ac.uk/cosmic">http://cancer.sanger.ac.uk/cosmic</a>	Julio 2018
BIOMUTA	Es una colección de variaciones de un solo nucleótido en el ADN y su efecto funcional en la proteína traducida.	<a href="https://hive.biochemistry.gwu.edu/biomuta">https://hive.biochemistry.gwu.edu/biomuta</a>	Julio 2018
FIREBROWSE	Es un recurso que permite de explorar datos genómicos y transcriptómicos de cáncer. Contiene datos de más de 80 mil muestras de cáncer donadas por más de 11,000 pacientes, que abarca 38 cohortes de distintos tipos de cáncer.	<a href="http://firebrowse.org/">http://firebrowse.org/</a>	Agosto 2018
UNIPROT	Es una plataforma en línea con una base de datos de secuencias de proteínas, conteniendo información de su función y estructura. Es resultado de la participación del European Bioinformatics Institute (EMBL-EBI), el Swiss Institute of Bioinformatics (SIB) y Protein Information Resource (PIR).	<a href="http://www.uniprot.org/">http://www.uniprot.org/</a>	Noviembre 2018
Genecodis	Es una plataforma en línea que permite describir la ontología de genes. Con la ayuda de determinados algoritmos se puede identificar anotaciones	<a href="http://genecodis.cnb.csic.es/analysis">http://genecodis.cnb.csic.es/analysis</a>	Noviembre 2018

---

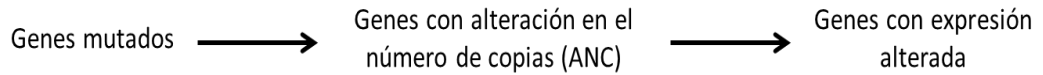
---

	<p>biológicas significativamente asociadas a un grupo de genes en evaluación comparándolos con una lista de referencias.</p>		
STRING	<p>Es una herramienta que permite graficar las interacciones proteína-proteína de un grupo de genes de interés comparándolos con una base de datos de interacciones conocidas.</p>	<p><a href="https://string-db.org/">https://string-db.org/</a></p>	<p>Diciembre 2018</p>
GEPIA	<p>Permite analizar datos de expresión génica de 9,736 tumores y 8,587 controles, de los proyectos TCGA y GTEx, respectivamente.</p>	<p><a href="http://gepia.cancer-pku.cn/">http://gepia.cancer-pku.cn/</a></p>	<p>Enero-Abril 2019</p>
XENA	<p>Es una herramienta de exploración e integración visual para datos genómicos, clínicas y fenotípicos de pacientes en diversos tipos de cáncer.</p>	<p><a href="http://xena.ucsc.edu/">http://xena.ucsc.edu/</a></p>	<p>Enero-Abril 2019</p>
VAPROS	<p>Es una herramienta que permite evaluar el efecto de variaciones en la estructura y función de proteínas a partir de la integración de información en biología estructural con genoma, transcriptoma, interactoma, etc.</p>	<p><a href="http://www.vapros.org/">http://www.vapros.org/</a></p>	<p>Agosto-Setiembre 2019</p>
I-MUTANT	<p>Permite predecir cambios de estabilidad de una proteína causadas por mutaciones puntuales a partir de la secuencia de la proteína.</p>	<p><a href="http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi">http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi</a></p>	<p>Agosto-Setiembre 2019</p>

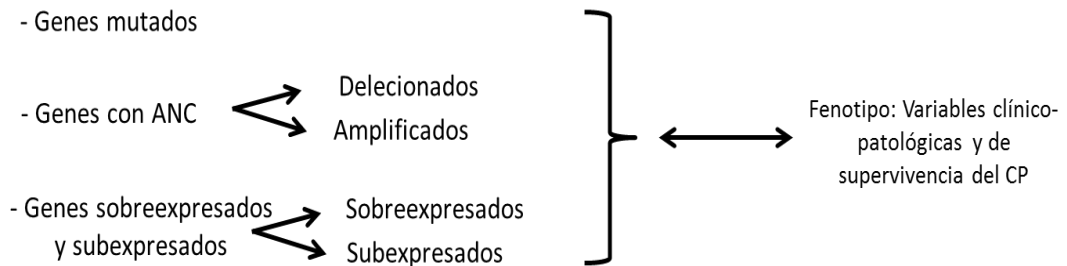
---

#### 4.5. Secuencia de análisis

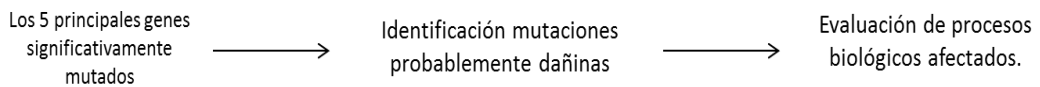
##### 1. Exploración de las alteraciones genéticas en el CP



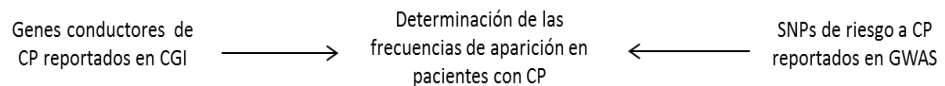
##### 2. Evaluación de la asociación de las alteraciones genéticas en el CP con rasgos fenotípicos.



##### 3. Estimación del daño en estructura y función de las proteínas de los principales genes significativamente mutados.



##### 4. Análisis complementarios: Determinación de las frecuencias de mutaciones conductoras y de polimorfismos SNPs de riesgo a CP.



#### 4.6. Técnicas y procedimientos de recolección de datos

En la primera etapa del estudio, se consultaron 14 estudios de CP (subtipo histológico adenocarcinoma) en cBioPortal, según se detalla en la Tabla 4. La descripción de la población de estudio fue tabulada en hojas de Excel utilizando el programa Microsoft Excel 2013.

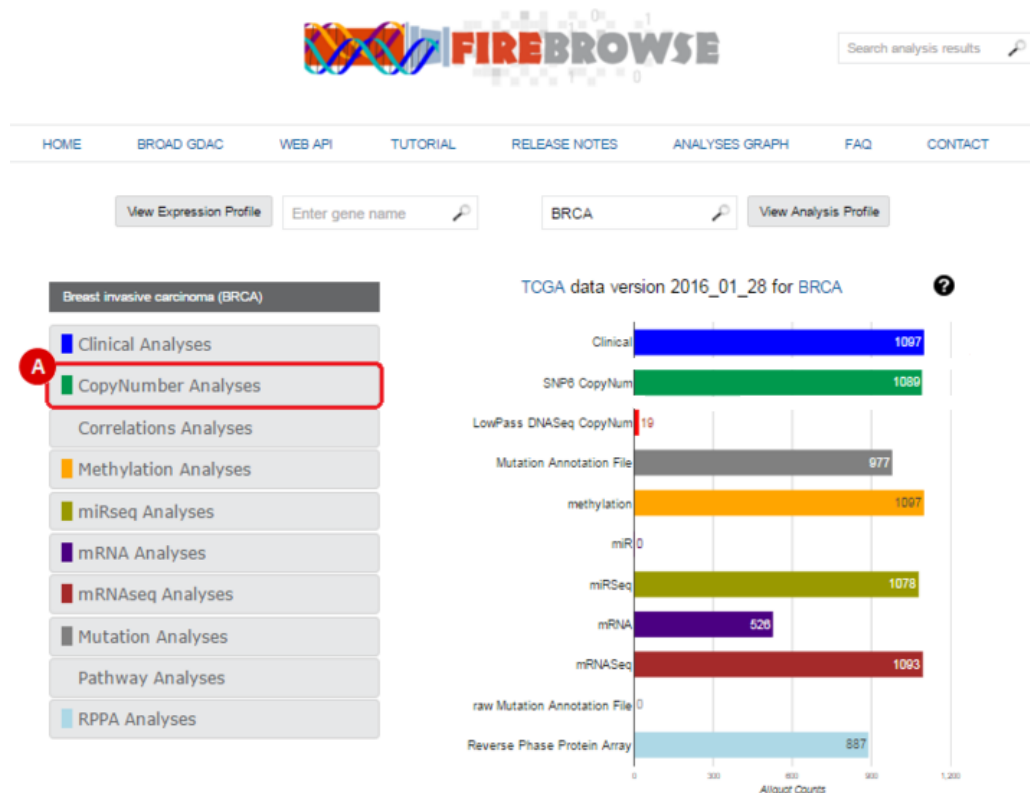
**Tabla 4. Estudios de cáncer de próstata albergados en cBioPortal**

Tipo histológico	Estadio del tumor	Estudio	Número de Pacientes	Número de Muestras		
Adenocarcinoma	Primario	Genomic Hallmarks of Prostate Adenocarcinoma (CPC-GENE, Nature 2017)	477	477		
		MSK-IMPACT Clinical Sequencing Cohort in Prostate Cancer (MSK, JCO Precision Oncology 2017)	451	504		
		Prostate Adenocarcinoma (Broad/Cornell, Cell 2013)	57	57		
		Prostate Adenocarcinoma (Broad/Cornell, Nat Genet 2012)	112	112		
		Prostate Adenocarcinoma (EurUrol, 2017)	65	65		
		Prostate Adenocarcinoma (Fred Hutchinson CRC, Nat Med 2016)	63	176		
		Prostate Adenocarcinoma (MSKCC, Cancer Cell 2010)	216	216		
		Prostate Adenocarcinoma (MSKCC/DFCI, Nature Genetics 2018)	1013	1013		
		Prostate Adenocarcinoma (TCGA, Provisional)	498	499		
		Prostate Adenocarcinoma CNA study (MSKCC, PNAS 2014)	104	104		
		Prostate Adenocarcinoma Organoids (MSKCC, Cell 2014)	7	12		
		Metastásico		Metastatic Prostate Cancer, SU2C/PCF Dream Team (Robinson et al., Cell 2015)	150	150
				Metastatic Prostate Adenocarcinoma (MCTP, Nature 2012)	59	61

Con el fin de tener una visión panorámica de las alteraciones genómicas o “paisaje mutacional” en CP, se investigaron tres tipos de modificaciones: mutaciones puntuales, ANCs y cambios en la expresión génica.

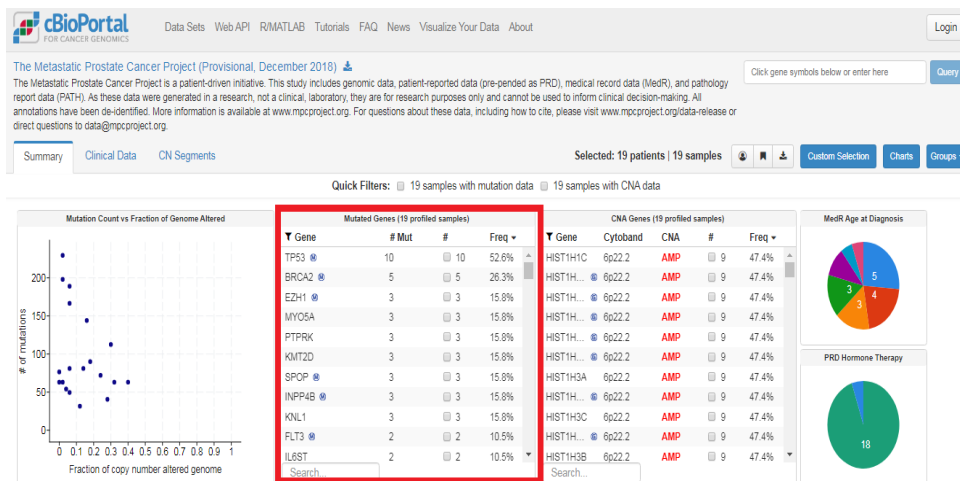
#### **4.6.1. Determinación del panorama mutacional en el CP.**

Se consultó Firebrowse para obtener los genes más mutados en Cpp. Esta base de datos contiene la información de la cohorte TCGA (Provisional, n=498 pacientes). Para correr el análisis, una vez en Firebrowse, se seleccionó la cohorte “Prostate Adenocarcinoma (PRAD)”, luego la opción Mutation Analyses y finalmente la alternativa MutSig2CV, cuyo algoritmo mide la carga de mutación y determina el p-valor para observar la cantidad de mutaciones no silenciosas en un determinado gen (Figura 12). La herramienta explora el estado mutacional de todos los genes contenido en el genoma humano, sin restricción alguna. En la sección de resultados, la información de interés se encontró bajo el término “Significantly Mutated Genes”, la misma que se descargó, mostrando un punto de corte de un q-valor  $\leq 0.1$  preestablecido en la plataforma (q-valor, tasa de descubrimiento falso).



**Figura 12. Interfaz de Firebrowse. A) El recuadro rojo muestra la lista de opciones para analizar alteraciones genéticas en base a la cohorte que se haya seleccionado. Gráfica tomada de la página web de Firebrowse (<http://firebrowse.org/>).**

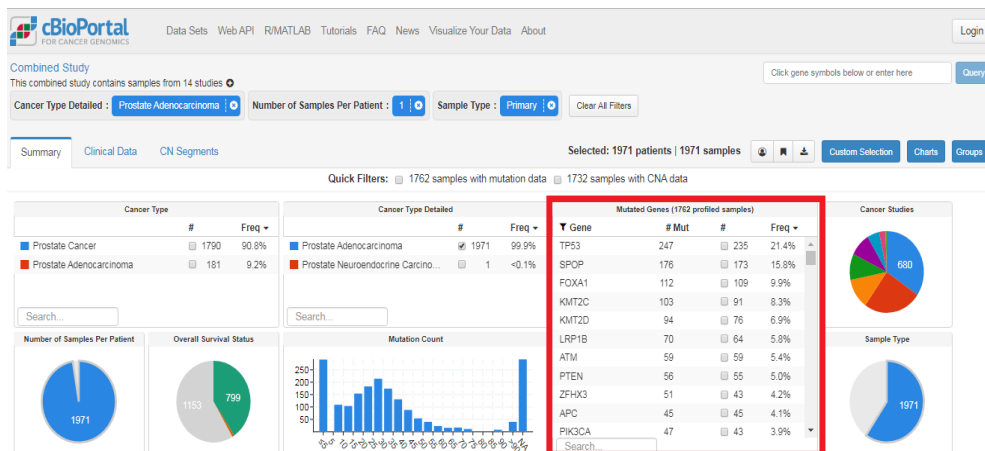
Para la identificación de GSMs en CPM se empleó cBioportal. Primero se seleccionó “Prostate” como órgano de análisis, lo cual generó una lista de 16 estudios. A continuación, se seleccionó el estudio “The Metastatic Prostate Cancer Project” (n=19 pacientes). Se activó la opción “View Summary” del panel superior para visualizar la información a detalle. Los GSMs se ubicaron en una tabla que contenía el nombre del gen, número de mutaciones y la frecuencia de aparición (Figura 13). Se descargó la información de la tabla resultante obtenida para una interpretación posterior.



**Figura 13. Panel de resumen de estudio “The Metastatic Prostate Cancer Project (Provisional, December 2018)” de CPM de cBioPortal. En el recuadro rojo se muestra la lista de GSMs para CPM. Gráfica tomada de la página web de cBioPortal (<https://www.cbioportal.org/>).**

Con el fin de indagar si los GSMs presentaban altas frecuencias de aparición, se deseó identificar los genes frecuentemente mutados (GFM) en estadio primario y metastásico del CP, para ello se empleó cBioPortal. Se seleccionó “Prostate” como órgano de análisis, lo cual generó una lista de 16 estudios. A continuación, se seleccionaron aquellos estudios que contenían información de mutaciones por gen, aquello se verificó seleccionando individualmente cada estudio y verificando que en la parte inferior se activara la opción “Mutations alterations”. Se activó la opción “View Summary” del panel superior para visualizar la información integrada y resumida de los estudios seleccionados. Se mostraron múltiples opciones de filtrado de la información bajo los siguientes criterios: subtipo histológico, tipo de muestra (tanto tumor primario como metastásico) y número de muestra por pacientes. Se aplicaron dichos criterios de filtro y se obtuvo una tabla con los GFMs junto con su frecuencia de aparición (%) en la cohorte de estudio

(Figura 14). Se descargó la información de la tabla resultante obtenida para una interpretación posterior.

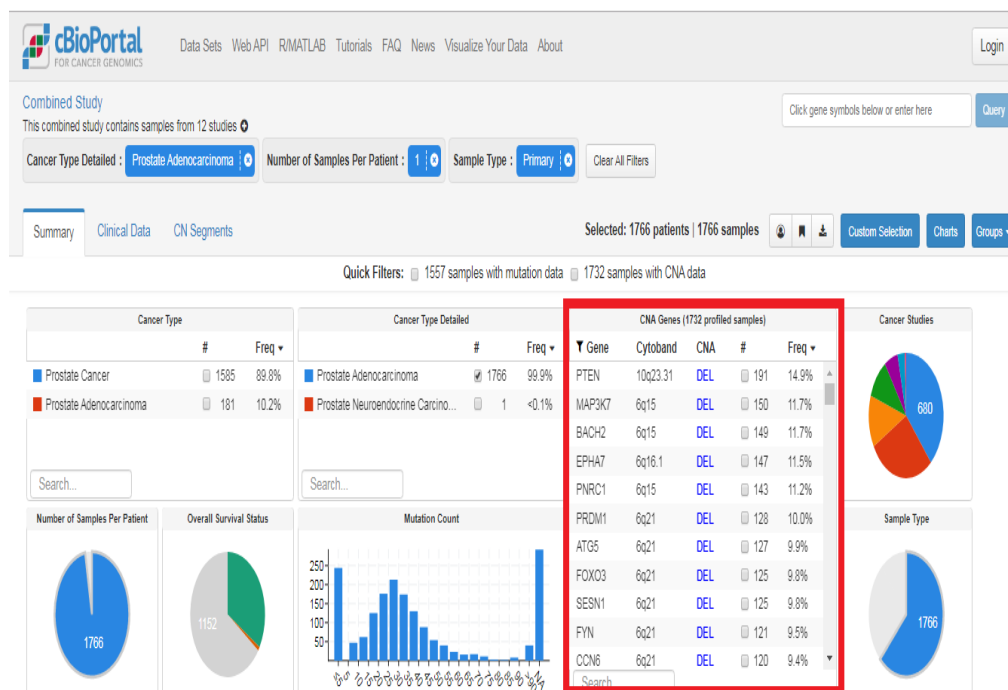


**Figura 14. Panel de resumen de estudios de CP en cBioPortal. En el recuadro rojo se muestra la lista de GFM para los estudios seleccionados de CP. Gráfica tomada de la página web de cBioPortal (<https://www.cbioportal.org/>).**

#### 4.6.2. Identificación de genes con alteración en el número de copias (ANC), amplificados y deletados.

Con el objeto de identificar genes con alteración del número de copias (amplificación o deleción), se consultó cBioPortal. Primero se seleccionó “Prostate” como órgano de análisis, lo cual generó una lista de 16 estudios. A continuación, se seleccionaron aquellos estudios que contenían información del número de copias por gen; esto se verificó seleccionando individualmente cada estudio y verificando que en la parte inferior se activara la opción “Putative copy-number alterations”.

Finalmente se activó la opción “View Summary” del panel superior para visualizar la información integrada y resumida de los estudios seleccionados. Esta interfaz contuvo múltiples opciones de filtrado de la información bajo los siguientes criterios: subtipo histológico, tipo de muestra (tanto tumor primario como metastásico) y número de muestra por pacientes. Se aplicaron dichos criterios de filtro y se obtuvo una tabla con los genes que presentaron ANC (amplificación y delección), junto con su frecuencia de aparición (%) en la cohorte de estudio y su localización cromosómica (Figura 15). Se descargó la información de la tabla resultante obtenida para una interpretación posterior.



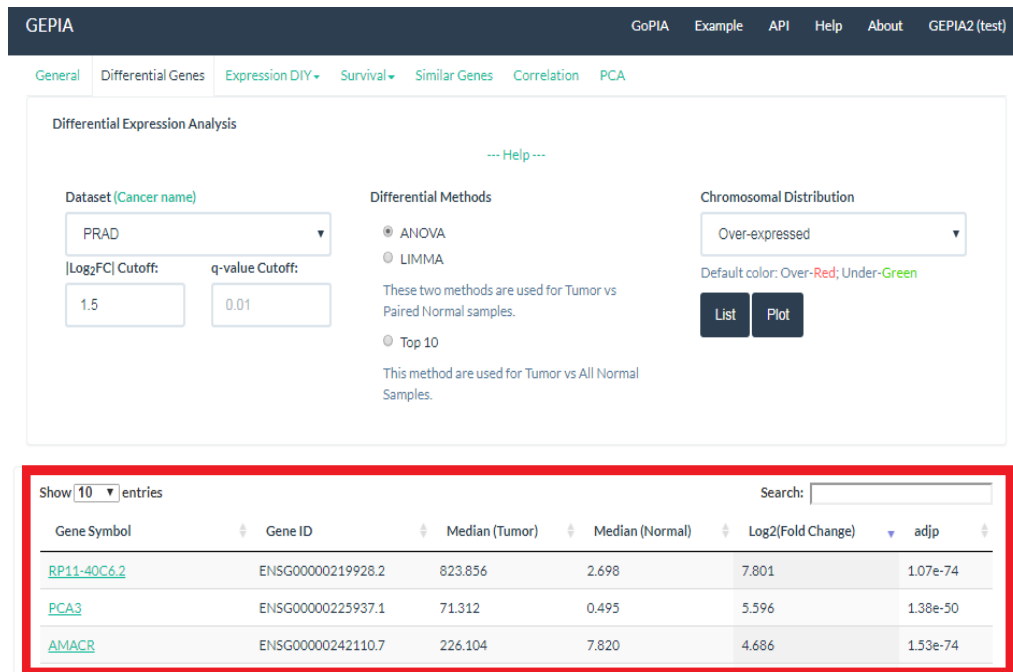
**Figura 15. Panel de resumen de estudios de CP en cBioPortal. En el recuadro rojo se muestra la lista de genes con ANC para los estudios seleccionados de CP. Gráfica tomada de la página web de cBioPortal (<https://www.cbioportal.org/>).**

#### **4.6.3. Evaluación de la expresión diferencial de genes en CP, genes con subexpresión y sobreexpresión.**

El perfil de expresión génica fue analizado con GEPIA. El recurso consideró 492 muestras de pacientes con CPr provenientes de la cohorte PRAD TCGA y un grupo de 152 controles incluyendo 52 muestras normales TCGA así como 100 muestras normales GETx (The Genotype-Tissue Expression). Se ingresó a la sección “Cancer Type Analysis”, luego se seleccionó la opción “Differential genes analysis” y a continuación, en Dataset (Cancer name) se seleccionó “PRAD”. En esta última ventana, en el casillero  $|\text{Log}_2\text{FC}|$  Cutoff: se tipeó 1.5 para seleccionar aquellos genes que tuviesen una expresión significativamente diferente con respecto a las muestras normales, las demás opciones se dejaron por defecto y se corrió el análisis en “List”. Los resultados se mostraron en una tabla en la parte inferior, que contuvo el nombre (símbolo) del gen, código de identificación del gen (Gene ID), la mediana de los transcritos por millón (TPM) del tumor, la mediana de los TPM de las muestras normales (controles), el  $\text{Log}_2$  Fold Change (FC) y el q-valor (Figura 13). Este último fue establecido en 0.01 y se tuvo en cuenta para el análisis estadístico, el mismo que evaluó las diferencias en los niveles de los TPM entre las muestras tumorales y controles.

Los genes fueron alistados considerando el valor del  $|\log_2\text{FC}|$  y los q-valores. Finalmente, se descargó los resultados de 227 genes que mostraron un  $\text{Log}_2$  Fold Change  $> 1.5$  (genes sobre-expresados) y de 754 genes que mostraron un  $\text{Log}_2$  Fold Change  $< -1.5$  (genes sub-expresados) (Figura 16). Se consideraron genes

codantes de proteínas y productores de ARNs de cadena larga no codificantes (lncRNAs).



**Figura 16. Panel de resultados de genes con expresión diferencial en CP de GEPIA. En el recuadro rojo se muestra la lista de GDEs obtenidos bajo los parámetros establecidos en materiales y métodos. Gráfica tomada de la página web de GEPIA (<http://gepia.cancer-pku.cn>).**

#### 4.6.4. Evaluación de la asociación de alteraciones genéticas con el fenotipo del CP

Con el fin de conocer si las alteraciones genómicas en PRAD se correlacionaban con el fenotipo de los individuos y/o con el fenotipo clínico de la enfermedad, se condujo un estudio de asociación genotipo-fenotipo. El genotipo estuvo dado por cada tipo de alteración genómica incluidas mutaciones, cambios en la expresión génica y ANCs.

El fenotipo incluyó variables demográficas de los sujetos y también variables clínico-patológicas del cáncer de estudio. A continuación, se describe el procedimiento por cada categoría de alteración genética.

#### **4.6.4.1. Asociación de mutaciones genéticas con el fenotipo.**

Para la evaluación de GSMs en CPp con el fenotipo se empleó cBioPortal. Se siguió el procedimiento detallado en la sección 4.6.1., una vez generado el panel de resumen con la información de los estudios de CP y habiendo aplicado los criterios de filtración señalados, se ubicó la opción “Quick Filters” y se seleccionó la opción que mostraba el número de muestras con información de mutación. Luego en la parte derecha se activó la opción “Group” que permitió crear un grupo con las muestras seleccionadas. Después se ubicó el cuadro que contenía los genes mutados y se ubicó el primer GSM y se repitió el paso anterior para crear un grupo muestras que tenían el gen mutado. Luego en la misma opción “Group” se seleccionaron los dos grupos previamente creados y se activó la opción “Compare” que generó una nueva ventana adicional que mostró diagramas de ambos grupos. Se seleccionó el área del diagrama que solo contenía muestras que no tenían mutación en el gen evaluado y se activó la opción de la parte inferior que permitía crear un nuevo grupo del área seleccionada (Figura 17).

Una vez realizado el paso anterior se retornó a la ventana inicial y en la opción “Group” se evidenció el nuevo grupo generado, se seleccionó solo este último y el grupo de muestras que contenía el gen mutado y se volvió activar la opción “Compare”. En la nueva ventana generada se ubicó la pestaña “survival” que

mostraba las gráficas de SG y SLE para los pacientes con el gen mutado y con el gen normal en evaluación. Se colectaron las gráficas que mostraron un p-valor significativo ( $p < 0.05$ ) para cualquiera de las dos variables. En el panel de resultados, en la pestaña “Clinical” se pudo conseguir la lista de variables clínicas-patológicas correlacionadas a la mutación del gen evaluado. En el recuadro de búsqueda se seleccionaron las variables de interés para el presente estudio (Estadio T, N, Puntaje de Gleason, PSA) y se colectaron aquellas correlaciones que mostraron un p-valor significativo. El mismo procedimiento se realizó para cada uno de los 10 primeros GFMs para CPp y CPM.



**Figura 17. Interfaz de cBioPortal para la comparación de dos grupos de muestras/pacientes. Seleccionando un área del diagrama se puede generar grupos de comparación con respecto a una variable genética o clínica. Gráfica tomada de la página web de cBioPortal (<https://www.cbioportal.org/>).**

#### 4.6.4.2. Asociación de genes con ANC con el fenotipo

La correlación de variables clínico-patológicas del CP con ANC se condujo a través de cBioPortal empleando la cohorte TCGA, el análisis solo se pudo llevar a cabo para muestras en estadio primario, ya que para los individuos con CPM los datos clínicos no estuvieron disponibles. Se siguió el procedimiento descrito en la sección 4.6.2. para los GFM con la diferencia que en este caso se evaluaría la correlación con genes que presentaron ANC, primero los genes frecuentemente deletados y luego los amplificados. Se ubicó la opción “Quick Filters” y se seleccionó la opción que mostraba el número de muestras con información de ANC. Luego los pasos siguientes fueron seguidos según la sección 4.6.4.1. hasta la generación de la ventana con la información de las variables de supervivencia (SG y SLE). En el mismo panel de resultados pero en ubicando la pestaña “Clinical” se pudo obtener la lista de variables clínicas-patológicas correlacionadas a la variable ANC para un determinado gen (Figura 18). En el recuadro de búsqueda se tipeó las variables de interés para el presente estudio (Estadio T, N, Puntaje de Gleason, PSA) y se colectaron aquellas correlaciones que mostraron un p-valor significativo  $P < 0.05$ ).

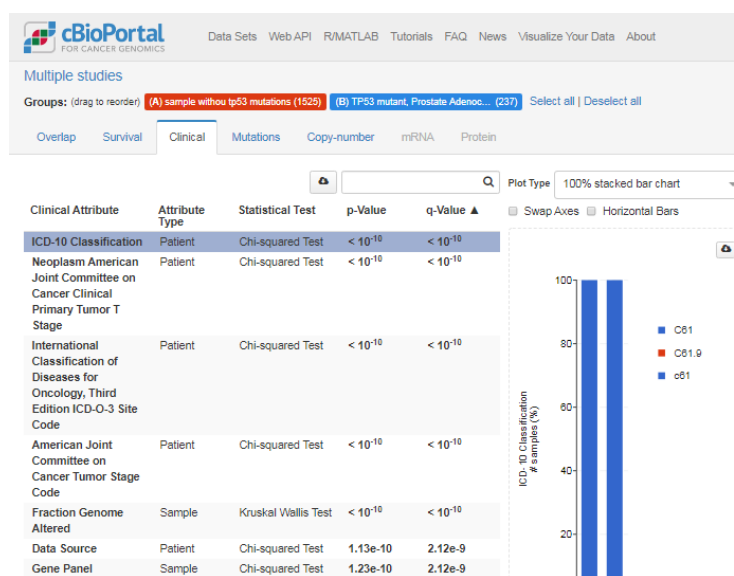


Figura 18. Interfaz de cBioPortal para los resultados de asociación de ANC con el fenotipo. Tipeando la variable clínica en el recuadro de búsqueda se puede obtener el la asociación con el gen con ANC en estudio. Gráfica tomada de la página web de cBioPortal (<https://www.cbioportal.org/>).

#### **4.6.4.3. Asociación de la expresión génica con el fenotipo**

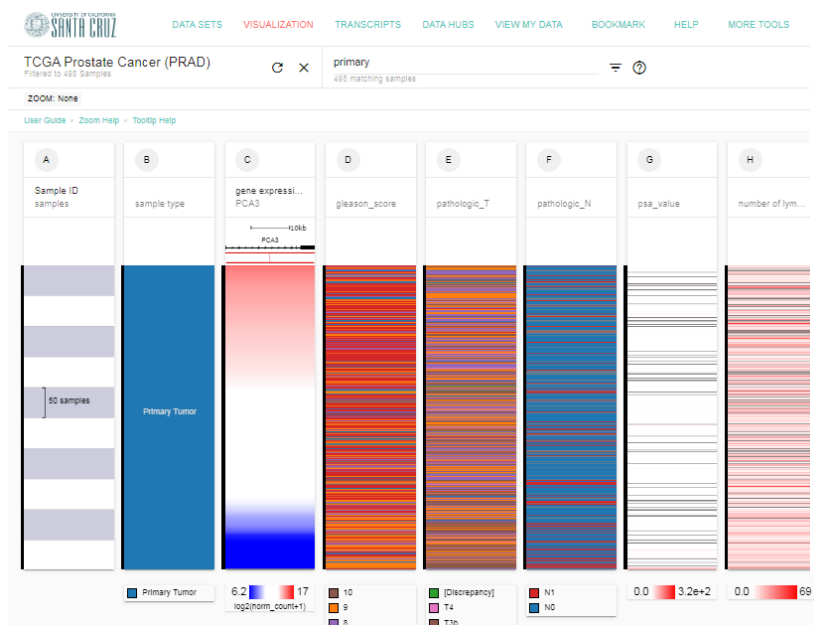
Con la finalidad de evaluar si la expresión génica de un gen estaba vinculada a alguna de las variables fenotípicas de estudio, se consultaron Gepia y Xena.

##### **4.6.4.3.1. Estudio de asociación con Xena**

Se ingresó al navegador a través de “Launch Xena” y en la primera columna en “Search for a study” se tipeó “TCGA Prostate Cancer (PRAD)”. En la siguiente columna, para definir la primera variable, se activaron las opciones “Phenotypic” y “Sample Type”, luego “Done”, para diferenciar las muestras de tumor primario y metastásico. En la tercera columna se repitió el paso anterior, lo cual generó una barra de búsqueda en la parte superior (Find Samples) donde se tipeó “Primary tumor” y luego “New column” en el menú superior, para correr el análisis sólo en las muestras de tumor primario. Luego se eliminó la tercera columna y se agregó una nueva para evaluar la asociación con la expresión génica.

En la nueva columna generada se activó la opción “Genomic”, luego se tipearon los genes a evaluar (aquellos sobre y sub expresados obtenidos de Gepia, sección 4.6.3.) y finalmente se activó la opción “Gene Expression (RNAseq Illum)”. Después, en una nueva columna se activó la opción “Phenotypic” y en la parte inferior se desplegó una lista con todas las variables fenotípicas disponibles. Se eligió evaluar las siguientes: Edad, Estadio T, Estadio N, Número de nodos linfáticos, Puntaje de Gleason y PSA. En la parte superior derecha, en el icono “view as chart”, los resultados se mostraron en gráficos de cajas o de dispersión dependiendo de la naturaleza de la variable evaluada.

Para el caso de variables categóricas como el estadio T, N y la escala de Gleason, se evaluaron los diagramas de cajas y se colectó el p-valor asociado al ANOVA (unidireccional) (Figura 19). Para las variables continuas como PSA y número de nodos linfáticos se evaluó los coeficientes de correlación de Spearman ( $r$ ) y Pearson ( $\rho$ ); así como el p-valor asociado a la correlación. Los resultados incluyeron las gráficas, el tamaño muestral, las pruebas estadísticas y la significancia estadística asociada al análisis. Se colectó la información anterior para todos los genes que presentaron un p significativo, y esta fue trabajada en una hoja de Excel para análisis subsecuentes.



**Figura 19.** Interfaz de XENA para la evaluación de asociación de a expresión genética con variables clínico-patológicas de Cpp. Gráfica tomada de la página web de XENA (<https://xena.ucsc.edu/>).

#### 4.6.4.3.2. Estudio de asociación con Gepia

Por otro lado, para conocer los genes asociados a la SG y SLE se consultó Gepia. En el menú superior “Survival” se seleccionó la opción “Most Significant

Survival Genes”. A continuación se seleccionó “PRAD” en “Datasets Selection” para ejecutar el estudio sólo en CP. Finalmente se eligió la opción Overall Survival dentro de métodos, en Group Cutoff se dejó por defecto la opción “Median” y se ejecutó la corrida en “List”. Los resultados aparecieron en una tabla que contuvo el nombre del gen, el código de identificación del gen (Gene ID) y el p-valor asociado al análisis de correlación entre la expresión génica y la supervivencia global (Figura 20). Los resultados fueron tabulados en hojas de Excel para su posterior análisis. El mismo procedimiento se realizó para determinar los genes correlacionados a la supervivencia libre de enfermedad, variando la opción OS por Disease Free Survival (RFS) en métodos. Los resultados fueron tabulados en Excel.

Una vez obtenidos los genes asociados con el pronóstico de la enfermedad, tanto para la SG como para la SLE, se procedió a evaluar sus perfiles de expresión con el fin conocer si eran genes sobre o sub-expresados. Para ello, en una hoja de Excel que contuvo ambas las lista de genes ( GDEs y aquellos asociados a SG y SLE), con la opción de “Formato condicional” se trabajó en la identificación de genes asociados a SG y SLE, y que a la vez estuviesen presentes en el grupo de genes sobre o sub expresados. Con estas acciones se obtuvo una lista integrada de genes con expresión diferencial y asociados a las variables de supervivencia señaladas (SG y SLE).

GEPIA GoPIA Example API Help About GEPIA2 (test)

[Click here to get the extension of tumor abbreviations.](#)

General Differential Genes Expression DIY Survival Similar Genes Correlation PCA

Most Differential Survival Genes --- Help ---

Datasets Selection: PRAD

Methods:
  Overall Survival
  Disease Free Survival (RFS)

Group Cutoff:
  Median
  Quartile
 List

Show 10 entries Search:

Gene Symbol	Gene ID	P-Value (Survival os)
<a href="#">RP11-178H8.7</a>	ENSG00000279696.1	3.29e-4
<a href="#">ZNF433</a>	ENSG00000197647.11	3.80e-4
<a href="#">TGDS</a>	ENSG00000088451.10	4.58e-4
<a href="#">RPAIN</a>	ENSG00000129197.14	6.89e-4
<a href="#">TRAPPC8</a>	ENSG00000153339.13	8.40e-4
<a href="#">EEF1A1P10</a>	ENSG00000243746.1	8.67e-4

**Figura 20. Interfaz de Gepia para la evaluación de asociación de la expresión genética con variables de supervivencia. En el recuadro inferior se muestra la lista de genes con expresión diferencial y su asociación con la SG en pacientes con CP. Gráfica tomada de la página web de GEPIA (<http://gepia.cancer-pku.cn>).**

#### 4.6.5. Identificación de rutas enriquecidas en CP.

La identificación de las rutas enriquecidas se realizó consultando Genecodis. Se evaluaron los genes que presentaron al menos un tipo de alteración genética. Así, se investigaron 3 categorías de genes: (i) los GSM, identificados con Firebrowse y cBioPortal, (ii) los genes con ANC, identificados con cBioPortal y (iii) los genes diferencialmente expresados, obtenidos de Gepia. En Genecodis, “Select the organism” se seleccionó “Homo sapiens” y luego se precisaron los términos de ontología y anotaciones a evaluar en el campo en “Select the annotations” incluyendo (i) Proceso Biológico, (ii) Función Molecular, (iii) Componente Celular y (iv) Rutas KEEG (Enciclopedia de Kyoto de genes y genomas) (Figura 21). A continuación, se ingresó el nombre de los genes por cada categoría (una

categoría por vez), en la tabla “Paste your lists of genes” y se corrió el análisis con “Submit”.

Los resultados se mostraron en una ventana emergente y contuvieron los cuatro términos de ontología solicitados para los genes ingresados. Se activó la opción “Interactive table” que permitió generar una tabla con los genes, sus anotaciones, el p-valor asociado al análisis de la anotación; así como el número de genes de la lista de referencia empleada para la evaluación. Finalmente, se descargó la información indicada en la tabla interactiva para cada componente y se trabajó en hojas de Excel para su tabulación.

The image shows the GeneCodis web interface for gene ontology and pathway analysis. At the top, there are logos for GeneCodis and Gene annotations co-occurrence discovery. Below the logos is the title "Modular and Singular Enrichment Analysis [?]".

1. Select the organism [?]  
Homo sapiens

2. Select the annotations [?] [Last update on Dec, 2011]

Annotations

- GO Biological Process
- GO Molecular Function
- GO Cellular Component
- GOSlim Process
- GOSlim Function
- GOSlim Component
- KEGG Pathways
- InterPro Motifs
- MicroRNA
- Omim Diseases
- Panther Pathways
- PharmGKB Drugs
- Pubmed
- Transcription Factors

3. Paste your lists of genes [see allowed IDs]

Paste list of genes [?]  
SPOP  
TP53  
PTEN  
NUDT11  
MLL2  
CTNNB1  
PIK3CA  
UTM1A

Or upload a file with the list of genes [?]  
Seleccionar archivo Ningún archi...seleccionado

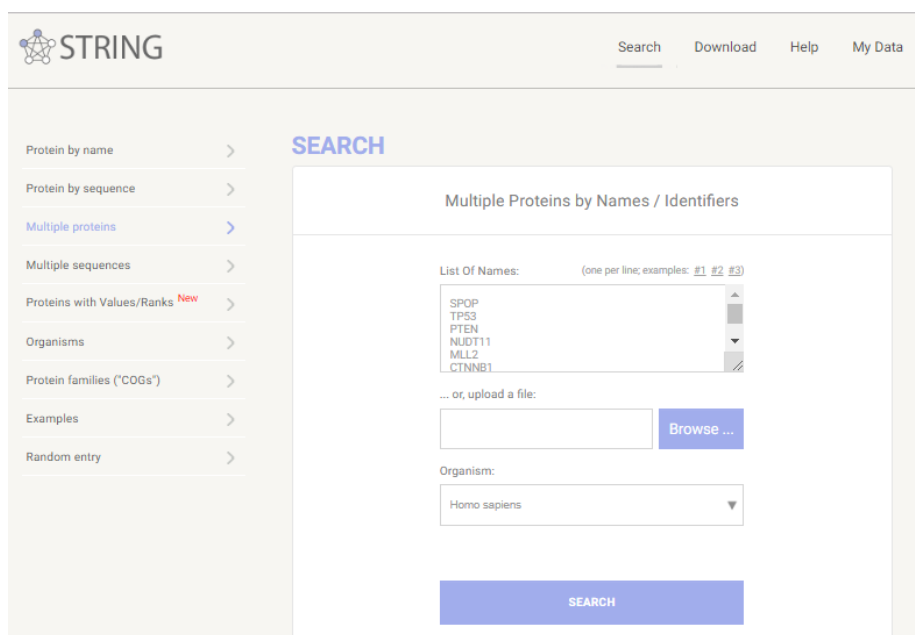
[Load yeast data] [Load human data]

**Figura 21. Interfaz de Genecodis para el análisis de ontología de genes (GO) y rutas enriquecidas KEEG. Se muestran las primeras opciones de selección para correr el análisis de un determinado grupo de genes ingresados. Gráfica tomada de la página web de Genecodis (<http://genecodis.cnb.csic.es/analysis>).**

#### **4.6.6. Redes de interacción proteína-proteína (interactomas)**

Se determinaron los interactomas de aquellas proteínas codificadas por los genes diferencialmente expresados, ya sea sobre o sub-expresados en CP. Para representar los interactomas globales (de grupos de proteínas) se consultó STRING, seleccionando la opción “Multiple proteins”, luego ingresando los nombres de los genes diferencialmente expresados en “List of Names” y precisando el organismo de interés (Homo sapiens). Finalmente se corrió el análisis con “Search” (Figura 22).

El resultado incluyó la imagen interactiva del interactoma o networking, que mostró conectividad entre proteínas. Aquellas proteínas sin contacto fueron eliminadas de la representación mediante la opción “Settings” y se dejó por defecto el valor de confianza de 0.4. El interactoma se analizó en función del número de nodos, donde cada nodo representa todas las proteínas producidas por un solo locus genético codificador de proteínas, color del nodo que representa el tipo de interacción y el contenido del nodo que indica si la proteína cuenta con estructura 3D conocida; así como el tipo del borde de la PPI que indica si son interacciones conocidas o predichas. De esta sección, se colectó información del enriquecimiento funcional (procesos biológicos, función molecular, componente celular y rutas KEGG) y significancia estadística del PPI, por cada interactoma.



**Figura 22. Interfaz de STRING para el análisis de interacción proteína-proteína (PPI). Se muestran las primeras opciones de selección para correr el análisis de un determinado grupo de genes ingresados. Gráfica tomada de la página web de STRING ([https://string-db.org/cgi/input.pl?sessionId=MWPVGJJKphyT&input\\_page\\_show\\_search=on](https://string-db.org/cgi/input.pl?sessionId=MWPVGJJKphyT&input_page_show_search=on)).**

#### **4.6.7. Predicción del daño en la estructura y función de las proteínas.**

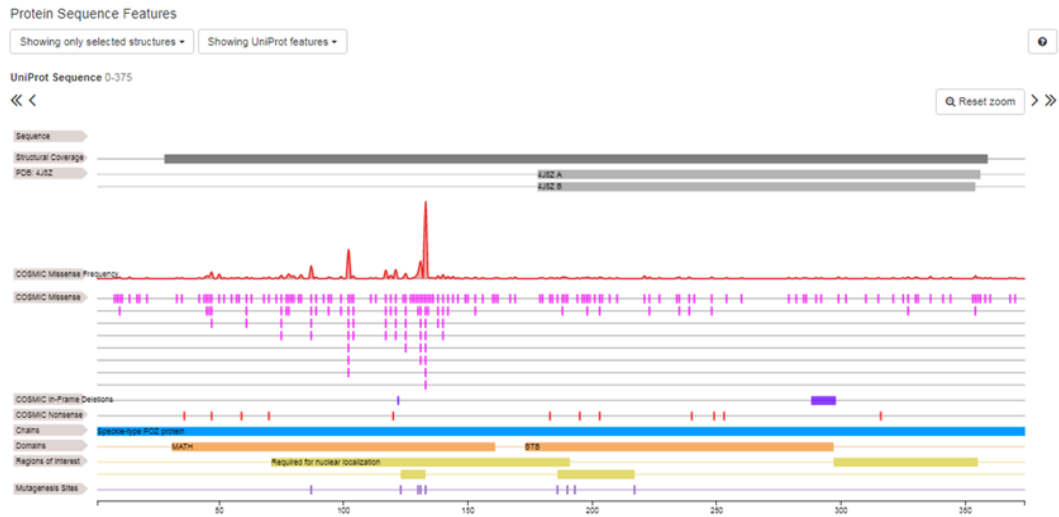
Uno de los objetivos del estudio fue conocer en qué medida las mutaciones genéticas afectaban la estructura y función de las proteínas codificadas por los GSMs en CPp. Para ello se se predijo de forma teórica el daño en la estructura y función de las proteínas, empleando los siguientes recursos: Biomuta, Vapros, Uniprot Cosmic, cBioportal e Imutant. Se tomaron en cuenta los cinco primeros GSMs en CPp: SPOP, TP53, FOXA1, MLL2 (KMT2D) y MLL3 (KMT2C).

En primer lugar se utilizó Biomuta para identificar las mutaciones probablemente patogénicas para cada gen. En la barra de búsqueda se ingresó el nombre del gen y se activó la opción “Search”, luego se seleccionó el código UniprotKB

correspondiente al gen de interés. Los resultados se mostraron en una tabla que contenía las lista de alteraciones para el gen ingresado, así como la predicción funcional con la probabilidad de ser una alteración dañina o benigna, el tipo de cáncer al cual está asociado la alteración y otros datos adicionales. Se descargó la tabla con la opción “Download table in CSV format” y se tabuló en una hoja Excel. Luego se filtró solo las alteraciones asociadas a CP y que presentaran una predicción funcional determinada como probablemente dañina.

Luego se empleó Uniprot y Vapros para identificar los dominios y regiones funcionales de las proteínas evaluadas, obteniendo también los rangos de aminoácidos que albergaban las alteraciones probablemente dañinas. Además se consultó la base de datos Cosmic para identificar aquellas regiones de interacción intermolecular ya fuera con otros productos proteicos o con iones metálicos así como dominios no identificados con las herramientas utilizadas anteriormente (Figura 23). A través de cBioPortal se obtuvieron representaciones de las estructuras tridimensionales de algunas regiones de importancia de las proteínas evaluadas. Siguiendo el procedimiento descrito en la sección 4.6.1., dentro del panel de resumen de estudios se ubicó en la parte izquierda en el recuadro de búsqueda donde se ingresó tipeó el nombre del gen y se activó la opción “Query”. Luego se generó una nueva ventana donde se activó la opción “Mutations” la cual mostraba la secuencia de la proteína con las mutaciones que albergaba en sus diferentes dominios. En la parte izquierda se ubicó la opción “View 3DStructure”, la cual fue activada para generar la estructura tridimensional disponible para la

proteína en cuestión. La imagen fue descargada para cada uno de los cinco genes en evaluación.



**Figura 23. Interfaz de Cosmic para el análisis de interacción proteína-proteína (PPI). Se muestran las primeras opciones de selección para correr el análisis de un determinado grupo de genes ingresados. Gráfica tomada de la página web de Cosmic (<https://cancer.sanger.ac.uk/cosmic>).**

Finalmente, se predijo el cambio de estabilidad en la proteína producido por cada una de las mutaciones probablemente dañinas, mediante I-Mutant 3.0. En la interfaz principal de esta herramienta se ubicó la sección “Prediction of protein stability changes upon single point mutation from” donde se seleccionó la opción “Protein Sequence” y se activó la opción “Enter”. En la nueva ventana generada se ingresó la secuencia de la proteína en formato FASTA, previamente descargada de Uniprot, luego se indicó la posición de la mutación y el nuevo residuo. Los parámetros de temperatura y pH se dejaron por defecto y se seleccionó la opción “DDG Ternary Classification”, por último se tipeó una dirección de correo electrónico a donde los resultados serían enviados (Figura 24). Una vez obtenidos

los resultados, se colectó la información correspondiente a la predicción del efecto desestabilizante de cada mutación, para ello se consideró el valor del cambio de energía libre (CEL) según un  $CEL < -0.5$ , cambio desestabilizador, un  $CEL > 0.5$ , un incremento en la estabilización y un  $-0.5 < CEL < 0.5$ , cambio neutral.

The screenshot shows the I-Mutant 3.0 web interface. At the top, it reads "I-Mutant  $\Delta\Delta G$  Predictor of stability change upon single point protein mutation". On the left, there is a navigation menu with links: "I-Mutant Suite Home", "I-Mutant Suite Help", "Biocomputing Unit", "Contact us", and "Last Update 25/12/06". The main area contains a "Protein Sequence" input field with a dropdown menu showing a list of amino acid sequences. Below this, there are input fields for "Position" (set to 53), "New Residue" (set to C), "Temperature" (set to 25), and "pH" (set to 7). There are two radio buttons for "Prediction": "DDG Value and Binary Classification" (selected) and "DDG Ternary Classification". An "e-mail" field contains "jorge.lara.z@upch.pe". On the right side, there are labels for "Sequence residue number", "New Residue", "Temperature in Celsius degrees [0-100]", "pH value [0-14]", and "[DDG >=0, DDG < 0]". At the bottom right, there is a "Submit" button.

**Figura 24. Interfaz de I-Mutant para evaluar el cambio de estabilidad de una proteína producido por una mutación. En la imagen se representa los pasos iniciales para introducir la secuencia de la proteína, la posición de la mutación, el aminoácido alterado y otros parámetros. Gráfica tomada de la página web de I-Mutant (<http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi>).**

#### **4.6.8. Frecuencia de mutaciones conductoras o drivers del CP.**

Un objetivo del presente estudio fue describir el paisaje mutacional en el CP e identificar la frecuencia de aparición de las mutaciones conductoras de esta neoplasia en el grupo de estudio. Para ello, se elaboró una lista de mutaciones conductoras del CP, disponibles en “Cancer Genome Interpreter” (CGI). Esta herramienta emplea diferentes base de datos y recursos para el análisis mutacional, así como los resultados de varios ensayos experimentales publicados: (i) DoCM, Una base de datos de mutaciones curadas en el cáncer; (ii) ClinVar, un archivo público de interpretaciones de variantes clínicamente relevantes y (iii) OncoK, un recurso que anota el efecto biológico y oncogénico y la significancia pronóstica y predictiva de las alteraciones moleculares somáticas.

El primer paso fue ingresar a la página web de “Cancer Genome Interpreter” y se ingresó a la opción “Datasets”, luego a la opción “Validated Oncogenic Mutations”. Se pudo visualizar un tabla con una lista de mutaciones oncogénicas con múltiples opciones de filtrado en la parte superior. Se ubicó la sección “Tumor Type” y se tipeó “Prostate”. Automáticamente se filtraron los resultados para las mutaciones conductoras de CP y con la opción “Download” se pudo descargar la información correspondiente a la mutación, el nombre del gen asociado, el tipo de mutación, etc., y la información anterior fue tabulada en una hoja Excel (Figura 25).

[Download](#)

Gene ^	GDNA	Protein change	Transcript	Context	Tumor type
Search he					prostate
<b>AKT1</b>	chr14:g.105246551C>T	p.E17K	ENST00000554581	somatic	1
<b>AR</b>	chrX:g.66931463T>A	p.L702H	ENST00000374690	somatic	1 Prostate adenocarcinom
<b>AR</b>	chrX:g.66943549T>C	p.F877L	ENST00000374690	somatic	1 Prostate adenocarcinom
<b>AR</b>	chrX:g.66943552A>G	p.T878A	ENST00000374690	somatic	1 Prostate adenocarcinom
<b>CHEK2</b>	chr22:g.29107974C>A	p.E239*	ENST00000328354	somatic	1 Prostate adenocarcinom
<b>EGFR</b>	chr7:g.55242433G>A	p.G735S	ENST00000275493	somatic	1 Non-small cell lung, Pros
<b>EPHB2</b>	chr1:g.23189553G>T	p.A279S	ENST00000374632	somatic	1 Prostate adenocarcinom
<b>EPHB2</b>	chr1:g.23233346G>A	p.D679N	ENST00000374632	somatic	1 Prostate adenocarcinom
<b>EPHB2</b>	chr1:g.23234473C>T	p.Q723*	ENST00000374632	somatic	1 Prostate adenocarcinom
<b>KLF6</b>	chr10:g.3824003A>G	p.L169P	ENST00000497571	somatic	1 Prostate adenocarcinom
<b>KLF6</b>	chr10:g.3824099G>T	p.S137*	ENST00000497571	somatic	1 Prostate adenocarcinom
<b>KLF6</b>	chr10:g.3824141G>T	p.A123D	ENST00000497571	somatic	1 Prostate adenocarcinom
<b>KLF6</b>	chr10:g.3824163A>G	p.S116P	ENST00000497571	somatic	1 Prostate adenocarcinom
<b>KLF6</b>	chr10:g.3824319A>G	p.W64R	ENST00000497571	somatic	1 Prostate adenocarcinom
<b>MAD1L1</b>	chr7:g.2265161G>A	p.R59C	ENST00000406869	somatic	1 Prostate adenocarcinom
<b>PTEN</b>	chr10:g.89711946T>A	p.Y188*	ENST00000371953	somatic	1 Prostate adenocarcinom

**Figura 25. Interfaz de CGI para la determinación de mutaciones conductoras del CP. Gráfica tomada de la página web de CGI (<https://www.cancergenomeinterpreter.org/home>).**

Luego de haber obtenido las mutaciones conductoras, se procedió a estimar su frecuencia de aparición en los participantes de diversos estudios globales de CP, albergados en CBioPortal, y que comprometen pacientes con CP de otros estudios no TCGA. Para esto, se ingresó al recurso, se eligió “Prostate” como órgano de estudio y se seleccionaron solo los estudios que contenían datos de mutaciones genéticas. Después, se ingresó a “View Summary” (según sección 4.6.1.) y en la sección “Cancer Type Detailed” se seleccionó la opción “Prostate Adenocarcinoma”, luego en la categoría “Number of Samples Per Patient” se eligió la opción “1” (que hace referencia a una sola muestra donada por paciente, para evitar repeticiones de muestras analizadas). Luego se ubicó la categoría “Sample Type”, donde se escogió la subcategoría “Primary”.

Una vez delimitado los filtros anteriores, se procedió a tipear los nombres de cada uno de los genes que albergaban una mutación conductora, según informó CGI, y se corrió el análisis dando clic en “Query”. En los resultados devueltos se buscó la mutación conductora de interés por inspección visual o tipeando la misma en la barra de búsqueda.

Si la población de estudio seleccionada presentaba la mutación conductora, se precisaba también la frecuencia de aparición. De esta forma se obtuvo el número de muestras que presentaron las mutaciones conductoras y se verificó que no hubiese muestras duplicadas en el análisis. Se registró el número total de muestras que presentaron la mutación puntual y el número total de muestras/pacientes, con el fin de calcular las frecuencias por mutación. La información fue tabulada en una hoja de Excel. Aplicando el mismo protocolo, se estimaron las frecuencias de aparición de mutaciones conductoras en la población de pacientes con Cpm (Fase IV de la enfermedad). En este caso, para seleccionar el grupo de estudio, en el campo “Sample Type”, se escogió las subcategorías “Metastasis” y “Metastasic”. El resto de pasos subsecuentes se siguió como antes fue descrito.

#### **4.6.9. Frecuencias de polimorfismos de un solo nucleótido (SNPs) de riesgo a CP.**

Se investigó la frecuencia de aparición de SNPs en Cpp y metastásico. Para ello, se consultó el catálogo de GWAS, ingresando el término “Prostate Carcinoma” en el casillero de búsqueda. La nueva ventana desplegó información de las asociaciones (entre polimorfismos y neoplasia) halladas para Cpp (EFO\_0001663) (Figura 26). En la parte inferior en “Trait information” se ubicó la opción “Download Catalog data” que permitió descargar el nombre del gen afectado, tipo de estudio, código del SNP, tipo de variante, entre otros datos adicionales. Luego se filtraron solo aquellos polimorfismos que fueron del tipo pérdida de sentido (missense), además se verifico que no hubiesen polimorfismos repetidos mediante el código del SPN (rs#). El mismo procedimiento se realizó para descargar los polimorfismos asociados a cáncer metastásico (EFO\_0000196). En este caso, se seleccionó la alternativa Metastasic prostate cáncer en la barra de búsqueda luego de tipear “Prostate Carcinoma” y el resto de pasos se siguieron como anteriormente fueron descritos.

Una vez obtenidos los SNPs de susceptibilidad a Cpp y metastásico, se procedió a estimar la prevalencia de cada uno de ellos en cBioPortal. Se siguió de forma de las mutaciones conductoras (sección 4.5.3). Se calcularon dichas frecuencias para cada estadio de la enfermedad y toda la información fue tabulada en una hoja de Excel.

The screenshot displays the GWAS Catalog interface. At the top, the logo and title 'GWAS Catalog' are visible, along with the subtitle 'The NHGRI-EBI Catalog of published genome-wide association studies'. A search bar contains the text 'prostate carcinoma'. Below the search bar, there are examples of search terms: 'breast carcinoma, rs7329174, Yao, 2q37.1, HBS1L, 6:16000000-25000000'. The main content area is titled 'Search results for prostate carcinoma' and lists two results:

- prostate carcinoma** (EFO\_0001663): A carcinoma that arises from epithelial cells of the prostate gland. One of the most common malignant tumors affecting men. The majority of carcinomas arise in the peripheral zone and a minority occur... [Show more >](#)  
Associations: 728 Studies: 66
- metastatic prostate cancer** (EFO\_0001166): A carcinoma that arises from the prostate gland and has spread to other anatomic sites.  
Associations: 12 Studies: 2

On the left side, there is a 'Refine search results' panel with a 'Traits' filter and 'Catalog stats' section:

**Catalog stats**

- Last data release on 2019-10-14
- 7296 publications
- 116109 SNPs
- 159202 associations
- Genome assembly GRCh38.p12
- dbSNP Build 151
- Ensembl Build 96

**Figura 26. Interfaz de GWAS para la determinación de SNPs de riesgo a Cpp y Cpm. Se muestran los resultados obtenidos para el término “Prostate Carcinoma”. Gráfica tomada de la página web de GWAS (<https://www.ebi.ac.uk/gwas/>).**

## **5. CONSIDERACIONES ETICAS**

El presente trabajo de investigación no presento ningún riesgo para la salud de los participantes. Los datos obtenidos de los recursos en línea no permitieron identificar a ningún participante. El proyecto fue aprobado por Comité Institucional de Ética en Investigación (CIEI), constancia N° 234-10-18, con fecha 25 de abril de 2018.

## **6. Resultados**

### **6.1. Población de estudio**

Los sujetos de estudio fueron pacientes con CPp y metastásico cuyos datos están disponibles en cBioPortal. Este recurso contiene datos genómicos de 15 estudios ejecutados a nivel mundial, con un total de 3891 pacientes con CP. La mayoría de los sujetos (45%) presentaron CPp (n=1739) (Tabla 5) mientras que una parte de la cohorte (20%) presentó CPm (n=794) y el 35% no tuvieron precisada dicha información (n=1358). Tal como se esperaba, la mayoría de individuos presentaron adenocarcinoma de próstata (85%), observándose con menos frecuencia otros subtipos histológicos como el CP del tipo neuroendocrino (n=28) y el de células pequeñas (n=4), estos dos últimos subtipos de CP no fueron considerados en la presente tesis.

En la Tabla 5 se presentan las principales características demográficas y clínicas de los sujetos de estudio, según si presentaban CPp. La edad promedio al momento del diagnóstico para los individuos con CPp fue de 61 años, mientras que para el grupo de pacientes con CPm fue de 66 años (dato no mostrado). El 84% de los individuos con CPp no tuvo disponible la raza, mientras que en aquellos sujetos que presentaron este dato predominó la raza blanca (14.7%), seguido por africanos y asiáticos. De forma similar para la categoría etnia, el 85% de los individuos no registraron dato para esta variable, y el 15% de la población que sí presentó información se precisó que eran no hispanos (no latinos). En cuanto a las variables de sobrevida como la SG, de los sujetos con registro de seguimiento, el 34% mostró una SG de hasta 5 años, el 9% una SG entre 5 y 10 años; mientras que solo el 1% mostro una sobrevida mayor a 10 años. De los

pacientes con registro de seguimiento para el tiempo libre de enfermedad, el 18% mostro una sobrevida libre de enfermedad de hasta 3 años, el 10% entre 3 y 5 años, y seguido por un 9% con una SLE mayor a 5 años. Dentro de las variables clínico-patológicas, se evaluó el puntaje de Gleason para el cual el 85.7% de los sujetos no presentaron la información. De los hombres para los cuales se registró el puntaje, el 6.7% mostro un Gleason alto (9), el 2.4% un Gleason 8 y seguido por un 4.8% con un Gleason de 7. El 74.6% de los hombres no tuvieron el valor de PSA, de los sujetos con el PSA disponible el 18.4% mostro un valor  $\leq 4$  ng/ml, mientras que el 7% tuvieron un PSA  $>4$  ng/ml. Cabe precisar que la mayoría de los profesionales de la salud toman como referencia un valor de 4,0 ng/ml o menos como valor normal del PSA, aunque no necesariamente es concluyente para diagnosticar la enfermedad.

Sobre el estatus patológico de la enfermedad, aproximadamente la mitad de la población no tuvo la calificación T para el tumor, de los datos disponibles el 25% de sujetos tuvo un T2, indicativo que el tumor se encontró en la próstata y no en otras partes del cuerpo, seguido por un 21% con categoría T3 y un 1% con categoría T4. Para la categoría N, más de los tres cuartos de los individuos no tuvieron el dato registrado, de los que tuvieron la calificación el 20% mostro un N0, lo que indico que el cáncer no se diseminó a los ganglios linfáticos regionales. El 4.5% de los sujetos tuvo un N2, donde el cáncer ya había invadido los ganglios linfáticos.

**Tabla 5. Características demográficas y clínicas de la población de pacientes con CPP albergados en los 15 estudios reportados en cBioPortal.**

Tipo de Variable.	Características	Primario N= 1739	
		N	(%)
	Edad promedio al momento del diagnóstico	61 ( $\pm 7.58$ )	
Demográficas	Raza		
	Blanco	255	14.7%
	Negro o Africano	12	0.7%
	Asiático	4	0.2%
	ND	1468	84.4%
	Categoría de Etnia		
	NO HISPANO O LATINO	265	15.2%
	ND	1474	84.8%
De sobrevida	Tiempo de SG		
	0-5 años	595	34.2%
	>5 años-10 años	160	9.2%
	>10 años	25	1.4%
	ND	959	55.2%
	Tiempo de sobrevida libre de enfermedad		
	0-3 años	311	17.9%
	>3años-5 años	171	9.8%
	>5 años	166	9.5%
ND	1091	62.8%	
Clínico-patológicas	Escala de Gleason		
	6	4	0.2%
	7	83	4.8%
	8	41	2.4%
	9	117	6.7%
	10	4	0.2%
	ND	1490	85.7%
	PSA (ng/ml)		
	$\leq 4$	320	18.4%
	$> 4$	122	7.0%
	ND	1297	74.6%
	Estadio T		
	T2	435	25.0%
	T3	373	21.4%
	T4	24	1.4%
	ND	907	52.2%
Estadio N			
N0	346	19.9%	
N1	79	4.5%	
ND	1314	75.6%	

ND: Información no disponible

## **6.2. Paisaje de las alteraciones genómicas en el CP**

### **6.2.1. Genes significativamente mutados**

#### **6.2.1.1. Genes mutados en Cpp.**

La firma mutacional del Cpp se obtuvo por Firebrowse y se consideró para el análisis la información genómica de 498 pacientes provenientes de la cohorte de PRAD-TCGA. Se identificaron 35 GSMs en Cpp (Tabla 6), aplicando MutSig 2CV v3. Los genes supresores de tumor SPOP y TP53 resultaron ser los dos genes más mutados, cada uno presente en el 11% de la cohorte TCGA. Curiosamente, ambos genes se localizan en el cromosoma 17. MLL2 y MLL3, Histona-lisina N-metiltransferasa 2D y 2C respectivamente, también fueron reconocidos como SMGs en el 6% de la población, cada uno. FOXA1, un factor transcripción involucrado en el desarrollo embrionario y diferenciación de tejidos, presentó mutaciones en el 6% del grupo de estudio, mientras que ATM, que codifica la proteína ATM serina/treonina quinasa que está vinculada a la regulación de procesos de control de la división celular y en la reparación de daños al ADN, estuvo mutado en el 4% de la población. Otros genes como MED15, un cofactor de transcripción del complejo mediador involucrado en la transcripción dependiente de la ARN polimerasa II, e IDH1, que codifica una isocitrato deshidrogenasa 1 (NADP +) que cataliza la descarboxilación oxidativa del isocitrato a 2-oxoglutarato, tuvieron bajas frecuencias de aparición (1%) en la cohorte TCGA.

**Tabla 6. Genes significativamente mutados en Cpp<sup>a</sup>**

Cromosoma	Símbolo del gen	Nombre	Función molecular	Frecuencia de aparición	p-valor	q-valor
17q21.33	SPOP	Proteína POZ tipo mota	Este gen codifica una proteína que puede modular las actividades de represión transcripcional de la proteína 6 asociada a la muerte.	11%	1.00E-16	9.10E-13
17p13.1	TP53	Antígeno tumoral celular p53	Proteína involucrada en la regulación del ciclo celular, actúa como un transactivador que modula la división celular.	11%	1.00E-16	9.10E-13
12q13.12	MLL2	Histona lisina N-metiltransferasa 2D38/5000	Proteína que funciona como un coactivador para el receptor de estrógeno, haciendo que se active la transcripción.	6%	3.50E-11	1.30E-07
7q36.1	MLL3	Histona lisina N-metiltransferasa 2C	Histona metiltransferasa, actúa metilando la 'Lys-4' de histona H3, que a su vez indica una señal para activar un proceso transcripcional epigenético.	6%	3.90E-07	7.10E-04
14q21.1	FOXA1	Factor nuclear del hepatocito 3-alfa	Es un factor de transcripción que está vinculado al desarrollo embrionario, también a la regulación de la expresión génica en tejidos diferenciados.	6%	1.40E-07	2.80E-04
11q22.3	ATM	Serina-proteína quinasa ATM	Serina / treonina proteína quinasa que activa la señalización del punto de control en las roturas de doble cadena (DSB), la apoptosis y las tensiones genotóxicas como la luz ionizante ultravioleta A (UVA), actuando así como un sensor de daño al ADN.	4%	1.60E-04	8.40E-02
10q23.31	PTEN	fosfatidilinositol-3,4,5-trisfosfato 3-fosfatasa	Es un supresor de tumor.	3%	1.80E-15	1.10E-11
16q22.2-q2	ZFH3	Proteína homeobox dedo de zinc homeobox proteína 3 Fosfatidilinositol 4,5-bisfosfato 3-cinasa subunidad catalítica alfa	Regulador transcripcional que puede actuar como activador o represor. Inhibe el elemento potenciador del gen AFP al unirse a su secuencia central rica en AT.	3%	1.10E-04	6.70E-02
3q26.32	PIK3CA	El gen PIK3CA induce la producción de la proteína p110 alfa (p110 $\alpha$ ), que es una subunidad de la enzima llamada fosfatidilinositol 3-quinasa (PI3K).		3%	1.50E-09	3.90E-06
3p22.1	CTNNB1	Catenina beta-1	Es una proteína que está asociada a la vía de señalización Wnt	3%	2.30E-10	6.90E-07
Xp11.3	KDM6A	Demetilas a 6A específica de lisina	Es una proteína Histona desmetilasa que actúa únicamente sobre Lys-27 de la histona H3	3%	8.30E-09	1.90E-05
Xq13.1	ZMYM3	Proteína 3 tipo YMY dedo de zinc	Proteína que tiene un rol en la modulación de la morfología celular y estructuración del citoesqueleto.	2%	1.20E-04	7.10E-02
Xp11.22	NUDT1	Difosfoinositol polifosfato fosfohidrolasa 3-beta Repetición de	Esta proteína rompe un beta-fosfato de los grupos difosfato en PP-InsP5 (difosfoinositol pentakisfosfato), indicio que podría desempeñar un papel en la transducción de señales.	2%	7.20E-14	3.30E-10
7p22.1	TNRC18	trinucleótidos que contiene 18	unión a la cromatina	2%	1.10E-06	1.60E-03

17q21.2	CNTNAP1	Proteína asociada a contactina 1	Actúa, con CNTNAP2, en la organización radial y longitudinal de los axones mielinizados.	2%	5.30E-06	5.60E-03
1p36.13	NBPF1	Miembro 1 del NBPF	Esta proteína es miembro de las proteínas de la familia del punto de ruptura del neuroblastoma, un grupo de proteínas que podrían estar involucradas en el desarrollo del sistema nervioso.	2%	1.20E-05	1.10E-02
1q25.3	SMG7	Proteína SMG7	Proteína involucrada en la desintegración del ARNm sin sentido	2%	3.50E-06	4.00E-03
Xq13.1	MED12	Mediador de la subunidad 12 de transcripción de la ARN polimerasa II	Esta proteína es un coactivador asociado a la regulación de la transcripción de casi todos los genes dependientes de la ARN polimerasa II.	2%	3.50E-05	2.80E-02
22q11.21	MED15	Mediador de la subunidad 15 de transcripción ARN polimerasa II	Componente del complejo mediador, un coactivador involucrado en la transcripción regulada de casi todos los genes dependientes de la ARN polimerasa II.	1%	6.80E-05	4.80E-02
2q34	IDH1	Isocitrato deshidrogenasa [NADP] citoplásmica	Actividad catalítica	1%	2.40E-06	3.10E-03
12p13.1	CDKN1B	Inhibidor de quinasa dependiente de ciclina 1B	Importante regulador de la progresión del ciclo celular.	1%	5.50E-06	5.60E-03
7q31.32	LMOD2	Leiomodina-2	Media la nucleación de los filamentos de actina y por lo tanto promueve la polimerización de actina	1%	1.60E-05	1.40E-02
Xp11.23	GAGE2A	Antígeno G 2A	Este gen pertenece a una familia multigénica expresada en una gran variedad de tumores, mientras que en los tejidos normales, la expresión está restringida a las células germinales.	1%	8.20E-07	1.40E-03
10q11.23	AGAP6	Arf-GAP con GTPasa, repetición ANK y proteína 6 que contiene el dominio PH	Protina que hipotéticamente activa la GTPasa.	1%	1.70E-06	2.40E-03
3p24.1	EOMES	Homólogo de eomesodermin	Funciona como un activador transcripcional que juega un papel crucial durante el desarrollo.	1%	7.60E-06	7.30E-03
3q26.3-q28	EHHADH	Enzima bifuncional peroxisomal	Proteína que forma parte del conjunto de enzimas de la vía de la oxidación beta peroxisomal.	1%	2.50E-05	2.10E-02
19p13.2	ZNF709	Proteína dedo de zinc 709	Proteína hipotéticamente involucrada en la regulación transcripcional.	1%	3.90E-05	3.00E-02
19q13.2	ERF	ERF, Factor de transcripción	Potente represor transcripcional que se une al elemento H1 del promotor Ets2. Puede regular otros genes implicados en la	1%	8.20E-05	5.40E-02

		ión que contiene el dominio ETS	proliferación celular.			
10p12.31	MLLT10	Proteína AF-10	Esta proteína está vinculada en varios reordenamientos cromosómicos ligados a leucemias	1%	9.40E-05	5.90E-02
15q15.3	STRC	Estereocilina	Esencial para la formación de conectores superiores horizontales entre los estereocilios de las células ciliadas externas.	1%	1.70E-04	8.60E-02
18q11.2	GATA6	Factor de transcripción GATA-6	Activador transcripcional	1%	3.30E-06	4.00E-03
1p36.11	TCEB3	Elongina-A	Subunidad A de elongina, un factor general de elongación de la transcripción que aumenta el alargamiento de la transcripción de la ARN polimerasa II más allá de los sitios de detención codificados por la plantilla.	1%	5.00E-05	3.60E-02
17q23.3	ERN1	Serina / treonina-proteína quinasa / endoribonucleasa IRE1	Serina / treonina-proteína cinasa y endoribonucleasa que actúa como un sensor clave para la respuesta de la proteína desplegada del retículo endoplásmico	1%	8.30E-05	5.40E-02
15q13-q14	FMN1	Formina-1	Juega un papel en la formación de la unión adherente y en la polimerización de los cables de actina lineales.	1%	1.30E-04	7.60E-02
14q32.33	AKT1	RAC-alfa serina / treonina-proteína cinasa	AKT1 es una de las 3 serina / treonina-proteína quinasa (AKT1, AKT2 y AKT3) llamada quinasa AKT, y que regula muchos procesos, incluido el metabolismo, la proliferación, la supervivencia celular, el crecimiento y la angiogénesis.	1%	1.40E-04	7.60E-02

a: Extraído de Firebrowse, cohorte TCGA (N=498), los valores de p y q fueron descargados del análisis de Mutsig.  
b.: La función molecular se obtuvo de Uniprot.

La ontología predicha sugirió que los productos de los GSMs en CPp participan en procesos como la regulación de la transcripción desde el promotor de la ARN polimerasa II, en la regulación negativa del proceso apoptótico y en vías de señalización intracelular como la señalización del receptor del FGF (factor de crecimiento de fibroblastos), en la ruta de señalización por fosfatidilinositol y en la señalización en cascada de la proteína quinasa B (Tabla 7). Los productos de estos genes son principalmente factores de unión a DNA y a proteínas, localizadas en el núcleo.

**Tabla 7. Los 5 principales términos de ontología de genes (GO) según Genecodis para los genes significativamente mutados en Cpp.**

Categoría	Termino GO	Descripción	Conteo	p-valor	Genes
PB	GO:0045944	Regulación positiva de la transcripción del promotor de ARN polimerasa II.	9	2.30E-06	MLL2,GATA6,EOMES,TP53,MLLT10,CTNNB1,AKT1,FOXA1,MED12
	GO:0043066	Regulación negativa del proceso apoptótico.	7	2.36E-06	GATA6,PTEN,TP53,CDKN1B,CTNNB1,AKT1,ATM
	GO:0008543	vía de señalización del receptor del factor de crecimiento de fibroblastos	4	1.42E-04	PTEN,PIK3CA,CDKN1B,AKT1
	GO:0048015	señalización mediada por fosfatidilinositol	4	1.45E-04	PTEN,PIK3CA,CDKN1B,AKT1
	GO:0043491	señalización en cascada de la proteína quinasa B	3	2.58E-04	PTEN,PIK3CA,AKT1
FM	GO:0044212	Unión al ADN región reguladora de la transcripción	6	1.99E-06	MLL2,GATA6,ZFHX3,TP53,CTNNB1,FOXA1
	GO:0019899	unión a enzimas	5	5.96E-05	PTEN,ZFHX3,TP53,CTNNB1,AKT1
	GO:0003677	Unión al ADN	11	6.37E-05	MLL2,TNRC18,EOMES,TP53,MLLT10,CTNNB1,MLL3,TCEB3,FOXA1,ZMYM3,ATM
	GO:0003700	actividad de factor de transcripción de unión a ADN específica de secuencia	8	8.85E-05	GATA6,EOMES,ZFHX3,TP53,MLLT10,CTNNB1,ERF,FOXA1
	GO:0005515	Unión a proteínas	15	3.34E-04	SPOP,MLL2,MED15,GATA6,PTEN,PIK3CA,TP53,MLLT10,CDKN1B,CTNNB1,CTNNB1,MLL3,AKT1,SMG7,ATM
CC	GO:0005634	núcleo	23	3.90E-09	SPOP,MLL2,KDM6A,MED15,GATA6,PTEN,EOMES,ZFHX3,TP53,MLLT10,CDKN1B,CTNNB1,MLL3,AKT1,ERF,TCEB3,FOXA1,SMG7,EHHADH,FMN1,MED12,ZMYM3,ATM

GO:0005654	nucleoplasma	8	1.05E-04	MED15,GATA6,TP53,CDKN1B,AKT1,TCEB3,MED12,ATM
GO:0005667	complejo de factor de transcripción	4	1.98E-03	GATA6,ZFX3,CTNNB1,FOXA1
GO:0030027	Lamelipodio	3	2.47E-03	PIK3CA,CTNNB1,AKT1
GO:0035097	complejo de histona metiltransferasa	2	2.77E-03	MLL2,MLL3

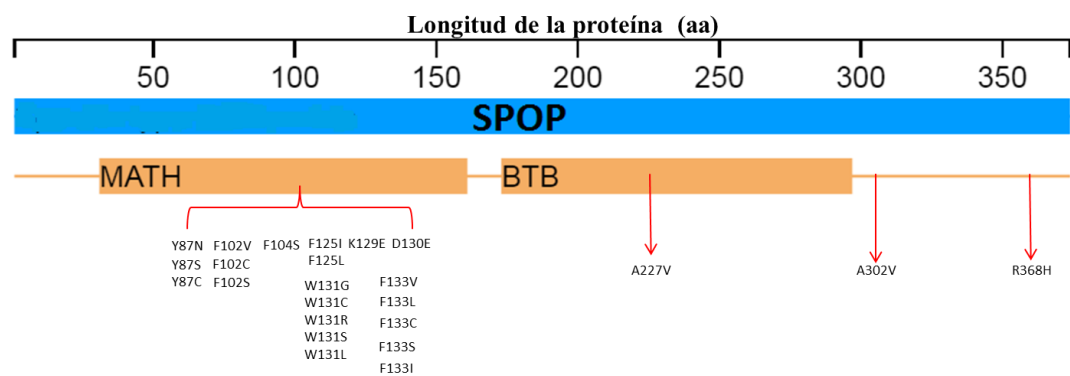
**GO Ontología de genes, PB Proceso biológico, FM Función molecular, CC Componente celular.**

El análisis de enriquecimiento de vías mostró que algunos GSMs, incluidos TP53 y PIK3CA, están vinculados no sólo con las rutas del desarrollo del CP sino también con otras neoplasias como el cáncer endometrial, cáncer de pulmón de células pequeñas y leucemia mieloide crónica (Tabla 8).

**Tabla 8. Las 5 primeras categorías más significativamente enriquecidas para los GSM en Cpp.**

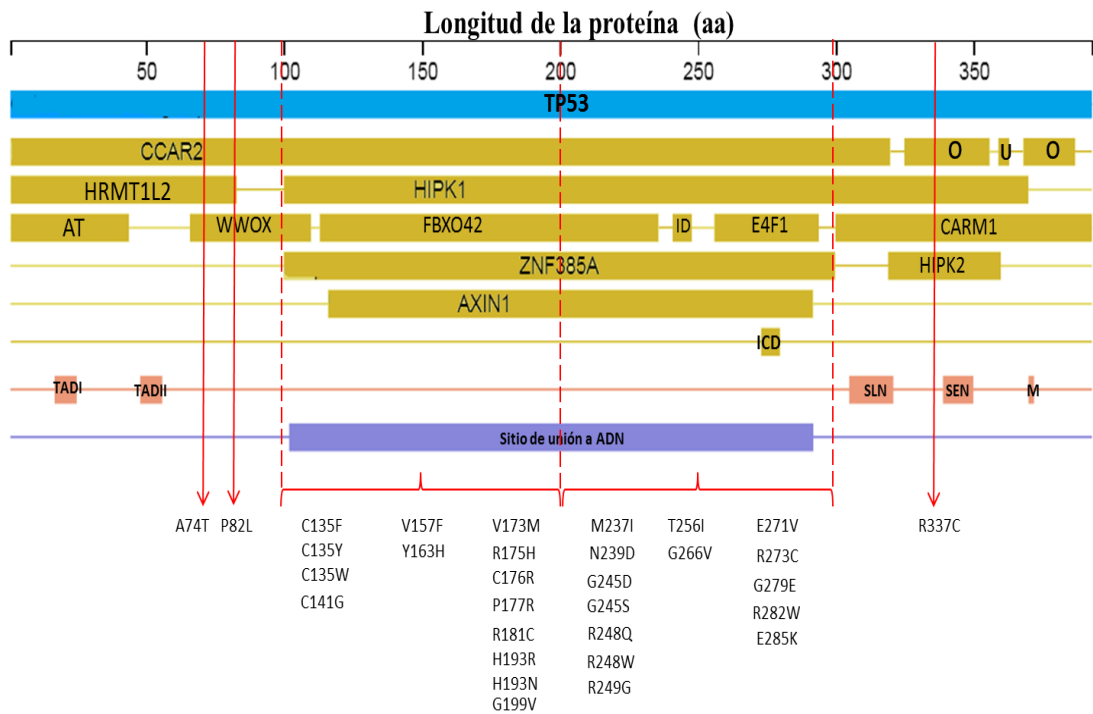
Termino KEGG	Nombre de la ruta	Conteo	p-valor	Genes
Kegg:05215	CP	6	2.72E-08	PIK3CA,PTEN,CDKN1B,AKT1,CTNNB1,TP53
Kegg:05213	Cáncer endometrial	5	7.62E-08	PIK3CA,PTEN,AKT1,CTNNB1,TP53
Kegg:05222	Cáncer de pulmón de células pequeñas	5	5.90E-07	PIK3CA,PTEN,CDKN1B,AKT1,TP53
Kegg:05214	Glioma	4	7.65E-06	PIK3CA,PTEN,AKT1,TP53
Kegg:05220	Leucemia mieloide crónica	4	8.67E-06	PIK3CA,CDKN1B,AKT1,TP53

Se investigaron los dos genes más mutados en CPp, con el fin de observar las zonas proteicas afectadas por el cambio aminoacídico. Según el análisis de mutaciones obtenido de cBioPortal para la cohorte TCGA (n=498), SPOP albergó gran parte de las mutaciones de cambio sentido (missense) en el dominio MATH, el cual media la interacción con sustratos de la proteína ubiquitina ligasa. (Figura 27). Sólo una mutación se encontró en el dominio BTB que media la dimerización y la interacción con CUL3.



**Figura 27. Distribución de mutaciones de cambio aminoacídico en la proteína SPOP en la cohorte TCGA(n=498), obtenido a través de cBioPortal para cáncer de próstata primario. Los números corresponden a la posición del aminoácido en el polipéptido, en naranja los dominios MATH y BTB. La línea horizontal superior indica la longitud de la proteína dada en aminoácidos (aa).**

TP53 presentó mutaciones missense principalmente en el dominio de unión al ADN así como en otros dominios de interacción intermolecular con proteínas como HIPK1, FBXO42, ZNF385A, CCAR2, AXIN1, E4F1, HRMT1L2 y WWOX. Sólo una alteración (R337C) ocurrió en la región de oligomerización (Figura 28).



**Figura 28. Distribución de mutaciones de cambio aminoacídico en la proteína TP53 en la cohorte TCGA(n=498), obtenido a través de cBioPortal para cáncer de próstata primario. Los números corresponden a la posición del aminoácido en el polipéptido. En amarillo, CCAR2, HRMT1L2, HIPK1, WWOX, FBXO42, HIPK2, ZNF385A, CARM1, AXIN1 y E4F1 indican las regiones en las cuales TP53 interactúa con dichas proteínas, región para la oligomerización (O), activación de la transcripción (AT), zona de interacción con el dominio 53BP2 SH3 (ID), zona de interacción con el DNA (ICD) y región de interacción con la proteína USP7 (U). En rosado los motivos de secuencia corta, TADI y TADII, señal de localización nuclear bipartita (SLN), señal de exportación nuclear (SEN), motivo KR-STA-K (M). En morado el sitio de unión a DNA. La línea horizontal superior indica la longitud de la proteína dada en aminoácidos (aa).**

### 6.2.1.2. Genes mutados en CPm

Empleando la base de datos de cBioPortal, se identificaron 79 GSMs en CPm. El análisis se condujo en la cohorte del estudio “**The Metastatic Prostate Cancer Project**”, considerando a 19 participantes con CPm. Cabe precisar que el estudio en mención no pertenece al grupo de estudios TCGA. Los resultados mostraron que TP53, al igual que en CPp, es el gen más mutado en pacientes con CPm, estando presente en el 53% del grupo de estudio. Los genes ZKSCAN1, codificante para una proteína hipotéticamente involucrada en la regulación de la transcripción y NBPF10, un miembro de la familia del punto de ruptura del neuroblastoma, estuvieron presentes en el 42% del grupo de estudio, cada uno de ellos (Tabla 9). SPOP, tal como en estadio primario (mutado en el 11% de la población), también estuvo presente en tumores metastásico pero en una frecuencia mayor (16%). Ningún otro GSM fue común a CPp y CPm.

**Tabla 9. Los 20 genes más significativamente mutados en CPm<sup>a</sup>**

Citobanda	Gene	Nombre	Función biológica	Frecuencia	Q-valor
17p13.1	TP53	Antígeno tumoral celular p53	Actúa como supresor de tumores en muchos tipos de neoplasias; induce la detención del crecimiento o la apoptosis según las condiciones fisiológicas y el tipo de célula.	53%	1.02E-12
7q22.1	ZKSCAN1	Proteína de dedo de zinc con dominios KRAB y SCAN 1	Proteína hipotéticamente involucrada en la regulación transcripcional.	42%	1.02E-12
1q21.1	NBPF10	Miembro 10 de la familia de punto de quiebre del neuroblastoma	Este gen es un miembro de la familia del punto de ruptura del neuroblastoma (NBPF), Codificado por una de las numerosas copias de los genes NBPF agrupados en las regiones p36, p12 y q21 del cromosoma 1.	42%	9.27E-09
19p13.11	ZNF253	Proteína de dedo de zinc 253	Hipotéticamente puede tener una actividad de represión transcripcional.	42%	8.33E-08
Xq27.2	MAGEC1	Antígeno asociado melanoma	C1 al Las proteínas de esta familia son antígenos específicos de tumores que pueden ser reconocidos por los linfocitos T citolíticos autólogos.	32%	3.42E-05
13q33.1	POGLUT2	Proteína O-glucosiltransferasa 2	proteína glucosiltransferasa que cataliza la transferencia de glucosa desde UDP-glucosa a un residuo de serina dentro de la secuencia de consenso péptido C-X-N-T-X-G-S-F-X-C	26%	1.24E-08

13q13.1	BRCA2	Proteína de susceptibilidad al cáncer de mama tipo 2	Participa en la reparación de roturas de ADN de doble hebra y / o en la recombinación homóloga. Se une a RAD51 y potencia la reparación del ADN recombinante al promover el ensamblaje de RAD51 en el ADN de una sola cadena.	26%	2.82E-07
6q13	MTO1	Homóloga de la proteína MTO1 h, mitocondrial	Participa en la modificación 5-carboximetilaminometilo (mnm5s2U34) de la base de uridina oscilante en los ARNt mitocondriales.	26%	1.98E-05
7p11.2	ZNF716	Proteína de dedo de zinc 716	Hipotéticamente asociada a la regulación transcripcional	26%	1.20E-03
1p22.1	EVI5	Homóloga de la proteína del Sitio de integración viral ecotrópico 5	Esta proteína puede actuar como un regulador del ciclo celular al actuar sobre FBXO5.	21%	1.75E-08
16q22.2	ZNF19	Proteína de dedo de zinc 19	Hipotéticamente esta proteína está asociada a la regulación transcripcional.	21%	9.84E-08
2p22.3	YIPF4	Proteína YIPF4	Participa en el mantenimiento de la estructura de Golgi.	21%	1.40E-07
18q22.1	CCDC102 B	Proteína 102B que contiene el dominio de bobina en espiral	La hernia diafragmática, congénita está asociada a esta proteína	21%	3.24E-04
12q21.2	GLIPR1L2	Proteína 2 tipo GLIPR1	Este gen codifica un miembro la superfamilia 1 relacionada con la patogénesis, la proteína secretora rica en cisteína y el antígeno 5.	21%	5.53E-04
1p36.12	NBPF3	Miembro 3 del NBPF	Esta proteína es miembro de las proteínas de la familia del punto de ruptura del neuroblastoma, un grupo de proteínas que podrían estar involucradas en el desarrollo del sistema nervioso.	21%	1.12E-03
19q13.43	ZNF776	Proteína de dedo de zinc 776	Proteína involucrada en la regulación transcripcional	21%	1.95E-03
19p12	ZNF708	Proteína de dedo de zinc 708	Proteína involucrada en la regulación transcripcional	21%	8.21E-02
1q42.13	ZNF678	Proteína de dedo de zinc 678	Proteína involucrada en la regulación transcripcional	21%	8.21E-02
19q13.43	ZNF814	Proteína de dedo de zinc 814	Proteína involucrada en la regulación transcripcional	21%	8.29E-02
8q21.2	CA3	Anhidrasa carbónica 3	Proteína motora de microtúbulos que se une a los microtúbulos con alta afinidad a través de cada heterodímero de tubulina y tiene una actividad ATPasa.	16%	1.60E-06

a: Extraído de cBioPortal, The Metastatic Prostate Cancer Project (n=19)

b.: La función molecular se obtuvo de Uniprot.

La ontología predicha para el grupo de GSMs en CPM mostró que los productos de estos genes se involucran mayoritariamente a la regulación de la transcripción, mecanismos de reparación al daño del ADN y en la regulación de la actividad de

la fosfoproteína fosfatasa. Siendo principalmente proteínas de unión tanto a ADN, como a otras proteínas y a iones metálicos, localizadas tanto intracelularmente como núcleo (Tabla 10). Tanto en CPp como en CPm se han identificado genes que están vinculados a procesos de regulación de la transcripción, sin embargo en CPp adicionalmente se han determinado genes como PTEN, PIK3CA, CDKN1B y AKT1 implicados en proliferación y supervivencia, mientras que en CPm TP53 y BRAC2 destacan por su participación en procesos de respuesta al daño del ADN.

**Tabla 10. Los 5 principales términos de ontología de genes (GO) según Genecodis para los genes significativamente mutado en CPm.**

Categoría	Termino GO	Descripción	Conteo	p-valor	Genes
PB	GO:0006355	Regulación de la transcripción, dependiente del ADN.	15	6.71E-04	ZNF814,RCAN1,TP53,ZNF776,EZH1,SOX30,ASCC1,ZNF708,ZKSCAN1,ZNF678,ZNF615,ZNF253,ZNF254,ZNF716,ZNF19
	GO:0043666	Regulación de la actividad fosfoproteína fosfatasa.	2	1.81E-02	RCAN1,HSP90B1
	GO:0045947	Regulación negativa de la iniciación traslacional.	2	2.36E-02	FMR1,RBM4
	GO:0006978	Respuesta al daño del ADN, transducción de señales por mediador de clase p53 que resulta en la transcripción de mediador de clase p21	2	2.50E-02	BRCA2,TP53
	GO:0042771	Respuesta al daño del ADN, transducción de señales por mediador de la clase p53 que resulta en la inducción de la apoptosis	2	3.33E-02	BRCA2,TP53
FM	GO:0003677	Unión al ADN	18	8.30E-06	BRCA2,DEAF1,HIF1,STRBP,RCAN1,TP53,ZNF776,EZH1,SOX30,ZNF708,ZKSCAN1,ZNF678,ZNF615,ZNF253,ZNF254,ZNF716,ZNF19,NUP153
	GO:0008270	Unión a iones de zinc	17	7.34E-05	DEAF1,ZNF814,TP53,ZNF776,CA3,ZN

					F708,ZKSCAN1,ZNF678,ZNF615,ZNF253,ZNF254,ANKMY2,ZNF716,ZNF19,TRIM60,NUP153,RBM4
					DEAF1,ATP9B,ZNF814,TP53,ZNF776,CA3,ZNF708,ZKSCAN1,HECTD1,ZNF678,ZNF615,ZNF253,ME1,ZNF254,ANKMY2,ZNF716,ZNF19,TRIM60,NUP153,RBM4
					BRCA2,MAGEC1,SPOP,DEAF1,MUC7,RHOBTB3,TNIK,ARF4,KIF14,RCAN1,TP53,EVI5,RP1L1,MPHOSPH10,HECTD1,COIL,EIF3H,FMR1,NUP153,SLC24A1,GOLGA4,RBM4,HSP90B1,NIN
					ZNF814,ZNF776,DX60L,ZKSCAN1,ZNF678,ZNF615,ZNF254,RBM4
					GNL2,ARF4,ZNF814,STRBP,ZNF776,EVI5,ZNF708,ZKSCAN1,HECTD1,ZNF678,ZNF615,ZNF253,ZNF254,ZNF716,ZNF19,TRIM60,NUP153
CC	GO:0005622	intracelular	17	1.14E-04	
					BRCA2,SPOP,DEAF1,TNIK,SESN3,GNL2,H1FO,KIF14,STRBP,RCAN1,TP53,ZNF776,EVI5,EZH1,MPHOSPH10,SOX30,ZNF708,ZKSCAN1,ZNF678,ZNF615,ZNF253,COIL,ZNF254,ZNF716,ZNF19,FMR1,NUP153,RBM4
	GO:0005634	núcleo	28	4.27E-04	
					BRCA2,NBPF3,INP5J,DEAF1,TNIK,ARF4,KIF14,KLHL1,STRBP,RCAN1,TP53,CA3,EVI5,RP1L1,BFSP1,ASCC1,ME1,EIF3H,FMR1,NUP153,GOLGA4,PPP1R
	GO:0005737	citoplasma	24	1.20E-02	

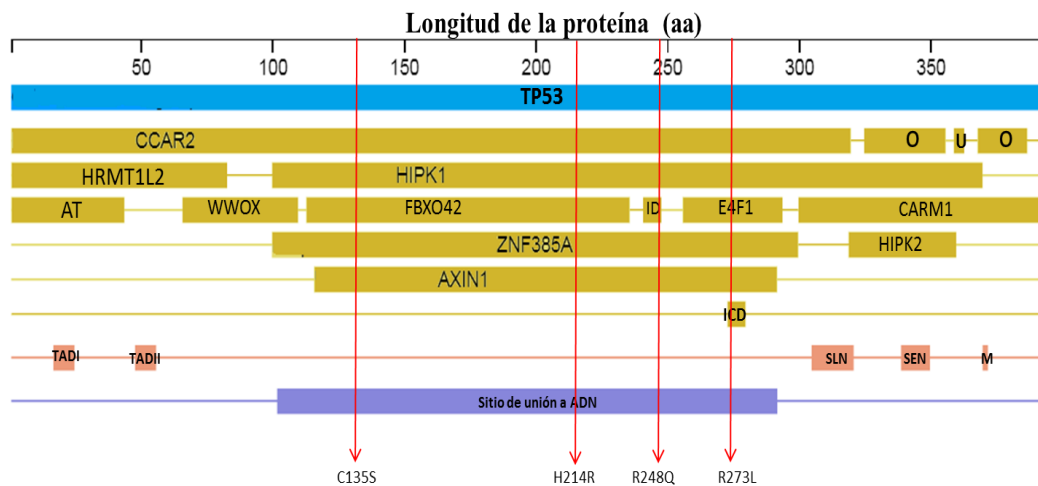
				9A,RBM4,NIN
GO:0005730	nucléolo	10	3.38E-02	BRCA2,TNIK,GNL2,TP53,EZH1,MPHOSPH10,COIL,FMR1,NUP153,RBM4
GO:0033593	Complejo BRCA2-MAGE-D1	1	3.83E-02	BRCA2

La evaluación de vías enriquecidas mostró que tanto la biogénesis del ribosoma como la fototransducción están vinculadas con algunos GSMs en CP metástasicos, tal como se muestra en la Tabla 11.

**Tabla 11. Resultados del análisis de enriquecimiento de vías para los GSM en Cpm.**

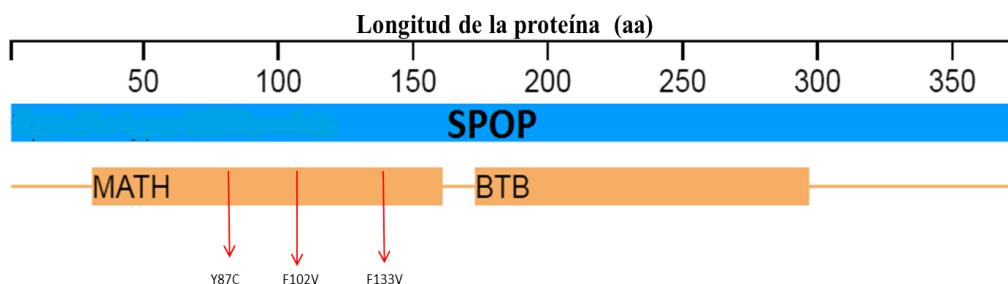
Termino KEGG	Nombre de la ruta	Conteo	p-valor	Genes
Kegg:03008	Biogénesis del ribosoma en eucariotas.	3	3.17E-02	GNL2,RIOK2,MPHOSPH10
Kegg:04744	Fototransducción	2	4.40E-02	SLC24A1,CNGA1

En la cohorte de Cpm, las mutaciones missense de TP53 estuvieron presentes únicamente en el dominio de unión al ADN (figura 29). La mutación R248Q estuvo presente tanto en cáncer de próstata primario como metastásico, ubicándose en la regiones de interacción con las proteínas CCAR2, HIPK1, ZNF385A, AXIN1 y con el dominio 53BP2 SH3. De las cuatro mutaciones encontradas en TP53 para Cpm, solo la variante R248Q fue identificada como una variante probablemente dañina y desestabilizadora.



**Figura 29. Distribución de mutaciones de cambio aminoacídico en la proteína TP53 en la cohorte The Metastatic Prostate Cancer Project (n=19), obtenido a través de cBioPortal para cáncer de próstata metastásico. Los números corresponden a la posición del aminoácido en el polipéptido. En amarillo, CCAR2, HRMT1L2, HIPK1, WWOX, FBXO42, HIPK2, ZNF385A, CARM1, AXIN1 y E4F1 indican las regiones en las cuales TP53 interactúa con dichas proteínas, región para la oligomerización (O), activación de la transcripción (AT), zona de interacción con el dominio 53BP2 SH3 (ID), zona de interacción con el DNA (ICD) y región de interacción con la proteína USP7 (U). En rosado los motivos de secuencia corta, TADI y TADII, señal de localización nuclear bipartita (SLN), señal de exportación nuclear (SEN), motivo KR-STA-K (M). En morado el sitio de unión a DNA. La línea horizontal superior indica la longitud de la proteína dada en aminoácidos (aa).**

En el caso de SPOP, las 3 mutaciones identificadas para CPm se localizaron en el dominio MATH. Las mutaciones F102V y F133V estuvieron presentes tanto en el CPp como en el estadio de metástasis. La variante Y87C fue la única mutación propia para CPm que fue encontrada en la cohorte analizada (Figura 30).



**Figura 30. Distribución de mutaciones de cambio aminoacídico en la proteína SPOP obtenido a través de cBioPortal para CPM. Los números corresponden a la posición del aminoácido en el polipéptido, en naranja los dominios MATH y BTB. La línea horizontal superior indica la longitud de la proteína dada en aminoácidos (aa).**

## 6.2.2. Genes con alteración en el número de copias.

### 6.2.2.1 Genes delecionados

Se utilizó cBioPortal para estimar la frecuencia de aparición de genes amplificados y delecionados, tanto en el CPp como el CPM, a través del análisis de 12 estudios globales (Tabla 12).

**Tabla 12. Estudios empleados en la identificación de genes frecuentemente amplificados y delecionados en CP, disponibles en cBioPortal.**

Estudio	Número de Pacientes	Número de Muestras
MSK-IMPACT Clinical Sequencing Cohort in Prostate Cancer (MSK, JCO Precision Oncology 2017)	451	504
Prostate Adenocarcinoma (Broad/Cornell, Cell 2013)	57	57
Prostate Adenocarcinoma (Broad/Cornell, Nat Genet 2012)	112	112
Prostate Adenocarcinoma (EurUrol, 2017)	65	65
Prostate Adenocarcinoma (Fred Hutchinson CRC, Nat Med 2016)	63	176
Prostate Adenocarcinoma (MSKCC, Cancer Cell 2010)	216	216
Prostate Adenocarcinoma (MSKCC/DFCI, Nature Genetics 2018)	1013	1013
Prostate Adenocarcinoma (TCGA, Provisional)	498	499
Prostate Adenocarcinoma CNA study (MSKCC, PNAS 2014)	104	104
Metastatic Prostate Cancer, SU2C/PCF Dream Team (Robinson et al., Cell 2015)	150	150
Metastatic Prostate Adenocarcinoma (MCTP, Nature 2012)	59	61

Un total de 1575 pacientes tuvieron la información genómica de ANC para Cpp. Para tumores primarios, el gen más frecuentemente deleciónado fue el supresor de tumor PTEN, que también resultó ser un GSM, estando presente en el 11% de la población. A continuación, los genes MAP3K7, cuyo producto es una enzima de la vía de transducción de señales de mitógeno MAP quinasa, y EPHA7, codante para un receptor tirosina quinasa involucrado en la comunicación celular, presentes ambos en el 10% de la población, respectivamente (Tabla 13). Interesantemente, un total de 11 genes deleciónados se localizaron en el cromosoma 11 brazo q.

**Tabla 13. Genes frecuentemente deleciónados en Cpp**

Citobanda	Símbolo del gen	Nombre	Función molecular	Frecuencia
10q23.31	PTEN	fosfatidilinositol-3,4,5-trisfosfato 3-fosfatasa	Supresor de tumor.	11%
6q15	MAP3K7	Proteína quinasa quinasa quinasa 7 activada por mitógeno	Es una proteína Serina / treonina quinasa asociada a la vía de transducción de señales de la MAP quinasa.	10%
6q16.1	EPHA7	Efrina tipo A receptor 7	El receptor tirosina quinasa se une a los ligandos de la familia Ephrin-A anclados en GPI que residen en células adyacentes, lo que lleva a la señalización bidireccional dependiente del contacto hacia las células vecinas.	10%
6q15	PNRC1	Coactivador del receptor nuclear 1 rico en prolina	Coactivador del receptor nuclear	10%
6q21	PRDM1	Proteína de dedo de zinc 1 Dominio PR	Factor de transcripción que media un programa transcripcional en varios tipos de linfocitos T residentes de tejido inmune innato y adaptativo	8%
6q21	FOXO3	Proteína O3 de caja de horquilla	Activador transcripcional que desencadena la apoptosis en ausencia de factores de supervivencia, incluida la muerte de células neuronales por estrés oxidativo.	8%
6q21	SESN1	Sestrin-1	Proteína que actúa como un sensor de leucina intracelular, regulando la vía de señalización TORC1.	8%
6q21	AK9	Adenilato quinasa 9	Participa en el mantenimiento de la homeostasis de los nucleótidos celulares mediante la catalización de la interconversión de los fosfatos de nucleósidos.	8%

6q21	FYN	Tirosina-proteína quinasa Fyn	Tirosina-proteína quinasa no receptora que desempeña un papel en muchos procesos biológicos, incluida la regulación del crecimiento y la supervivencia celular, la adhesión celular, la señalización mediada por integrinas, la remodelación del citoesqueleto, la motilidad celular, la respuesta inmune y la guía de los axones. Responsable de la desacetilación de los residuos de lisina en la parte N-terminal de las histonas centrales (H2A, H2B, H3 y H4).	8%
6q21	HDAC2	Histona desacetilasa 2	Hipotéticamente esencial para desarrollo esquelético postnatal.	8%
6q21	WISP3	Proteína 3 de la vía de señalización inducible por WNT1	Factor de transcripción probable involucrado en el desarrollo del huso muscular.	8%
8p21.3	EGR3	Proteína 3 de respuesta temprana de crecimiento	Es un proteína que actúa como factor de modelación de la cromatina dependiente de ATP.	7%
5q15-q21.1	CHD1	Proteína 1 de unión a ADN-cromodominio-helicasa	Proteína que forma parte del componente estructural de los músculos estriados.	7%
1q42.13	OBSCN	Obscurin	Es una proteína que regula la división celular y actúa como un supresor de tumores.	67%
13q14.2	RB1	Proteína asociada al retinoblastoma	Juega un papel en el tráfico de proteínas intracelulares y la degradación.	7%
6q22.1	GOPC	Proteína que contiene el motivo de bobina enrollada y PDZ asociada al aparato de Golgi	Proteína de unión a la actina. Juega un papel en la activación de las células T en respuesta a la coestimulación a través de TCR / CD3 y CD28 o CD28	6%
13q14.13	LCP1	Plastin-2	Acetiltransferasa requerida para el establecimiento de cohesión cromátida hermana	6%
8p21.1	ESCO2	N-acetiltransferasa ESCO2	Proteínas potenciales de la superficie celular que se unen e internalizan ligandos en el proceso de endocitosis mediada por receptores.	6%
2q22.1-q22.2	LRP1B	Proteína 1B relacionada con el receptor de lipoproteínas de baja densidad	Este receptor contiene un dominio de muerte. Hipotéticamente regula la fisiológica de la muerte celular programada.	6%
10q23.31	FAS	Miembro 6 de la superfamilia del receptor del factor de necrosis tumoral		6%

En el CP metastásico, el número de pacientes con información de ANC fue de 705 individuos. Al igual que en CPP, PTEN fue el gen más frecuentemente deletado, estando presente en el 11% de los pacientes. También estuvieron deletados los genes RB1, regulador de la división celular y FAS, una proteína

de membrana tipo II miembro de la familia del factor de necrosis tumoral, presentes en 9% y 8% de la población, respectivamente (Tabla 14). Los genes TP53 y BRCA2, que resultaron GSM en CPm, también presentaron deleciones en CPm.

**Tabla 14. Genes frecuentemente deletados en CPm.**

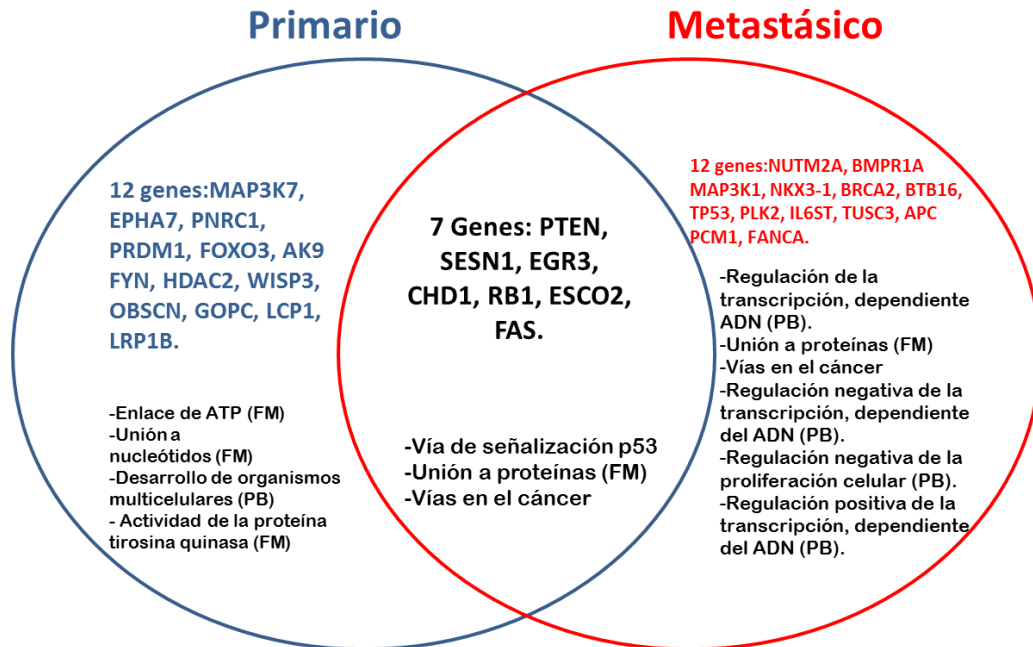
Citobanda	Símbolo del gen	Nombre	Función molecular	Frecuencia
10q23.31	PTEN	fosfatidilinositol-3,4,5-trisfosfato 3-fosfatasa	Supresor de tumor.	25%
13q14.2	RB1	Proteína asociada al retinoblastoma	Regulador clave de la entrada en la división celular que actúa como un supresor de tumores.	9%
10q23.31	FAS	Miembro 6 de la superfamilia del receptor del factor de necrosis tumoral	Receptor para TNFSF6 / FASLG.	8%
10q23.2	NUTM2A	Miembro 2A de la familia NUT	Las enfermedades asociadas con NUTM2A incluyen el sarcoma del estroma endometrial y el nódulo del estroma endometrial. Factor de remodelación de la cromatina dependiente de ATP que funciona como componente de reconocimiento de sustrato del complejo de acetilación de histonas reguladoras de la transcripción (HAT) SAGA.	8%
5q15-q21.1	CHD1	Proteína 1 de unión a ADN-cromodominio-helicasa	En la unión del ligando, forma un complejo receptor que consiste en dos serina / treonina quinasa transmembrana de tipo II y dos de tipo I.	6%
10q23.2	BMPR1A	Receptor de proteína morfogenética ósea tipo 1 <sup>a</sup>	Serina / treonina quinasa que actúa como un componente esencial de la vía de transducción de señales de la MAP quinasa.	5%
5q11.2	MAP3K1	Proteína quinasa quinasa quinasa 7 activada por mitógeno	Factor de transcripción involucrado en el desarrollo del huso muscular.	5%
8p21.3	EGR3	Proteína 3 de respuesta temprana de crecimiento	Factor de transcripción, que se une preferentemente a la secuencia de consenso 5'-TAAGT [AG] -3' y puede comportarse como un represor transcripcional.	5%
8p21.2	NKX3-1	Proteína Homeobox Nkx-3.1	Participa en la reparación de roturas de doble cadena y / o recombinación homóloga	5%
13q13.1	BRCA2	Proteína de susceptibilidad tipo 2 al cáncer de mama		
11q23.2	ZBTB16	proteína 16 dedo de zinc que contiene el dominio BTB	Actúa como represor transcripcional.	4%

---

17p13.1	TP53	Antígeno tumoral celular p53	Actúa como supresor de tumores en muchos tipos de tumores; induce la detención del crecimiento o la apoptosis según las circunstancias fisiológicas y el tipo de célula.	4%
5q11.2	PLK2	Serina / treonina-proteína quinasa PLK2	Supresor de tumores serina / treonina-proteína quinasa involucrada en la plasticidad sináptica, la duplicación del centriolo y la transición de fase G1 / S.	4%
5q11.2	IL6ST	Subunidad beta del receptor de Interleucina 6	La proteína codificada por este gen actuaría como un transductor de señal de muchas citoquinas.	4%
8p22	TUSC3	Candidato 3 de supresor d tumores	Esta proteína se localiza en el retículo endoplásmico y actúa como un componente del complejo de oligosaccharil transferasa que es responsable de la glicosilación de la proteína unida a N	4%
8p21.1	ESCO2	N-acetiltransferasa ESCO2	Acetiltransferasa requerida para el establecimiento de cohesión cromátida hermana	4%
6q21	SESN1	Sestrin-1	Funciona como un sensor de leucina intracelular que regula negativamente la vía de señalización TORC1 a través del complejo GATOR.	4%
5q22.2	APC	APC, Regulador de la vía de señalización WNT	Supresor de tumor.	4%
8p22	PCM1	Proteína de material pericentriolar 1	Proteína necesaria para el ensamblaje y la función del centrosoma.	3%
16q24.3	FANCA	Grupo A de Complementación de anemia Fanconi	Puede estar involucrado en la reparación de enlaces cruzados de ADN entre cadenas y en el mantenimiento de la estabilidad cromosómica normal.	3%

---

A continuación, se compararon los genes deletados de los grupos de CPp y CPm. Se observó que 7 genes están comúnmente deletados en ambas poblaciones de estudio incluyendo PTEN, SESN1, EGR3, CHD1, RB1, ESCO2, FAS. El análisis de ontología (GO) mostró que los productos de estos genes están involucrados en la vía de señalización de p53 y vías del cáncer, siendo principalmente productos de unión a otras proteínas y de localización nuclear (figura 31). Por otro lado, 12 genes resultaron únicos a CPp involucrándose en el desarrollo de organismos multicelulares, siendo proteínas de unión a ATP, nucleótidos y con actividad quinasa. De igual forma, 12 genes fueron únicos a CPm; los cuales se asocian a mecanismos de regulación de la transcripción y de la proliferación celular, siendo factores de unión a proteínas principalmente (figura 31).



**Figura 31. Diagrama de Venn que representa el número de genes frecuentemente deletados y términos GO comunes a CPp y CPm. GO Ontología de genes, PB Proceso biológico, FM Función molecular, CC Componente celular.**

### 6.2.2.2. Genes Amplificados

La frecuencia de aparición de genes amplificados fue evaluada de igual forma que la de los genes delecionados. Para Cpp, el gen más frecuentemente amplificado fue NDRG1, una proteína que participa en las respuestas hormonales, crecimiento y diferenciación celular; presente en el 6.03% de la población. Seguido por AGO2, una proteína involucrada en el silenciamiento génico mediado por ARN, y PTK2, una proteína que desempeña un papel importante en procesos como la regulación de la migración, adhesión y diseminación celular, ambos genes presentes en el 5.14% de los pacientes (Tabla 15). Se observó que un total de 14 genes amplificados estaban localizados en el cromosoma 8 brazo q.

**Tabla 15. Genes frecuentemente amplificados en Cpp**

Citobanda	Símbolo del Gen	Nombre	Función biológica	Frecuencia
8q24.22	NDRG1	Proteína NDRG1	Proteína sensible al estrés involucrada en las respuestas hormonales, el crecimiento celular y la diferenciación.	6%
8q24.3	AGO2	Proteína argonauta-2	Requerido para el silenciamiento génico mediado por ARN.	5%
8q24.3	PTK2	Quinasa 1 de adherencia focal	Proteína tirosina quinasa no receptora que desempeña un papel esencial en muchos procesos biológicos como la regulación de la migración celular, la adhesión, la diseminación, etc.	5%
8q21.13	HEY1	Factor de transcripción de la familia BHLH, relacionado a Hes, con el motivo 1 YRPW	Represor transcripcional	5%
8q21.13	PAG1	Fosfoproteína asociada con microdominios enriquecidos en glicosfingolípidos 1	Regula negativamente la señalización mediada por TCR (receptor de antígeno de células T) en células T y la señalización mediada por FCER1 (receptor de inmunoglobulina de alta afinidad épsilon) en mastocitos.	5%
8q24.21	MYC	Proteína protooncogénica myc	Proteína que actúa como un factor de transcripción uniéndose al ADN.	5%
8q21.11	ELOC	Elongin-c	Es un factor de alargamiento de la transcripción general que aumenta el alargamiento de la transcripción de la ARN polimerasa II más allá de los sitios de detención codificados por plantilla	5%
8q24.3	RECQL4	Helicasa Q4 dependiente de ATP	ATPasa dependiente de ADN	4%

1p21.1	COL11A1	Cadena de colágeno alfa-1 (XI)	Puede jugar un papel importante en la fibrillogénesis al controlar el crecimiento lateral de las fibrillas de colágeno II.	4%
8q24.3	PLEC	Plectina	Proteína que une los filamentos intermedios con microtúbulos y microfilamentos.	4%
8q13.3	PRDM14	Proteína 14 de dedo de zinc de dominio PR	Factor de transcripción que tiene roles tanto positivos como negativos de transcripción.	4%
8q13.3	NCOA2	Coactivador 2 de receptores nucleares	Es una proteína coactivadora de la expresión de genes receptores de esteroides y receptores nucleares.	4%
8q24.3	EPPK1	Epiplakin	Proteína enlazadora del citoesqueleto que se conecta a los filamentos intermedios y controla su reorganización en respuesta al estrés	4%
8q13.2	PREX2	Factor 2 de intercambio de Rac dependiente de Fosfatidilinositol 3,4,5 trisfosfato	Funciona como un factor de intercambio de nucleótidos de guanina RAC1 (GEF), que activa las proteínas Rac al intercambiar el PIB unido por el GTP libre.	4%
8q21.3	NBN	Nibrina	Es una proteína que forma parte del componente del complejo MRN, involucrado en respuestas a daño al ADN.	4%
6p22.2	HIST1H2BC	Histone Cluster 1C Miembro C de la familia H2B	Componente central del nucleosoma.	4%
6p22.2	HIST1H2AC	Histona H2A tipo 1-C	Componente central del nucleosoma.	4%
6p22.2	HIST1H1E	Histona H1.4	La proteína histona H1 se une al ADN enlazador entre los nucleosomas que forman la estructura macromolecular conocida como la fibra de cromatina.	4%
6p22.2	HIST1H3D	Histona H3.1	Componente central del nucleosoma	4%
3q25.1	WWTR1	Proteína 1 regulador de la transcripción que contiene el dominio WW	Coactivador transcripcional	4%

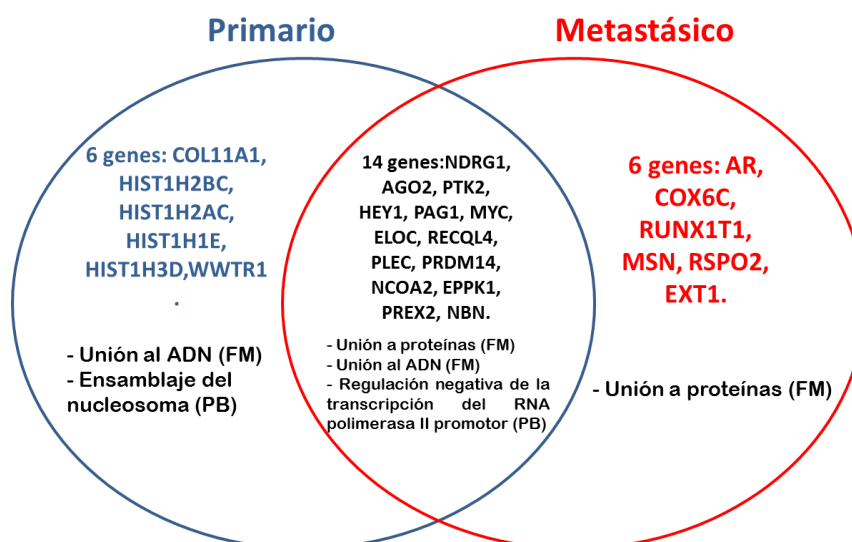
En CP metastásico, el receptor de andrógeno fue el gen más frecuentemente amplificado, con una presencia en el 39.72% de los pacientes. Seguido por los factores de transcripción ELOC y MYC, presentes en el 14.33% y 12.20% de la población respectivamente (Tabla 16). Nuevamente, la mayoría de genes amplificados pertenecían al cromosoma 8 brazo q, adicionalmente se pudo observar que el Cpm presentó frecuencias más altas de genes amplificados y deletados comparado a las frecuencias reportadas en Cpp.

**Tabla 16. Genes frecuentemente amplificados en CPM.**

Citobanda	Símbolo del Gen	Nombre	Función biológica	Frecuencia
Xq12	AR	Receptor de andrógeno	Los receptores de hormonas esteroides son factores de transcripción activados por ligando que regulan la expresión de genes eucarióticos y afectan la proliferación y diferenciación celular en los tejidos diana.	40%
8q21.11	ELOC	Elongin-c	Es un factor de alargamiento de la transcripción general que aumenta el alargamiento de la transcripción de la ARN polimerasa II más allá de los sitios de detención codificados por plantilla	14%
8q24.21	MYC	Proteína protooncogénica myc	Es un factor de transcripción que se añade el ADN de una manera no específica.	12%
8q13.3	NCOA2	Coactivador 2 de receptores nucleares	Coactivador transcripcional para receptores de esteroides y receptores nucleares.	12%
8q13.3	PRDM14	Proteína 14 de dedo de zinc de dominio PR	Factor de transcripción que tiene roles tanto positivos como negativos de transcripción.	12%
8q21.13	PAG1	Fosfoproteína asociada con microdominios enriquecidos en glicosfingolípidos 1	Regula negativamente la señalización mediada por TCR (receptor de antígeno de células T) en células T y la señalización mediada por FCER1 (receptor de inmunoglobulina de alta afinidad épsilon) en mastocitos.	11%
8q21.13	HEY1	Factor de transcripción de la familia BHLH, relacionado a Hes, con el motivo 1 YRPW	Represor transcripcional	11%
8q24.22	NDRG1	Proteína NDRG1	Proteína sensible al estrés involucrada en las respuestas hormonales, el crecimiento celular y la diferenciación.	11%
8q22.2	COX6C	Citocromo c oxidasa subunidad 6C	Actividad de la citocromo-c oxidasa	11%
8q21.3	NBN	Nibrina	Componente del MRE11-RAD50-NBN (complejo MRN) que desempeña un papel fundamental en la respuesta celular al daño del ADN y al mantenimiento de la integridad del cromosoma.	11%
8q24.3	RECQL4	Helicasa Q4 dependiente de ATP	ATPasa dependiente de ADN	10%
8q13.2	PREX2	Factor 2 de intercambio de Rac dependiente de Fosfatidilinositol 3,4,5 trisfosfato	Funciona como un factor de intercambio de nucleótidos de guanina RAC1 (GEF), que activa las proteínas Rac al intercambiar el PIB unido por el GTP libre.	10%
8q21.3	RUNX1T1	Proteína CBFA2T1	Corepresor transcripcional que facilita la represión transcripcional a través de su asociación con factores de transcripción que se unen al ADN y el reclutamiento de otros corepresores y enzimas modificadoras de histonas.	10%
8q24.3	AGO2	Proteína argonauta-2	Requerido para el silenciamiento génico mediado por ARN.	10%

8q24.3	PTK2	Quinasa 1 de adherencia focal	Proteína tirosina quinasa no receptora que desempeña un papel esencial en muchos procesos biológicos como la regulación de la migración celular, la adhesión, la diseminación, etc.	9%
Xq12	MSN	Moesin	Probablemente esté involucrado en las conexiones de las estructuras citoesqueléticas principales a la membrana plasmática.	9%
8q24.3	EPPK1	Epiplakin	Proteína enlazadora del citoesqueleto que se conecta a los filamentos intermedios y controla su reorganización en respuesta al estrés	9%
8q24.3	PLEC	Plectina	Proteína que une los filamentos intermedios con microtúbulos y microfilamentos. .	9%
8q23.1	RSPO2	R-espondina-2	Proteína que activa la vía de señalización Wnt .	9%
8q24.11	EXT1	Exostosina-1	Glicosiltransferasa requerida para la biosíntesis de sulfato de heparano.	9%

Se hallaron 11 genes comúnmente amplificados tanto en CPp como en CPM, incluyendo a NDRG1, PTK2, HEY1, PAG1, MYC, PLEC, PRDM14, NCOA2, EPPK1, PREX2, NBN (Fig. 32). Según la ontología predicha, los productos de estos genes participan en la regulación de la transcripción desde el promotor de la ARN polimerasa II, siendo factores de unión a DNA y proteínas, principalmente nucleares (figura 16).



**Figura 32. Diagrama de Venn que representa el número de genes frecuentemente amplificados y términos GO comunes a CPp y metastásico. GO Ontología de genes, PB Proceso biológico, FM Función molecular. CC Componente celular.**

### 6.2.3. Genes con expresión alterada.

El tercer tipo de aberración genética investigada en CP fue la expresión diferencial de genes. Para ello, se utilizó Gepia, recurso que alberga 492 muestra de pacientes con CP de TCGA, y bajo los parámetros establecidos en la sección 4.6.3., se identificaron genes sobre-expresados y sub-expresados para CPp. No se llevó a cabo el análisis en CPM debido a que no hubo información disponible de la expresión genética para ningún gen.

#### 6.2.3.1. Genes sobreexpresados

Empleando Gepia se identificaron 227 genes sobre-expresados en CPp, los cuales presentaron un  $\log FC > 1.5$  (Tabla 17). El gen más sobre-expresado fue PCA3 con un  $\log FC > 5$ , cuyo producto es un ARN largo no codificante que anteriormente ha sido reportado como altamente expresado en este tipo de neoplasia (49). Seguido por AMACR, un gen involucrado en la racemización de ésteres de CoA de ácidos grasos y DLX1, cuyo producto está asociado a la activación o represión transcripcional. También se identificaron pseudogenes altamente expresados como RP11-40C6.2, MTND4P12 y PCAT14; algunos de ellos hipotéticamente generan ARNs largos no codificantes.

**Tabla 17. Los 20 genes más significativamente sobre-expresados en CPp.**

Símbolo del gen	Nombre del Gen	Función del gen	$\log_2(FC)$	Q- valor
PCA3*	Asociado a CP 3	Este gen produce un ARN largo no codificante.	5.596	1.38E-50
AMACR	Alfa-metilacil-CoA racemase Proteína	Proteína involucrada en la racemización de ésteres de CoA.	4.686	1.53E-74
DLX1	Homeobox DLX-1	Juega un papel como activador o represor transcripcional	4.159	4.92E-56
OR51E2	Receptor olfativo 51E2	Probable receptor acoplado a la proteína G que se activa por el acetato de propionato y ácidos grasos de cadena corta	4.115	4.17E-45
PCAT14*	Transcrito 14 asociado al CP	Está afiliado a la clase de ARN no codificante.	4.095	5.02E-38
GOLM1	Proteína 1 de la	Proteína de respuesta celular a la infección viral.	3.286	4.23E-

	membrana de Golgi			84
HPN	Serina proteasa hepsina	Es una enzima que rompe los sustratos extracelulares.	3.225	1.99E-90
GLYATL1	Proteína 1 similar a Glicina N-aciltransferasa	Es una enzima Aciltransferasa que proporciona un grupo acilo al extremo N de la glutamina.	3.223	4.92E-48
ACSM1	Acil-coenzima A sintetasa ACSM1, mitocondrial	Enzima activadora de lipoato dependiente de GTP que genera el sustrato para la lipiltransferasa	3.153	3.03E-34
TRGC1	Receptor de células T gamma constante 1	Región constante de la cadena gamma del receptor de células T (TR) que participa en el reconocimiento del antígeno	2.99	1.20E-60
TMEFF2	Tomoregulina-2	Puede ser un factor de sobrevivencia para las neuronas hipocámpales y mesencefálicas.	2.984	1.13E-21
GJB1	Proteína beta-1 de unión a brecha	Una unión de brecha consiste en un grupo de pares de canales transmembrana muy compactos a través de los cuales los materiales de bajo peso molecular se difunden de una celda a otra.	2.917	4.16E-68
PIGY	Subunidad Y de fosfatidilinositol N-acetilglucosaminil transferasa	Es parte de un complejo que cataliza la transferencia de GlcNAc de UDP-GlcNAc a un aceptor fosfatidilinositol.	2.893	5.34E-21
CGREF1	Regulador de crecimiento celular con proteína 1 en dominio de mano EF	Media la adhesión célula-célula de una manera dependiente del calcio (por similitud).	2.889	3.95E-88
CRISP3	Proteína secretora 3 rica en cisteína	Este gen se expresa en el tracto reproductivo masculino, donde desempeña un papel en la función del esperma y la fertilización.	2.785	2.37E-10
ABCC4	Proteína 4 asociada a la resistencia a múltiples fármacos	Podría constituir una bomba de aniones orgánicos relevante para la desintoxicación celular.	2.75	1.77E-62
EPCAM	Molécula de adhesión de células epiteliales	Puede actuar como una molécula de interacción hemofílica física entre las células epiteliales intestinales (IEC) y los linfocitos intraepiteliales (IEL) en el epitelio de la mucosa para proporcionar una barrera inmunológica como primera línea de defensa contra la infección de la mucosa	2.691	1.54E-79
TMSB15A	Timosina beta-15a	Juega un papel importante en la organización del citoesqueleto. Se une y secuestra monómeros de actina (actina G) y, por lo tanto, inhibe la polimerización de actina (por similitud).	2.677	3.94E-59
TFF3	Factor de trébol 3	Proteína involucrada en la regulación de la reparación de la mucosa intestinal. Promueve la movilidad de las células epiteliales en los procesos de curación.	2.609	6.34E-21
SIM2	Homólogo 2 único	Es un factor de transcripción.	2.579	1.34E-58

\* lncRNA (Genes largos no codificantes)

Según el análisis de la ontología resultante (tabla 18), los productos de los genes sobre-expresados en Cpp están asociados a procesos como la adhesión celular y

metabolismo de ácidos grasos y lipoproteínas, siendo principalmente proteínas de unión a otras moléculas e iones calcio, estando localizados en citoplasma, membrana y en ambiente extracelular.

**Tabla 18. Los 5 principales términos de ontología de genes (GO) predicho por Genecodis para los genes significativamente sobre-expresados en CPP.**

Categoría	Termino	Descripción	Conteo	p-valor	Genes
PB	GO:0007155	Adhesión celular	14	1.17E-04	THBS4,EMB,CADM1,DSC2,SDK1,SPON2,NME2,CNTNAP2,PODXL2,SPOCK1,F5,CGREF1,DSG2,COMP
	GO:0035338	Proceso biosintético de acilo graso de cadena larga.	3	1.30E-02	ELOVL2,FASN,ELOVL7
	GO:0016338	adherencia celular-celular independiente del calcio	3	1.46E-02	CADM1,CLDN8,CLDN3
	GO:0034626	alargamiento del ácido graso, ácido graso poliinsaturado	2	1.72E-02	ELOVL2,ELOVL7
	GO:0034447	Aclaramiento de partículas de lipoproteínas de muy baja densidad	2	1.73E-02	APOC1,APOE
FM	GO:0005509	unión de iones de calcio	15	6.89E-05	THBS4,GALNT3,DSC2,CAMKK2,MMP10,TGM3,SPOCK1,CANT1,CALR,REPS2,PPP3CA,CGREF1,DSG2,TPD52,COMP
	GO:0016491	actividad oxidorreductasa	12	7.79E-05	DHRS7,PDIA5,FASN,RDH11,SCD,ALDH3B2,UGDH,PYCR1,STEAP1,PRDX4,TM7SF2,STEAP4
	GO:0002163	Unión a alfa-distroglucano	2	1.63E-03	AGR2,AGR3
	GO:0005516	unión a calmodulina	6	3.33E-03	MYO6,CAMKK2,IQGA,P2,TRPM4,PPP3CA,MAKSL1
	GO:0005529	Unión a azúcar	6	5.10E-03	GALNT3,SFTP2,COL12,GALNT7,CALR,ZP3
CC	GO:0005615	espacio extracelular	23	4.95E-11	THBS4,PRSS8,GDF15,SFTP2,MMP10,PLA2G7,CRISP3,APOE,CHIT1,ST14,CCL18,SPOCK1,SMPDL3B,CXCL11,CALR,F5,PRDX4,NPY,ZP3,MMP9,FAM3B,PCSK6,COMP

GO:0005576	región extracelular	35	5.97E-11	APOC1,AGR2,THBS4,TF3,PRSS8,AGR3,GDF15,TMEFF2,APOF,SPO N2,SFTPA2,MMP10,PLA2G7,CRISP3,APOE,C HIT1,ST14,CCL18,SPO CK1,SMPDL3B,LUZP2, CXCL11,CALR,F5,PRD X4,COL10A1,NPY,DN ASE2B,CGREF1,ZP3,MP9,FAM3B,PCK6,C OMP,CST2
GO:0005783	retículo endoplásmico	22	2.64E-08	APOC1,AGR2,GJB1,SE Z6L2,ELOVL2,ERGIC1 ,PDIA5,PDIA3,RDH11, CLDN8,SCD,CANT1,E NTPD5,TRPM4,CALR, ELOVL7,CYP2J2,DHC R24,ZP3,TM7SF2,TPD5 2,PCK6
GO:0016021	integral a la membrana	47	1.15E-07	OR51E1,HPN,GJB1,SE Z6L2,ELOVL2,GALNT 3,EMB,EPCAM,ST6GA LNAC1,PRSS8,CADM1 ,B3GAT1,NKAIN1,DSC 2,ERGIC1,SDK1,TMEF F2,RDH11,MAL2,CLD N8,CLDN3,PMEPA1,T MEM45B,TSPAN1,SCD ,CNTNAP2,OR51E2,M ANEAL,TLCD1,NAAL ADL2,ABCC4,GABRB 3,ABHD2,COLEC12,G ALNT7,CANT1,TRGC1 ,TRPM4,GCNT1,ELOV L7,GPR160,TRPM8,DH CR24,F2RL1,DSG2,ZP3 ,STEAP4
GO:0005794	Aparato de Golgi	20	2.46E-07	GALNT3,ST6GALNAC 1,B3GAT1,ERGIC1,FA SN,MYO6,CNTNAP2,A POE,MANEAL,GALNT 7,CANT1,TRPM4,GCN T1,RAP1GAP,CALR,G OLM1,DHCR24,F2RL1, ZP3,STEAP4

En el análisis de rutas enriquecidas se apreció que los genes sobreexpresados de Cpp se asocian tanto con la biosíntesis de mucin tipo O-glicano y a la vía del fagosoma (Tabla 19).

**Tabla 19. Resultados del análisis de enriquecimiento de vías para los genes sobre-expresados en CPp.**

Termino	Nombre de la ruta	Conteo	p-valor	Genes
Kegg:0051 2	Biosíntesis de mucin tipo O-glicano	4	8.23E-04	GALNT3,GCNT1, ST6GALNAC1,GALNT7
Kegg:0414 5	Fagosoma	5	1.65E-02	COMP,CALR,TUBB2A,T HBS4,COLEC12

KEGG, Enciclopedia de Kyoto de genes y genomas

### 6.2.3.2. Genes subexpresados

Se identificaron un total de 755 genes sub-expresados en CPp, que presentaron un log FC < -1.5. El gen ADAM33, codante para una enzima metaloendopeptidasa, resultó el más sub-expresado con un log FC < -4 (Tabla 20). También se observó que los genes GSTM2 (Glutation S-transferasa Mu 2) y GABRE (Receptor de ácido gamma-aminobutírico) están subexpresados, ambos con un log FC < -3.8. Al igual que en el grupo de genes sobre-expresados, se hallaron pseudogenes sub-expresados, tales como RP11-394O4.5, MEG3 y ACTA2-AS1, siendo todos ARN no codificantes de cadena larga (lncRNA de sus siglas en inglés), como se muestra en la Tabla 20.

**Tabla 20. Lista de los 20 genes más significativamente sub-expresados en CPp**

Símbolo del gen	Nombre del gen	Función del gen	log2(FC)	Q-Value
RP11-394O4.5*	ARN no codificante asociado al potenciador del mesodermo cardíaco	Función no determinada, con una alta expresión en endometrio y próstata.	-4.65	1.19E-72
ADAM33	Desintegrina y proteína 33 que contiene el dominio metaloproteínas a	actividad metaloendopeptidasa y unión a iones de zinc	-4.438	2.82E-82
MEG3*	Expresado maternalmente 3	Es un supresor de tumor, también interactúa con el supresor tumoral p53 y regula la expresión del gen diana p53.	-4.108	6.43E-74

GSTM2	Glutación S-transferasa Mu 2	Proteína que se encarga de mediar la conjugación de glutación.	-3.894	5.07E-85
GABRE	Receptor de ácido gamma-aminobutírico, subunidad épsilon	Esta proteína pertenece a la familia de canales iónicos activados por ligando.	-3.834	1.52E-59
ACTA2-AS1*	ACTA2 ARN antisentido 1	Funcion no determinada. Las enfermedades asociadas con este gen incluyen la susceptibilidad al cáncer de pulmón 3 y el carcinoma hepatocelular.	-3.715	4.56E-74
CYP3A5	Citocromo P450 3A5	Esta proteína está asociada a la vía de transporte de electrones dependiente de NADPH.	-3.556	8.12E-61
CORO6	Coronina-6	Esta proteína está asociada a la unión a filamentos de actina	-3.419	2.72E-59
GOLGA8A	Miembro de la subfamilia A de Golgi 8A	Esta proteína está vinculada a la regulación del mantenimiento del aparato de Golgi.	-3.368	1.16E-39
DIO3OS*	DIO3 filamento opuesto ARN aguas arriba	Puede tener un rol en la regulación de la expresión monoalélica de DIO3, un gen que a la vez regula la inactivación de la hormona tiroidea durante el desarrollo embriológico.	-3.342	1.18E-71
SLC26A10	Miembro 10 de la familia de portadores de solutos 26	Proteína que intercambia iones de cloruro / bicarbonato	-3.195	9.55E-72
HIF3A	Factor inducible por hipoxia 3-alfa	Proteína que regula la transcripcional bajo condiciones de estrés por oxígeno.	-3.161	1.78E-75
FAM153A	Miembro A de la Familia con similitud de secuencia 153	Función no determinada		
KRT13	Queratina, citoesqueleto tipo I 13	Actividad de molécula estructural	-3.118	1.56E-26
COL27A1	Cadena de colágeno alfa-1 (XXVII)	Juega un papel durante la calcificación del cartílago y la transición del cartílago al hueso.	-3.105	3.65E-66
GSTM1	Glutación S-transferasa Mu 1	Proteína asociada a la conjugación de glutación.	-3.08	3.76E-07
COL7A1	Cadena de colágeno alfa-1 (VII)	Proteína estratificada de la membrana basal epitelial escamosa que forma fibrillas de anclaje que pueden contribuir a la organización y adherencia de la membrana basal epitelial al interactuar con proteínas de la matriz extracelular (MEC), como el colágeno de tipo IV.	-3.028	1.48E-56
LCAT	Fosfatidilcolina-esterol aciltransferasa	Proteína importante en el metabolismo extracelular de lipoproteínas	-3.003	2.44E-78

MIR143HG*	Gen huésped MIR 143	Función no determinada.	-2.998	2.72E-76
CFD	Factor de complemento D	El factor D escinde el factor B cuando este último está en complejo con el factor C3b, activando el complejo C3bbb, que luego se convierte en la convertasa C3 de la ruta alternativa.	-2.975	1.51E-43

\* lncRNA (Genes largos no codificantes)

El análisis de ontología predicho mostro que la transducción de señales y procesos de regulación de la proliferación, adhesión y diferenciación celular están principalmente ligado a los genes subexpresados. Sus productos proteicos unen a iones y otras moléculas y son constituyentes de la matriz extracelular, citoplasma y membrana (Tabla 21).

**Tabla 21. Las 5 principales términos de ontología de genes (GO) con el p-valor más bajo para los genes sub-expresados en Cpp.**

Categoría	Término	Descripción	Conteo	p-valor	Genes
PB	GO:0007165	transducción de señales	64	2.57E-15	P2RX1,NPR2,L YVE1,MAP3K 6,ASB2,WISP2 ,CDK18,KCNI P3,S100A6,TN FRSF25,ANKD D1A,MPP2,TR AF5,GRASP,P ENK,MIB2,CN TFR,ACVRL1, CCK,RASSF4, RAPGEFL1,SF RP5,ARAP3,S YNGAP1,FGF 17,ACAP3,CY TL1,GAS6,PG F,RAPGEF3,G NRH1,IL11RA,

---

				PLEKHG5,AD CY3,IGFBP6,K CNIP2,ROR2,T RIP10,S100A9, MPP3,SYDE1, GPR173,PDE2 A,CAPN3,IRA K3,RIN1,TRAF 1,PLXNB1,FG F7,ARHGAP23 ,STAC,TLE2, MICAL1,NDR G2,ARHGAP3 3,DLG4,SCGB 1A1,SMOC1,R ASL12,EDN3,S TARD8,LAT,G ATA3,PDGFR B FES,ROBO3,B MP1,FHL1,TB X2,L1CAM,PD LIM7,HOXB3, SEMA6D,TBX 4,DPY19L2,FG F17,DZIP1,HO XB2,HOXD13, HOXD10,NOT CH4,NUMBL, PGF,GNRH1,N TRK3,HOXD9, ROR2,DLL1,IS L1,PLXNA2,ID 1,ITGB4,PLXN B1,SERPINF1, OBSCN,EBF4, GPM6B,NDRG 2,TWIST2,SN AI2,ID3,EYA4, SMOC1,ADA MTS9,MEIS1,P DGFRA,EDN3, HIC1,COL13A 1,PDGFRB FGFR2,WISP2, RARRES1,PT
GO:000 7275	desarrollo de organismos multicelulares	46	5.36E-09	
GO:000 8285	Regulación negativa de la	25	7.41E-08	

	proliferación celular.				H1R,ALDH1A2,ACVRL1,SFRP5,GABBR1,BCL6,BNIPL,DPT,GNRH1,IGFBP6,ROR2, COL18A1,PMP22,SPEG,TGFB3,TWIST2,PODN,MEG3,TGFB1I1,COL4A3,RARG,GATA3 MUC4,LYVE1,EFS,WISP2,L1CAM,COL9A1,COL7A1,NFASC,TRO,HAPLN2,ITGA7,FLRT2,COL4A6,DPT,ITGA5,THBS3,NID2,NCAM1,COL18A1,NID1,COL5A3,ITGB4,COL16A1,STAB1,COL27A1,TGFB1I1,COL4A3,COL6A1,COL5A1,DST ROBO3,BMP1,FHL1,L1CAM,LAMA5,PDLIM7,SFRP5,SEMA6D,DPY19L2,DZIP1,OSR1,NOTCH4,PGF,SRD5A2,NTRK3,ROR2,DLL1,SOX15,MALL,PMP22,OBSN,GPM6B,NDRG2,SMOC1,MEG3,NFATC4,TGFB1I1,CAPRIN2,COL13
GO:0007155	Adhesión celular	30	1.05E-06		
GO:0030154	diferenciación celular	29	1.15E-06		

					A1
FM	GO:0005201	constituyente estructural de la matriz extracelular	13	1.63E-08	MUC4,COL4A6,COL5A3,TINAGL1,COL6A2,FBLN1,ELN,COL27A1,LAMA4,EFEMP2,COL4A3,COL5A1,COL13A1BMP1,KCNIP3,S100A6,S100A4,RASGRP2,JMJD7-PLA2G4B,PLCH2,LPCAT4,RGN,MASP1,NOTCH4,GAS6,EGFL8,NINL,S100A8,S100B,KCNIP2,DLL1,DUOX2,THBS3,NID2,S100A9,S100A14,PLCD4,NID1,MEGF6,DUOX1,CAPN3,S100P,FBLN1,SMOC1,SGCA,S100A16,EFEMP2,DSPT,EFHD1HSPB7,FGFR2,FHL1,CRYAB,S100A6,S100A4,COL7A1,TRAF5,PDLIM7,SHANK3,STK36,HBA2,FGFR1,ACVRL1,PILRB,RASSF4,FS
	GO:0005509	Unión a iones de calcio	36	1.92E-08	TL3,SFI1,HSPB6,METTL17,LCAT,LENG8,CCDC130,DZIP1,ZNF219,DMPK,SMARCD3,SYNM,MA
	GO:0005515	Unión a proteínas	120	5.94E-08	

SP1,HBB,BCL  
6,NUMBL,PPP  
1R12C,TPCN1,  
SOX10,AGER,  
BNIPL,HSPB8,  
S100A8,ITGA5  
,KSR1,MYOT,  
LDB3,S100B,A  
CACB,PKD1,L  
3MBTL1,NPH  
P3,KCNIP2,KR  
T15,DLL1,APB  
B3,TNS1,LPA  
R1,NID2,S100  
A9,DNM1,DES  
,ID1,PDE2A,C  
OL18A1,MAL,  
TP63,FLNA,IR  
AK3,RIN1,TR  
AF1,ITGB4,PL  
XNB1,COL16  
A1,TGFB3,TN  
S4,SORBS1,SI  
N3B,OBSCN,G  
PRASP1,S100P  
,AKAP8L,TLE  
2,MICAL1,FB  
XO17,MRV11,  
COL6A2,NDR  
G2,ARHGAP3  
3,TWIST2,AD  
CY4,DLG4,ZN  
F655,ELN,WS  
B1,ADAMTSL  
4,MEIS1,MAP  
3K12,NLRP1,K  
RT5,EFEMP2,  
GAS1,ID4,RBP  
MS,JPH2,GST  
P1,TGFB111,C  
DK11A,COL4  
A3,STX16,SN  
AP25,RSPO1,H  
IC1,TRIM29,V  
AMP1,DST,CO  
L13A1,ACTA1

GO:0005524	Unión a ATP	52	4.47E-06	,DOC2A,RAR G,LAT,GATA3 ,PDGFRB,BEG AIN FES,P2RX1,SG K1,NPR2,FGF R2,ACTC1,MA P3K6,RIMKLB ,CDK18,STK36 ,P2RX2,FGFR1 ,ATAD3C,AC VRL1,PDK4,M YH3,DMPK,IP 6K3,MYO15B, ADCY3,KSR1, NTRK3,TRAN K1,ACACB,C HKB,ROR2,N AV2,NRBP2,T TLL3,IRAK3,S PEG,OBSCN,A TP1A2,PNCK, FBXO17,ADC Y5,ADCY4,M AP3K12,NLRP 1,ABCC3,PDG FRA,ACSF2,S TARD9,ABCA 8,CDK11A,CK M,UBA7,DST, KIFC3,ACTA1, ATAD3B,PDG FRB RNPC3,FES,S GK1,NPR2,FG FR2,ACTC1,T UBG2,MAP3K 6,RIMKLB,CD K18,RBM6,CE LF6,ISYNA1,S TK36,TDRD10 ,FGFR1,ATAD 3C,ACVRL1,P DK4,MYH3,R BFOX3,DMPK ,IP6K3,RDH5,
GO:0000166	Unión a nucleótidos	65	7.15E-06	

					RAPGEF3,AD CY3,NTRK3,A CACB,CHKB, ROR2,NAV2,D NM1,RBMS3, NRBP2,HSD11 B1,LTB4R,PD E2A,PABPC1L ,IRAK3,SPEG, OBSCN,ATP1 A2,BDH2,PNC K,FBXO17,AD H1B,ADCY5,A DCY4,GNAO1, MAP3K12,NL RP1,ABCC3,P DGFRA,ACSF 2,RASL12,STA RD9,RBPMS,A BCA8,CDK11 A,CKM,UBA7, KIFC3,ACTA1, ATAD3B,PDG FRB
CC	GO:000 5737	cytoplasm	185	6.62E-27	HSPB7,PRSS5 0,PLA2G6,FES ,SGK1,FGFR2, ACTC1,C9orf3, TUBG2,AMT,F HL1,EFS,WISP 2,CRYAB,RIM KLB,KCNIP3, S100A6,ITIH4, CELF6,JMJD7- PLA2G4B,ISY NA1,TRAF5,P DLIM7,SHAN K3,GRASP,ST K36,PTH1R,T RO,CDC25B,D MTF1,MIB2,F GFR1,PER1,A LDH1A2,PLC H2,SFI1,HSPB 6,SFRP5,SMT N,ARAP3,RGN

---

,MYH3,GABBR1,SEMA6D,RBFOX3,ENGASE,SNCG,SYNGAP1,PLIN4,GSTM5,GPX2,DZIP1,SLC18A2,HAAO,EPB41L2,DMPK,SMARCD3,SYNM,IP6K3,LMOD1,MYO15B,SEC31B,NUMBL,PPP1R12C,SOX10,AGER,PLEKHG5,TSPYL2,ADCY3,HSPB8,IGFBP6,NINL,S100A8,SRD5A2,KSR1,MYOT,LDB3,PPP1R14A,NTRK3,S100B,ACACB,CHKB,MAPK8IP3,AHSA2,KCNIP2,GSTM2,APBB3,CRIP3,TNS1,ISL1,LPAR1,TRIP10,SOX15,S100A9,DNM1,S100A14,GFPT2,DES,RBMS3,NRBP2,ID1,PLCD4,PDE2A,TTLL3,AOX1,NID1,TP63,FLNAD,DDIT4,CAPN3,IRAK3,RIN1,TRAF1,TGFB3,PAN2,TNS4,SORBS1,SIN3B,OBSCN,TINAGL1,ATP1A2,GSTM1,STAC,ARHGEF25,G

---

---

GO:000  
5576

región  
extracelular

89

7.35E-19

PRASP1,S100P  
,GSDMB,BDH  
2,SAP25,AKA  
P8L,PLEKHH2  
,PRCD,PNCK,  
BHMT2,ENO2,  
MICAL1,CLIP  
3,FBXO17,MR  
VII,ADH1B,N  
DRG2,KCNAB  
3,TWIST2,ALS  
2CL,ADCY4,B  
ZRAP1,CSDC2  
,DLG4,XAF1,S  
NAI2,ID3,SCG  
B1A1,AKR1C2  
,ZNF655,EYA4  
,PODN,PTGDS  
,SGCA,MAP3  
K12,NLRP1,P  
DGFRA,ARHG  
EF40,S100A16,  
MEG3,NFATC  
4,ID4,TRIM22,  
RBPMS,GSTP  
1,TGFB1I1,CD  
K11A,STX16,T  
AZ,SNAP25,C  
KM,CMAHP,T  
RIM29,HIF3A,  
ATG16L2,DST  
,CAPRIN2,KIF  
C3,ACTA1,PD  
GFRB,BEGAI  
N  
MUC4,FGFR2,  
BMP1,A1BG,  
WISP2,COL9A  
1,LAMA5,ITIH  
4,PROK1,JMJ  
D7-  
PLA2G4B,TNF  
RSF25,COL7A  
1,C2orf40,HAP  
LN2,PENK,FG  
FR1,LGI4,PLA

---

---

				C9,GPX3,LCN 10,CCK,FSTL3 ,SFRP5,LCAT, SERPINA5,FG F17,MASP1,C YTL1,COL4A6 ,AGER,GAS6, PGF,DPT,CPA MD8,GNRH1,I GFBP6,S100A 8,CES4A,S100 B,WFDC2,C7, LCN6,PCOLC E2,DLL1,DEF B1,THBS3,PI1 6,NID2,S100A 9,NCAM1,AD AMTS10,AMY 2B,SCGB3A1, COL18A1,NID 1,MEGF6,FLN A,COL5A3,PL XNB1,COL16 A1,SERPINF1, TGFB3,FGF7, NTF4,TINAGL 1,DMKN,PRSS 1,PAPLN,COL 6A2,SCGB1A1 ,FBLN1,SMOC 1,ELN,COL27 A1,ADAMTS9, PODN,PTGDS, ADAMTSL4,C 1QTNF1,LAM A4,EFEMP2,E DN3,RARRES 2,ABO,COL4A 3,LYNX1,COL 6A1,RSPO1,C OL5A1 FES,P2RX1,M UC4,SGK1,NP R2,FGFR2,LY VE1,FHL1,WI SP2,L1CAM,C
GO:000 5886	membrana de plasma	110	6.65E-11	

---

---

RYAB,KCNIP  
3,GNB3,S100A  
6,LGR6,GPR17  
,ITIH4,FXYD1,  
TNFRSF25,SH  
ANK3,GRASP,  
PTH1R,NFAS  
C,TRO,FGFR1,  
CNTFR,COL23  
A1,PLCH2,LI  
ME1,SFRP5,GI  
PR,ARAP3,SE  
MA6D,ITGA7,  
CLEC2D,TCIR  
G1,PLIN4,SLC  
12A4,SLC18A  
2,EPB41L2,D  
MPK,AGER,M  
PZ,CPAMD8,R  
APGEF3,PLEK  
HG5,TSPYL2,  
ADCY3,S100A  
8,LIMS2,ITGA  
5,NTRK3,PKD  
1,SLC2A5,MA  
PK8IP3,KCNIP  
2,LMBR1L,DL  
L1,DUOX2,LP  
AR1,JPH4,TRI  
P10,S100A9,N  
CAM1,FXYD6,  
S100A14,SLC4  
7A1,PLXNA2,  
LTB4R,SLC16  
A5,GPR173,PD  
E2A,GABRE,D  
UOX1,FLNA,R  
IN1,ITGB4,PL  
XNB1,TGFB3,  
SORBS1,PRIM  
A1,ATP1A2,A  
RHGEF25,EN  
O2,CLIP3,ADC  
Y5,KCNAB3,A  
DCY4,DLG4,G  
NAO1,SLC7A6

---

---

					,MRGPRF,MAP3K12,ABCC3,PDGFRA,S100A16,PEAR1,GAS1,JPH2,GSTP1,ABCA8,LYNX1,TAZ,SNAP25,CPLX3,HIF3A,COL13A1,SLC14A1,LAT,PDGFRB
GO:0005581	colágeno	14	7.61E-09		COL9A1,COL7A1,COL23A1,COL4A6,COL18A1,COL5A3,COL16A1,COL6A2,COL27A1,C1QTNF1,COL4A3,COL6A1,COL5A1,COL13A1
GO:0031012	la matriz extracelular	17	8.01E-09		FGFR2,LAMA5,DPT,MMP23B,NID2,COL18A1,NID1,SERPINF1,TGFB3,TINAGL1,COL6A2,FBLN1,ADAMTSL4,RARRES2,TGFB1I1,COL6A1,COL5A1

---

Los genes sub-expresados se vinculan principalmente a las vías de digestión y absorción de proteínas, a la interacción receptor-Matriz Extracelular (ECM), a la vía de señalización MAPK y adhesión focal así como a la ruta de miocardiopatía dilatada (Tabla 22).

**Tabla 22. Resultados del análisis de enriquecimiento de vías para los genes sub-expresados en Cpp. Se muestran sólo los 5 primeros términos KEGG más significativos.**

Termino	Nombre de la ruta	Conteo	p-valor	Genes
Kegg:0497 4	Digestión y absorción de proteínas.	11	4.66E-06	COL4A6,COL5A3,ELN,PGA3, COL18A1,COL6A2,COL5A1,P RSS1,COL9A1,COL6A1,ATP1 A2
Kegg:0451 2	Interacción ECM-receptor	11	5.12E-06	COL4A6,COL5A3,LAMA5,IT GA5,LAMA4,COL6A2,COL5A 1,ITGB4,COL6A1,THBS3,ITG A7 JMJD7-
Kegg:0401 0	Vía de señalización MAPK	18	5.80E-06	PLA2G4B,CACNB1,TGFB3,FL NA,PDGFRA,FGF17,NFATC4, MAP3K12,MAP3K6,PLA2G6, CDC25B,FGFR1,PDGFRB,MA PK8IP3,RASGRP2,FGF7,NTF4, FGFR2
Kegg:0541 4	Miocardiopatía dilatada	11	6.27E-06	ACTC1,ADCY5,CACNB1,TGF B3,ITGA5,ADCY3,ADCY4,IT GB4,SGCA,DES,ITGA7 PGF,COL4A6,COL5A3,LAMA 5,FLNA,PDGFRA,ITGA5,LAM A4,COL6A2,PDGFRB,COL5A 1,ITGB4,COL6A1,THBS3,ITG A7
Kegg:0451 0	Adhesión focal	15	1.37E-05	

### 6.3. Asociación del genotipo con el fenotipo del CP.

#### 6.3.1. Correlación de las mutaciones puntuales en GSMs y fenotipo.

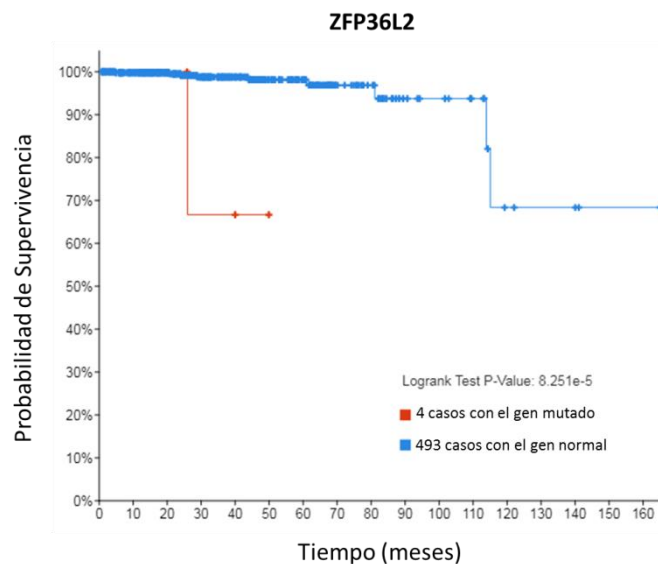
Se evaluó la correlación de la presencia de mutaciones puntuales, en los 35 GSMs del Cpp, con variables fenotípicas de la enfermedad. Empleando cBioPortal se observó que mutaciones en tres genes se correlacionaban con al menos a una variable fenotípica (Tabla 23). Las asociaciones se evaluaron empleando diferentes pruebas estadísticas, dependiendo de la naturaleza de las variables clínicas y/o genéticas. Así se aplicó la prueba de Chi-cuadrado para las variables

estadio T y N y escala de Gleason. En tanto que el trazado de la Kaplan Meier se aplicó para la variable de SG.

**Tabla 23. Correlaciones entre el genotipo y fenotipo (variables clínicas) en CPp.**

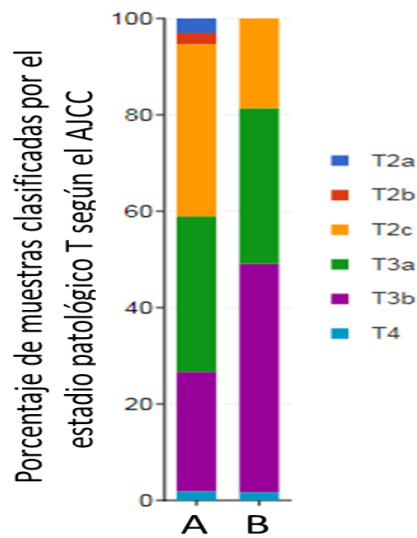
Variable clínica	Prueba estadística	P-valor	Gen mutado
SG	Kaplan-Meier	8.251e-5	ZFP36L2
Estadio T	Chi-squared Test	5.153e-3	TP53
Estadio N	Chi-squared Test	1.595e-5	TP53
Escala de Gleason	Chi-squared Test	2.05e-9	TP53
		5.805e-3	ZMYM3

Mutaciones en el gen ZFP36L2, que regula la respuesta a factores de crecimiento, se asociaron a la SG de los pacientes con CPp. ( $P < 0.05$ ). La gráfica de Kaplan-Meier mostró que los individuos con mutaciones en el gen ZFP36L2 tuvieron una supervivencia menor en comparación con pacientes sin mutaciones en dicho gen (figura 33).

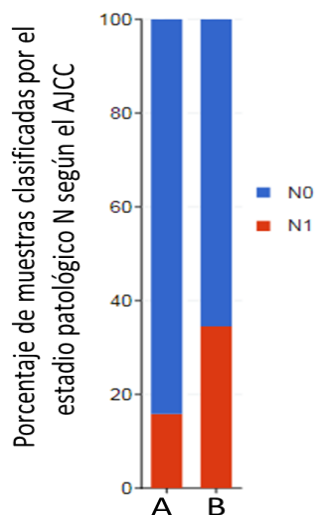


**Figura 33. Gráfica de KAPLAN-MEIER mostrando la supervivencia global de pacientes con y sin mutaciones en el gen ZFP36L2. Cohorte TCGA (N= 497) para CPp.**

El gen TP53, un supresor de tumor, se correlacionó al estadio patológico (T y N) de Cpp. La categoría T, que mide la extensión del tumor, se asoció al estatus mutacional de TP53 con un  $p < 0.05$ , empleando la prueba de Fisher. Así, los pacientes con categorías más bajas en el estadio T (T1 y T2), lo que significa que el tumor aún se encontraba localizado presentaron tp53 normal (figura 34). De igual forma para la categoría N, parámetro clínico que mide si el tumor ha invadido a los ganglios linfáticos cercanos. La mayoría de pacientes quienes mostraron categoría N0 (n=309), lo que significa que no hubo expansión del tumor a los nódulos linfáticos, presentaron tp53 normal (Prueba de Fisher,  $P < 0.05$ ) (figura 35).



**Figura 34. Gráfica de barras mostrando el estadio patológico T de pacientes con y sin mutaciones en el gen TP53 con. Cohorte TCGA (N= 497) para Cpp. (A) Muestras con el gen normal. (B) Muestras con el gen mutado.**



**Figura 35. Gráfica de barras mostrando el estadio patológico N de pacientes con y sin mutaciones en el gen TP53 con. Cohorte TCGA (N= 497) para Cpp. (A) Muestras con el gen normal. (B) Muestras con el gen mutado.**

Por otro lado, mutaciones en los genes ZMYM3 y TP53 estuvieron asociadas con la escala de Gleason ( $P < 0.05$ ) (figuras 36 y 37). ZMYM3 codifica una proteína de dedos de Zinc que se piensa es un componente del complejo histona desacetilasa, modificadoras de la estructura de la cromatina y responsables del silenciamiento génico. El grupo de pacientes con ZMYM3 y TP53 mutados tuvieron una escala de Gleason promedio de 8.5 (DS 0.9) y 8.4 (DS 0.9) respectivamente, mientras que el grupo de pacientes sin mutaciones en el gen, tuvieron una escala de Gleason promedio de 7.6 (DS 1.0) para el gen ZMYM 3 y de 7.5 (DS 1.0) para TP53. Ya que la escala de Gleason mide el grado de agresividad del tumor basándose en las características que presentan las células de una biopsia, una mayor escala, indica un cáncer más agresivo. Es decir, aquellos pacientes con mutaciones en ZMYM3 y TP53, mostraron un cáncer más avanzado en comparación con aquellos pacientes sin mutaciones.

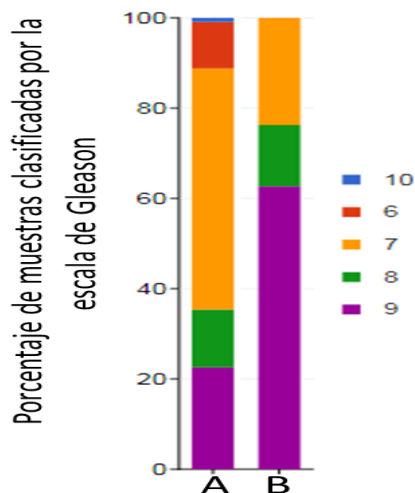


Figura 36. Gráfica de barras mostrando el puntaje de Gleason de pacientes con y sin mutaciones en el gen TP53 con. Cohorte TCGA (N= 497) para CPp. (A) Muestras con el gen normal. (B) Muestras con el gen mutado.

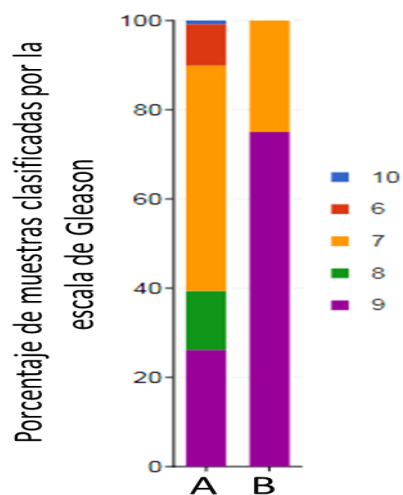
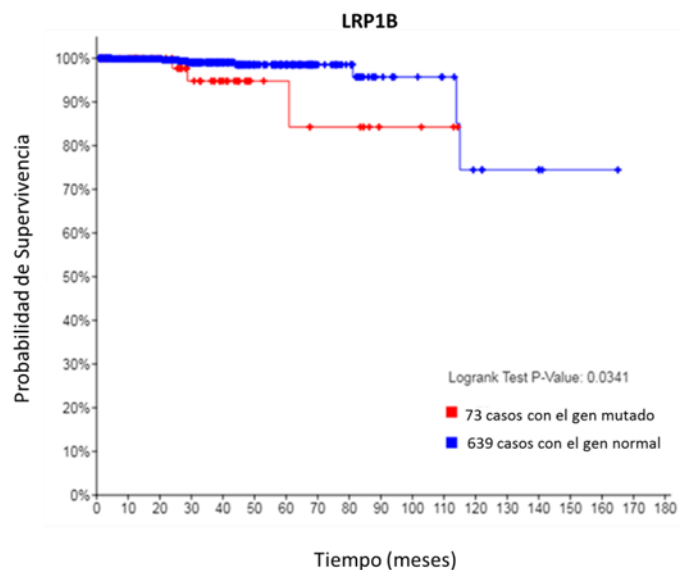


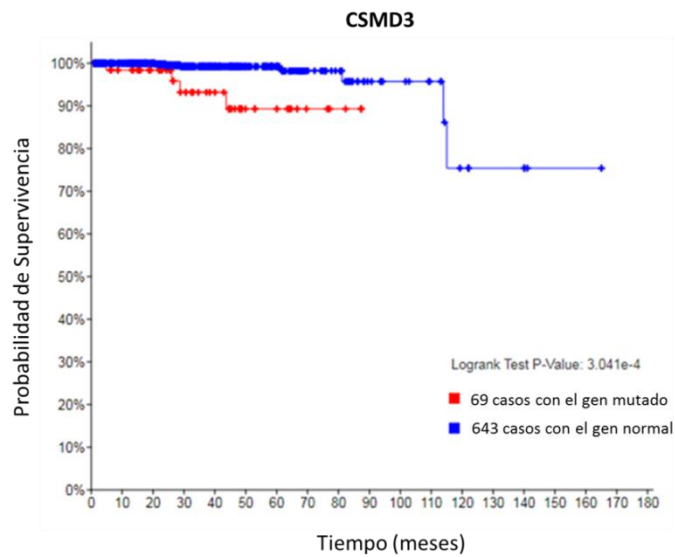
Figura 37. Gráfica de barras mostrando el puntaje de Gleason de pacientes con y sin mutaciones en el gen ZMYM3 con. Cohorte TCGA (N= 498) para CPp. (A) Muestras con el gen normal. (B) Muestras con el gen mutado

### 6.3.2. Correlación entre la mutación puntual de los GFM's y el fenotipo.

Se evaluó la correlación de mutaciones puntuales en los 20 GFM's del CPp con rasgos fenotípicos de los pacientes. Se encontró que mutaciones en dos GFM's estuvieron asociadas a la SG de los pacientes. Individuos con mutaciones en los genes LRP1B, proteína 1B relacionada con el receptor de lipoproteínas de baja densidad y CSMD3, proteína 3 que contiene dominios CUB y sushi, presentaron una menor supervivencia (hasta 110 meses para los pacientes con el gen LRP1B mutado y hasta 90 meses para pacientes con el gen CSMD3 mutado) en comparación a los pacientes que no presentaron mutaciones en dichos genes (hasta 170 meses aproximadamente) (figuras 38 y 39, respectivamente).

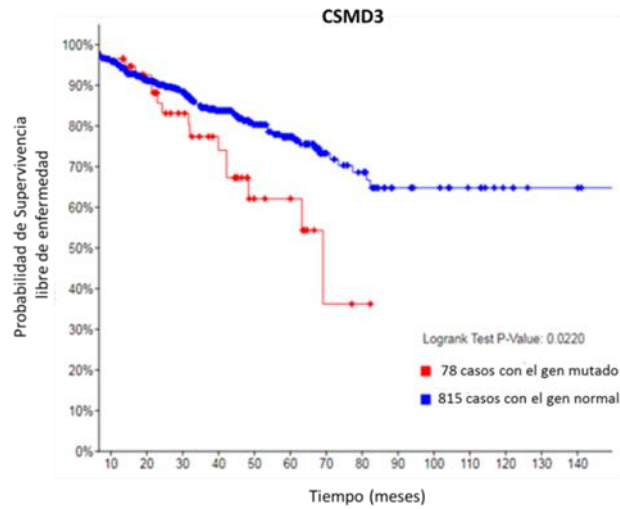


**Figura 38. Grafica de KAPLAN-MEIER mostrando la supervivencia global de pacientes con y sin mutaciones en el gen LRP1B. Estudios de cBioPortal para CPp (N= 712).**



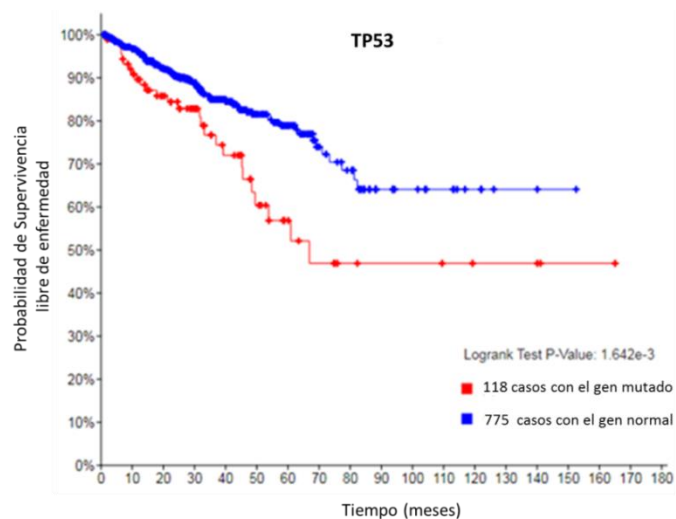
**Figura 39. Grafica de KAPLAN-MEIER mostrando la sobrevida global de pacientes con y sin mutaciones en el gen CSMD3. Estudios de cBioPortal para CPp (N= 712).**

Por otro lado, 6 GFM's resultaron asociados al SLE pacientes con CPp. Individuos con mutaciones puntuales en los genes TP53, SPOP, PTEN, CSMD3, SPTA1 y HMCN1 tuvieron un menor tiempo de progresión libre de la enfermedad en comparación con aquellos pacientes que presentaron dichos genes normales. Pacientes con el gen CSMD3 mutado vivieron 85 meses libres de progresión de la enfermedad en comparación a los 150 meses que tuvieron los pacientes con el gen normal (Figura 40).



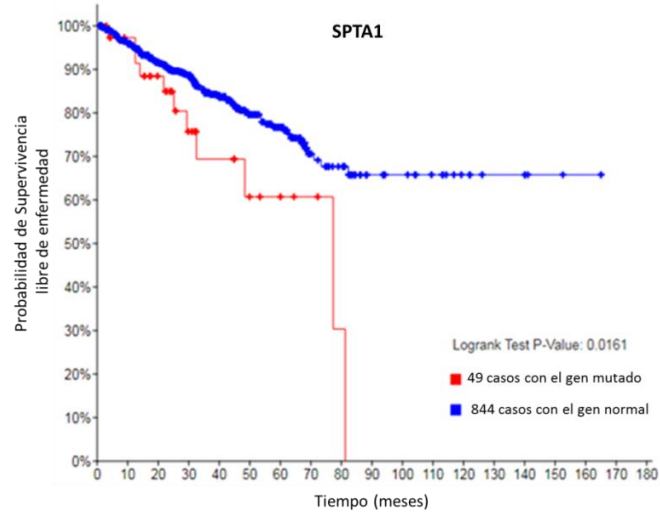
**Figura 40. Grafica de KAPLAN-MEIER mostrando la sobrevida libre de enfermedad de pacientes con y sin mutaciones en el gen CSMD3. Estudios de cBioPortal para CPp (N= 893).**

Así mismo, individuos con mutaciones del gen TP53 tuvieron una SLE de hasta 150 meses, menor en comparación a los pacientes con el gen normal (170 meses) (Figura 41).

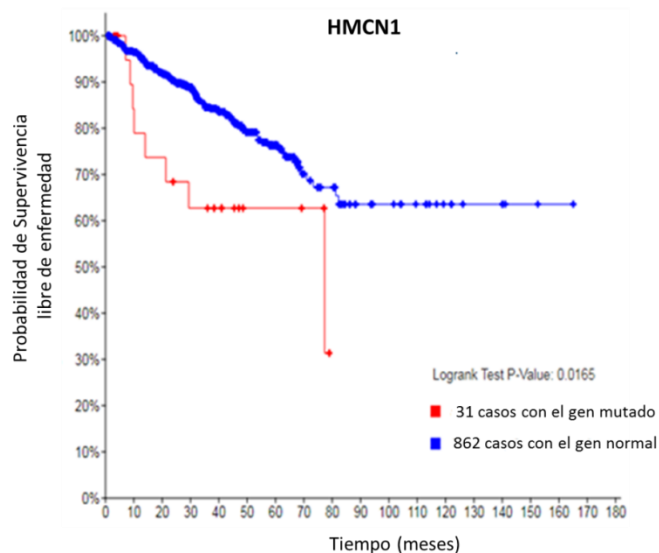


**Figura 41. Grafica de KAPLAN-MEIER mostrando la sobrevida libre de enfermedad de pacientes con y sin mutaciones en el gen TP53. Estudios de cBioPortal para CPp (N= 893).**

Por otro lado, pacientes con mutaciones en los genes SPTA1 y HMCN1 fallecieron antes de cumplir los 80 meses libres de progresión de la enfermedad, mientras que individuos con el gen normal lograron sobrevivir más de 160 meses (Figuras 42 y 43).

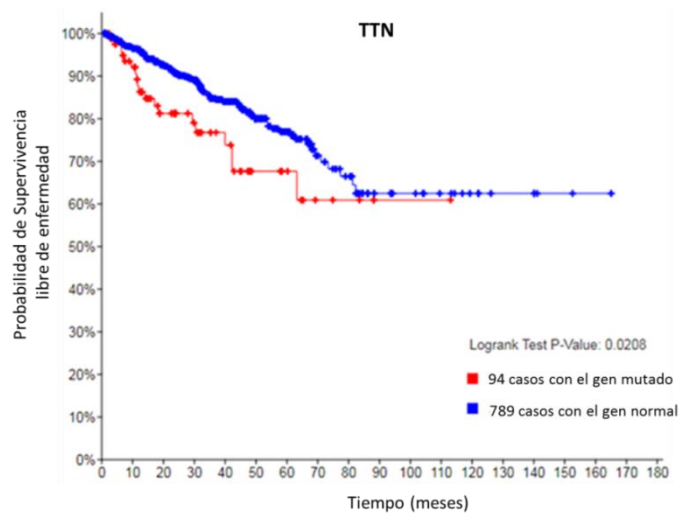


**Figura 42. Grafica de KAPLAN-MEIER mostrando la sobrevida libre de enfermedad de pacientes con y sin mutaciones en el gen SPTA1. Estudios de cBioPortal para CPp (N= 893).**



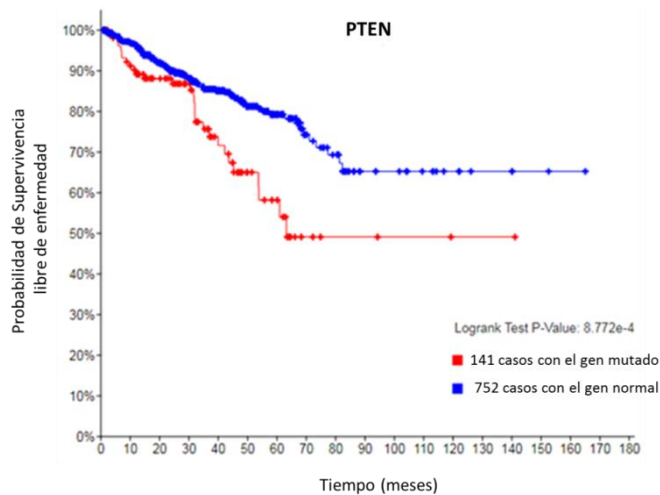
**Figura 43. Grafica de KAPLAN-MEIER mostrando la sobrevida libre de enfermedad de pacientes con y sin mutaciones en el gen HMCN1. Estudios de cBioPortal para CPp (N= 893).**

En el caso de pacientes con el gen TTN mutado lograron tener aproximadamente 120 meses libre de progresión mientras que los pacientes con el gen normal superaron los 170 meses (Figura 44).



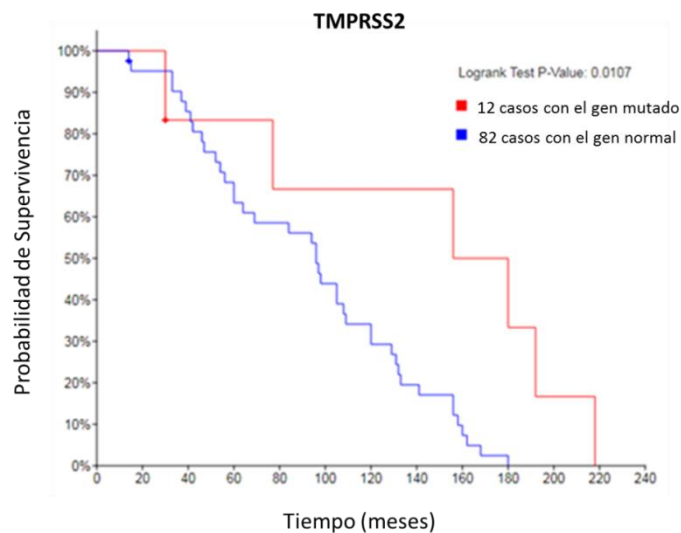
**Figura 44. Grafica de KAPLAN-MEIER mostrando la sobrevida libre de enfermedad de pacientes con y sin mutaciones en el gen TTN. Estudios de cBioPortal para Cpp (N= 883).**

En pacientes con PTEN mutado la SLE fue cerca de 140 meses, tiempo menor en comparación a los 170 meses que sobrevivieron libre de enfermedad los individuos con el gen no mutado (Figura 45).



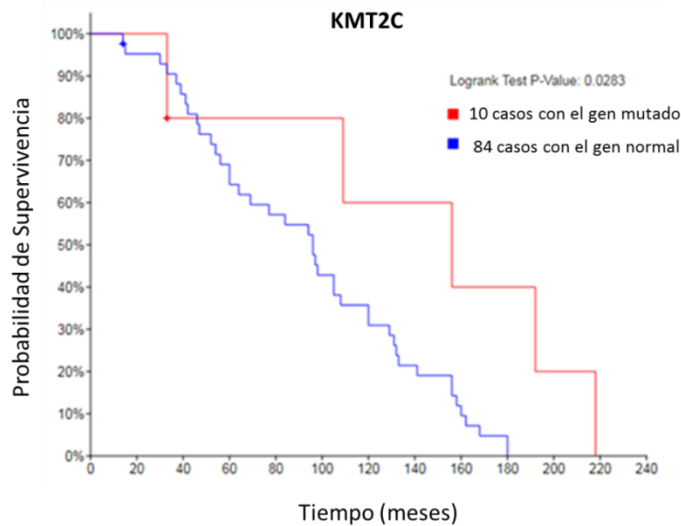
**Figura 45. Grafica de KAPLAN-MEIER mostrando la sobrevida libre de enfermedad de pacientes con y sin mutaciones en el gen PTEN. Estudios de cBioPortal para CPd (N= 893).**

En CPM, mutaciones puntuales en tres GFM resultaron asociadas a la SG. Así, pacientes con mutaciones en el gen *TPRSS2*, una serina proteasa que facilita las fusiones de virus-membrana celular, tuvieron una SG mayor (de hasta 220 meses) que los pacientes sin mutaciones en este gen (hasta 180 meses) (Figura 46).

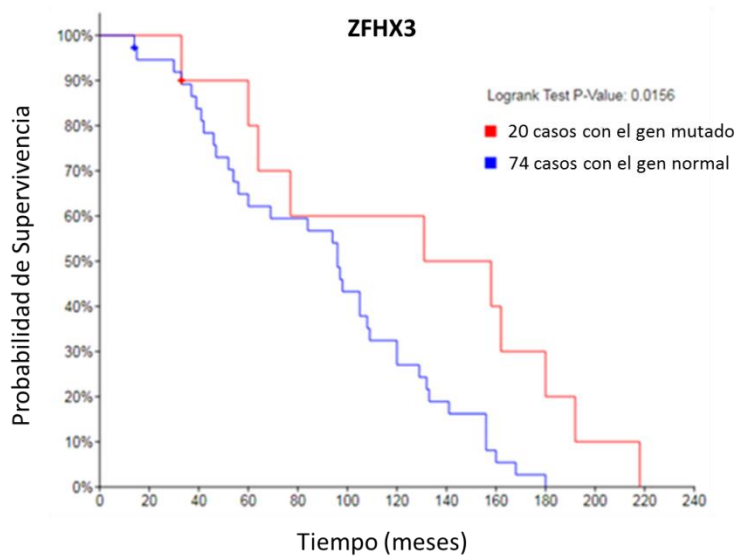


**Figura 46. Grafica de KAPLAN-MEIER mostrando la sobrevida global de pacientes con y sin mutaciones en el gen *TPRSS2*. Estudios de cBioPortal para CPM (N= 94).**

De forma similar ocurrió para los genes KMT2C, Histona-lisina N-metiltransferasa 2C y ZFHX3, proteína homeobox dedo de zinc 3. Aquellos pacientes con mutaciones en KMT2C o ZFHX3 mostraron un pronóstico más favorable en cuanto al tiempo de sobrevida, en comparación con aquellos individuos que no presentaban mutaciones en dichos genes (Figura 47 y 48). Adicionalmente, el tiempo máximo de SG en sujetos con mutaciones en TMPRSS2, KMT2C o ZFHX3 fue de 220 meses, siendo 40 meses más en comparación con los pacientes sin mutación en dichos genes.



**Figura 47. Grafica de KAPLAN-MEIER mostrando la sobrevida global de pacientes con y sin mutaciones en el gen KMT2C. Estudios de cBioPortal para CPM (N= 94).**



**Figura 48. Grafica de KAPLAN-MEIER mostrando la sobrevida global de pacientes con y sin mutaciones en el gen ZFH3. Estudios de cBioPortal para CPM (N= 94).**

### **6.3.3. Correlación entre los genes con alteración en el número de copias y el fenotipo**

#### **6.3.3.1 Genes Deleccionados en cáncer de próstata primario**

Se investigó, a través de cBioPortal, si deleciones en los 10 genes más frecuentemente deleccionados en Cpp se asociaban con variables clínicas en la cohorte TCGA (Tabla 24). El gen PTEN, quien fue el gen más frecuentemente deleccionado en la población, estuvo asociado con el estadio patológico T ( $P=0.012$ ) y con la escala de Gleason ( $P<0.01$ ). Los pacientes que tuvieron la deleción de gen PTEN mostraron una categoría más alta en el estadio T (T2, T3 y T4) y un puntaje de Gleason más elevado ( $>8$ ) en comparación a los individuos sin el gen deleccionado. Por otro lado, deleciones en los genes GTF2F2, TPT1 y GPALPP1 se asociaron con un mayor estadio patológico T, mostrando categorías T3 y T4 principalmente.

El Gen GTF2F2, factor de transcripción general IIF subunidad 2, codifica una subunidad del TFIIF, que es un factor de iniciación de la transcripción general que se une a la ARN polimerasa II y ayuda a reclutarlo para el complejo de iniciación en colaboración con TFIIB. TPT1, proteína tumoral controlada traslacionalmente, es una proteína involucrada en la unión a calcio y la estabilización de microtúbulos; mientras que GPALPP1, también conocida como LSR7, proteína de respuesta específica de lipopolisacárido no tiene una función claramente definida.

**Tabla 24. Genes deletados en cáncer de próstata primario asociados a variables clínicas.**

Variable clínica	Gen	P-valor	prueba estadística
Estadio patológico T	PTEN	1.20E-02	Chi-squared Test
	GTF2F2	3.60E-02	
	TPT1	3.60E-02	
	GPALPP1	3.60E-02	
Escala de Gleason	PTEN	1.92E-04	Chi-squared Test

### 6.3.3.2. Genes amplificados en cáncer de próstata primario

También se investigó si amplificaciones de los 10 genes más frecuentemente amplificados en CPp estuvieron asociadas a ciertas variables clínicas (Tabla 25). Así, la amplificación de los genes CASC8, MYC, POU5F1B, CCAT1, PCAT1, PCAT2, ZBTB10, ZNF704, RN7SL107P y FABP4 estuvo asociada a la SG y al estadio patológico T. Los pacientes con CPp que tenían amplificación en alguno de los nueve genes mencionados anteriormente, mostraron una sobrevida menor, de aproximadamente 110 meses, en comparación a aquellos individuos que no mostraron amplificación de dichos genes, quienes tuvieron una sobrevida de hasta 165 meses. Similarmente, las categorías de estadio patológico T3 y T4 fueron más frecuentes en aquellos pacientes con amplificación en alguno de los 9 genes en cuestión. CASC8, candidato a susceptibilidad al cáncer 8, produce un ARN no codificante, mientras que MYC es un factor de transcripción que tiende a la unión al ADN activando la expresión de genes asociados al crecimiento. POU5F1B es un gen que puede codificar una proteína funcional similar al factor de transcripción homeobox 1 de POU clase 5, siendo un activador transcripcional débil que puede desempeñar un papel en la carcinogénesis y el desarrollo ocular. CCAT1, es un gen que produce un ARN largo no codificante que promueve la formación de tumores

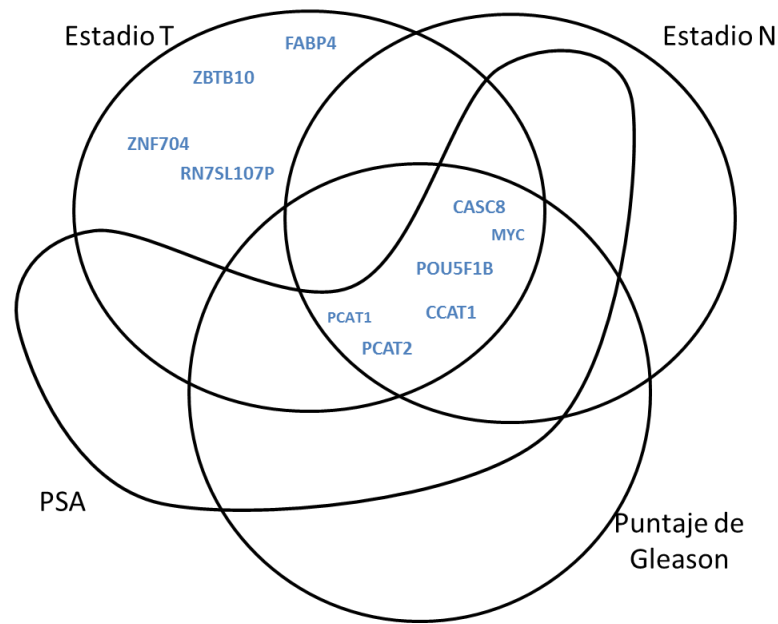
y está sobre regulado en el cáncer de colon y otros tipos de células cancerosas. PCAT 1 y 2, son igualmente genes productores de ARN no codificantes asociados a cáncer de próstata, PCAT1 regula negativamente la proteína supresora de tumores BRCA2 y regula positivamente la oncoproteína Myc. ZBTB10, proteína 10 conteniendo dominio BTB y dedo de zinc, hipotéticamente estaría involucrado en la regulación transcripcional. ZNF704, proteína de dedo de zinc 704, es un factor de transcripción que se adhiere a elementos de secuencia RE2. Mientras que RN7SL107P, pseudogen citoplasmático 107 ARN, 7SL, produce un ARN largo no codificante.

**Tabla 25. Genes amplificados en cáncer de próstata primario asociados a variables clínicas.**

Variable	Gen	P-valor	prueba estadística
SG	CASC8	3.61E-04	Kaplan Meier
	MYC	2.47E-04	
	POU5F1B	3.48E-04	
	CCAT1	2.61E-04	
	PCAT1	1.35E-04	
	PCAT2	1.35E-04	
	ZBTB10	2.98E-02	
	ZNF704	2.98E-02	
	RN7SL107P*	2.98E-02	
Estadio patológico T	CASC8	2.36E-02	Chi-squared Test
	MYC	1.17E-02	
	POU5F1B	0.0249	
	CCAT1	0.0169	
	PCAT1	0.0117	
	PCAT2	0.0117	
	ZBTB10	0.0141	
	ZNF704	0.0141	
	RN7SL107P*	0.0141	
Estadio patológico N	CASC8	0.003939	Chi-squared Test
	MYC	0.001799	
	POU5F1B	0.0337	
	CCAT1	0.00877	
	PCAT1	0.001799	
	PCAT2	0.001799	

Escala de Gleason	CASC8	0.0001317	Chi-squared Test
	MYC	0.0000615	
	POU5F1B	0.0003097	
	CCAT1	0.0001107	
	PCAT1	0.00003355	
	PCAT2	0.00003355	
PSA	CASC8	0.0003377	Kruskal Wallis Test
	MYC	0.0003381	
	POU5F1B	0.0003377	
	CCAT1	0.0003387	
	PCAT1	0.0001657	
	PCAT2	0.0001657	

Por otro lado, la amplificación de los genes CASC8, MYC, POU5F1B, CCAT1, PCAT1, PCAT2 estuvo asociado al estadio patológico N, donde la ganancia del número de copias en aquellos genes estuvo asociada a una categoría mayor de este estadio (N1), indicativo de un estado más avanzado de la enfermedad. Así mismo, un puntaje de Gleason alto (>8) y un valor de PSA alto estuvieron asociados a la amplificación en estos seis últimos genes (Figura 49).



**Figura 49. Genes frecuentemente amplificados en Cpp que comparten asociaciones en común con variables clínicas.**

#### 6.3.4. Genes con expresión diferencial

Se evaluó la correlación de la expresión génica diferencial con las variables fenotípicas (demográfica y clínico-patológica) del cáncer próstata primario. Se emplearon los recursos GEPIA y XENA para sondear dicha correlación en los 20 primeros GDEs. En primer lugar se investigaron los GDEs subexpresados (Tabla 26). Se identificaron 8 genes correlacionados a la variable PSA de los cuales ADAM33, GSTM2, HIF3A, FAM153A, LCAT y FAM153B se correlacionaron negativamente con esta variable, es decir, cuanto menor era la expresión de estos genes, mayores eran los valores del PSA. En tanto que la subexpresión de GOLGA8A y GOLGA8B se correlacionó positivamente con PSA, es decir, cuanto menor era la expresión de dichos genes, menores eran los valores de PSA. Con esta variable se hallaron correlaciones débiles con rangos de coeficientes de

Pearson desde -0.09 a -0.17 para las correlaciones negativas; y entre 0.11 y 0.12 para las correlaciones positivas. ADAM33, desintegrina y proteína que contiene el dominio metaloproteína 33, es una proteína que forma parte de la familia ADAM (un dominio de desintegrina y metaloproteasa). Esta familia alberga proteínas que se relacionan a procesos de interacciones célula-célula y matriz-célula. GSTM2, Glutación S-transferasa Mu 2, produce una glutación S-transferasa perteneciente a la clase mu. Las proteínas perteneciente a la clase mu están involucradas en procesos de desintoxicación de compuestos electrófilos, además de sustancias tóxicas medioambientales y desechos por estrés oxidativo.

Los genes GOLGA8A y GOLGA8B, miembros de la subfamilia A de Golgi, hipotéticamente están involucrados en el mantenimiento de la estructura de Golgi. HIF3A, factor inducible por hipoxia 3-alfa, está vinculado a la regulación de la transcripción bajo condiciones de estrés por oxígeno. Los genes FAM153A y FAM153B, codante para la proteína FAM153A y FAM153B respectivamente, están asociados al carcinoma renal, mientras que LCAT, una fosfatidilcolina-esterol aciltransferasa, codifica para una enzima central en el metabolismo extracelular de las lipoproteínas plasmáticas.

**Tabla 26. Asociación de genes subexpresados con variables clínicas en Cpp.**

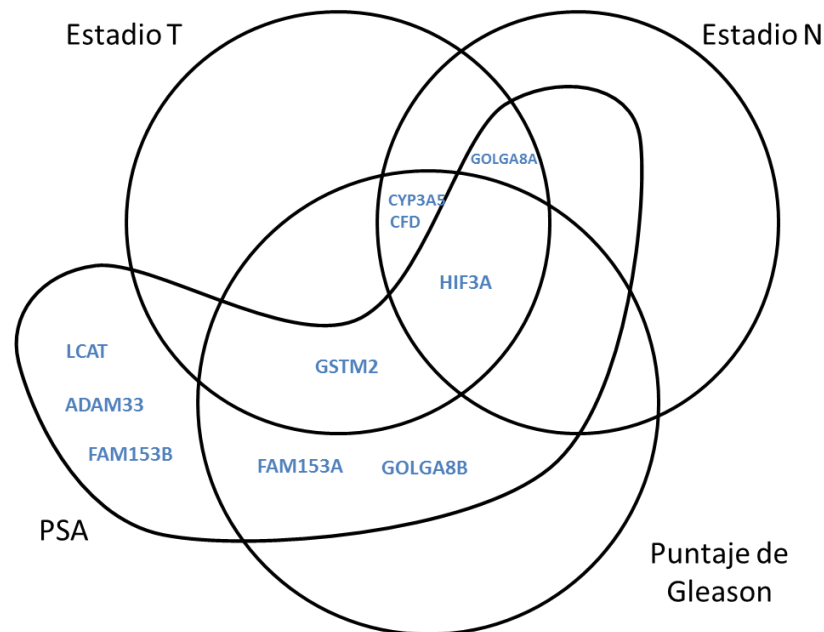
Variable	Gene	p-valor*	prueba estadística
PSA	ADAM33	7.17E-04 (-1.61E-01)	Correlación de Spearman
	GSTM2	7.81E-04 (-1.60E-01)	
	GOLGA8A	9.00E-03 (0.1244)	
	HIF3A	3.75E-04 (-1.69E-01)	
	FAM153A	2.52E-04 (-1.74E-01)	
	LCAT	4.75E-02 (-9.46E-02)	
	GOLGA8B	1.69E-02 (0.1139)	
	FAM153B	5.16E-02 (-9.28E-02)	
Escala de Gleason	GSTM2	4.58E-04	Anova
	CYP3A5	4.34E-08	
	GOLGA8A	5.06E-06	
	HIF3A	1.306e-7	
	FAM153A	2.97E-02	
	CFD	3.73E-03	
	GOLGA8B	0.0001597	
Estadio T	GSTM2	0.01434	Anova
	CYP3A5	2.95E-04	
	HIF3A	9.14E-04	
	CFD	1.69E-02	
Estadio N	CYP3A5	4.80E-03	Anova
	GOLGA8A	0.009499	
	HIF3A	0.00005175	
	CFD	1.15E-04	

\*Se indica el p-valor asociado a la prueba estadística y entre paréntesis el coeficiente de correlación de Spearman.

También se observó que la subexpresión de siete genes estuvo asociado a la escala de Gleason (Tabla 26). Así, cuanto menor era la expresión de los genes GSTM2, CYP3A5, GOLGA8A, HIF3A, FAM153A, CFD y GOLGA8B, los puntajes de Gleason eran mayores (>8). El gen CYP3A5 codifica la monooxigenasa del citocromo P450 involucrada en el metabolismo de las hormonas esteroides y las vitaminas. CFD, factor de complemento D, codifica un miembro de la familia de serina peptidasas S1, o quimotripsina. Esta proteasa

cataliza la escisión del factor B, también funciona como una adipocina que regula la secreción de insulina en ratones.

En tanto que la subexpresión de los genes GSTM2, CYP3A5, HIF3A y CFD se asoció a estadios patológicos T altos, principalmente T3 y T4 y a la categoría N1 del estadio patológico N. De tal forma que una menor expresión en cualquier de estos 4 genes se encontraría vinculada a un estadio más avanzado de la enfermedad (Figura 50).



**Figura 50. Genes subexpresados en Cpp que comparten asociaciones en común con variables clínicas.**

A continuación se investigó la correlación de los GDEs sobreexpresados con el fenotipo en Cpp (Tabla 27). Resultó que la sobreexpresión de seis GDEs se correlacionó negativamente a la variable PSA. Con ello, una mayor expresión en cualquiera de los genes incluyendo PCA3, OR51E2, GOLM1, ACSM1, GJB1 y ABCC4 se correlacionaron a valores altos de PSA (>4 ng/ml). Se hallaron

correlaciones débiles, donde el rango del coeficiente de correlación de Pearson fue de -0.13 a -0.2. El receptor olfativo 51E2, OR51E2, que es activado por el odorante, la beta-ionona, un terpenoide sintético, fue el gen cuya sobreexpresión tuvo la correlación negativa más fuerte con un coeficiente de correlación de -0.2074. La función de este receptor está regulada por las proteínas G, que a su vez modulan la concentración intracelular de  $Ca^{2+}$ , cAMP y activan a las enzimas quinasas PKA y MAPK3 / MAPK1. PCA3, que produce un ARN empalmado, largo y no codificante que se sobreexpresa en la mayoría de los tipos de células de cáncer de próstata, fue el segundo gen cuyos niveles altos de expresión tuvieron la segunda correlación más fuerte con un coeficiente de correlación de -0.19. Mientras que los niveles de expresión de los otros genes tuvieron correlaciones más débiles con el PSA.

GOLM1, proteína de membrana de Golgi 1, es codante de una proteína de respuesta celular a infección viral. El gen ACSM1, acil-coenzima A sintetasa ACSM1, mitocondrial, es codante para la enzima que cataliza la activación de ácidos grasos por CoA para producir un acil-CoA, el primer paso en el metabolismo de los ácidos grasos. GJB1, proteína beta-1 de unión de brecha, codifica una proteína que forma parte de un grupo de pares de canales transmembrana, a través de los cuales los materiales de bajo peso molecular se difunden de una célula a una célula vecina. ABCC4, proteína 4 asociada a resistencia a múltiples fármacos, hipotéticamente puede conformar una bomba de aniones orgánicos relevante para la desintoxicación celular.

**Tabla 27. Asociación de genes sobreexpresados con variables clínicas en Cpp**

Variable	Gene	p-valor	prueba estadística*
PSA	PCA3	6.05E-05 (-0.1900)	Correlación de Pearson
	OR51E2	1.16E-05 (-0.2074)	
	GOLM1	1.72E-03 (-0.1491)	
	ACSM1	4.15E-03 (-0.1364)	
	GJB1	0.0009485 (-0.157)	
	ABCC4	0.005953 (-0.1309)	
Escala de Gleason	PCA3	5.82E-11	Anova
	AMACR	2.64E-02	
	OR51E2	6.01E-13	
	GOLM1	7.50E-05	
	HPN	1.61E-06	
	GLYATL1	1.49E-08	
	TMEFF2	0.02019	
	GJB1	7.78E-10	
	PIGY	9.73E-03	
	ABCC4	0.0004481	
	TFF3	2.45E-05	
	PODXL2	0.0002682	
Estadio T	PCA3	1.07E-05	Anova
	AMACR	0.005673	
	DLX1	0.01863	
	OR51E2	3.791E-07	
	GOLM1	0.001536	
	GLYATL1	0.000002724	
	TMEFF2	0.0001398	
	GJB1	2.114E-09	
	CRISP3	4.65E-02	
	ABCC4	9.13E-04	
	TFF3	1.15E-07	
	PODXL2	0.01213	

\*Se indica el p-valor asociado a la prueba estadística y entre paréntesis el coeficiente de correlación de Pearson.

Por otro lado, se halló que la sobreexpresión de 12 genes se asoció con la escala de Gleason. Así, la sobreexpresión de PCA3, AMACR, OR51E2, GOLM1, HPN, GLYATL1, TMEFF2, GJB1, PIGY, ABCC4, TFF3 y PODXL2, se asoció significativamente a puntajes altos de Gleason (8, 9 y 10). El gen AMACR, alfa-metilacil-CoA racemasa, codifica una enzima que cataliza la interconversión de

estereoisómeros (R) y (S) de ésteres de acil-CoA grasos de cadena ramificada de alfa-metilo. HPN, serina proteasa hepsina, que codifica para una enzima que rompe sustratos del exterior de la célula y que regula la actividad proteolítica de factores de crecimiento, como HGF y MST1 / HGFL. El gen GLYATL1 codifica la proteína de tipo glicina N-aciltransferasa 1 que transfiere un grupo acilo al extremo N de glutamina. TMEFF2, tomoregulina-2, hipotéticamente puede funcionar como un factor de supervivencia para las neuronas hipocámpales y mesencefálicas, también podría regular la proliferación de células cancerosas al promover la fosforilación de ERK1 / 2.

El gen PIGY, fosfatidilinositol N-acetilglucosaminiltransferasa subunidad Y, codifica una proteína que es parte del complejo GPI-N-acetilglucosaminiltransferasa que inicia la biosíntesis de glicosilfosfatidilinositol (GPI). GPI se forma en el retículo endoplásmico y funciona como un anclaje para diversas proteínas, que podrían estar asociados a interacciones célula-célula. TFF3, factor 3 del trébol, participa en el mantenimiento y reparación de la mucosa intestinal; además de promover la movilidad de las células epiteliales en los procesos de curación. El gen PODXL2, proteína 2 tipo podocalixina, produce un ligando para selectinas vasculares. Media la rodadura rápida de los leucocitos sobre las superficies vasculares a través de interacciones dependientes de cationes divalentes de alta afinidad con las selectinas E, P y L.

También se observó que categorías altas del estadio T estuvieron asociadas a la sobreexpresión de 12 genes: PCA3, AMACR, DLX1, OR51E2, GOLM1,



### 6.3.5. GDEs asociados a la supervivencia

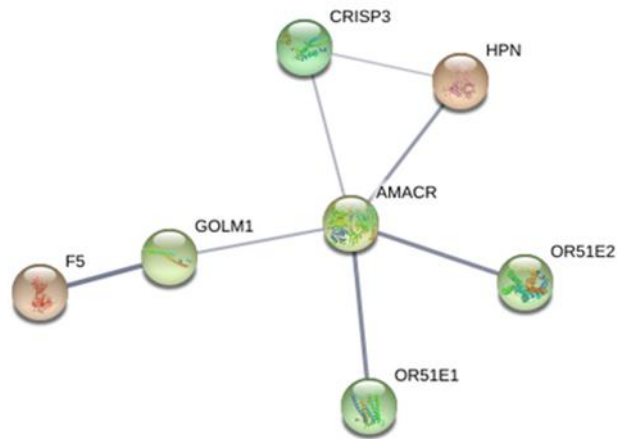
A través de GEPIA se identificaron GDEs asociados a la SG y a la SLE (Tabla 28). UBE2SP2, un gen codante para una enzima conjugadora de ubiquitina, fue el único gen sobreexpresado asociado a la SG. El gen SNHG12, gen que produce un ARN largo no codificante, fue el único gen subexpresado asociado a la SG. En tanto que 12 genes subexpresados se asociaron a la SLE, incluyendo a YJEFN3, que hipotéticamente podría desempeñar un papel en la espermatogénesis y la ovogénesis. Dicho gen tuvo el p-valor más significativo asociado a la SLE.

**Tabla 28. Evaluación de la asociación de GDEs en Cpp con variables pronósticas.**

Patrón de expresión	Variable de supervivencia	Gen	Nombre del gen	P-valor
sobreexpresados	SG	UBE2SP2	Enzima conjugadora de ubiquitina E2 S, pseudogen 2	1.01E-03
		SNHG12	Gen 12 del huésped del ARN nuclear pequeño	2.88E-03
		YJEFN3	Proteína 3 que contiene el dominio N-terminal YjeF	9.77E-08
		AHSA2	Activador putativo de la proteína de choque térmico de 90 kDa ATPasa homólogo 2	2.39E-07
		ATG16L2	Proteína 16-2 relacionada con la autofagia	1.18E-06
		EGFL8	Proteína 8 tipo Factor de crecimiento epidérmico	1.19E-06
		TTLL3	Tubulina monoglicilasa TTLL3	6.55E-07
		LY6G5B	Antígeno de linfocitos 6, miembro de la familia G5B	2.27E-06
Subexpresados	SLE	SMG1P7	Pseudogen SMG1	2.38E-06
		ATHL1	Proteína-glicosilgalactosilhidroxilisina glucosidasa	2.42E-06
		TAZ	Tafazzina	2.08E-07
		FAM13A-AS1	FAM13A ARN antisentido 1	2.00E-06
		LINC00342	ARN 342 intergénico largo no codificante	2.07E-06
		CCDC84	Proteína 84 que contiene el dominio de bobina enrollada	1.97E-06

## 6.4. Interactomas

Empleando STRING, se esquematizaron las redes de contacto (PPI) o interactomas para los GDEs en cáncer de próstata primario. Considerando los 50 genes más sobre-expresados en CP se obtuvo una red de PPI compuesta de siete proteínas: AMACR, CRISP3, OR51E1, OR51E2, GOLM1, F5 y HPN (Figura 52). Las otras 43 proteínas no mostraron interacción mutua en el diagrama de contactos.



**Figura 52. Red de interacción proteína-proteína (PPI) mediante STRING para genes sobreexpresados en CPp. Se muestra la red resultante al ingresar los 50 genes más sobreexpresados en CP. Se obtuvieron 29 nodos, de los cuales solo 7 mostraron conexión, con un grado medio de nodo de 0.483, un coeficiente promedio de agrupación local de 0.176 y un p-valor de enriquecimiento de  $4.74 \times 10^{-5}$ .**

Para los 7 genes sobre-expresados y conectados entre sí, STRING no arrojó alguna categoría estadísticamente significativa en el análisis GO ni en enriquecimiento de rutas KEGG, sin embargo se llevó el análisis a través de Genecodis, para los siete genes de la Fig. 52. Los procesos biológicos enriquecidos fueron la regulación de respuesta a la hormona tiroidea, el reflejo pilomotor, desmontaje de la membrana y regulación en la proliferación de hepatocitos, todos ellos vinculados al gen HPN. En cuanto a funciones biológicas enriquecidas, destacan actividades catalíticas del tipo racemasa y exopeptidasa, así como actividad de canales de potasio y receptores olfativos (Tabla 29).

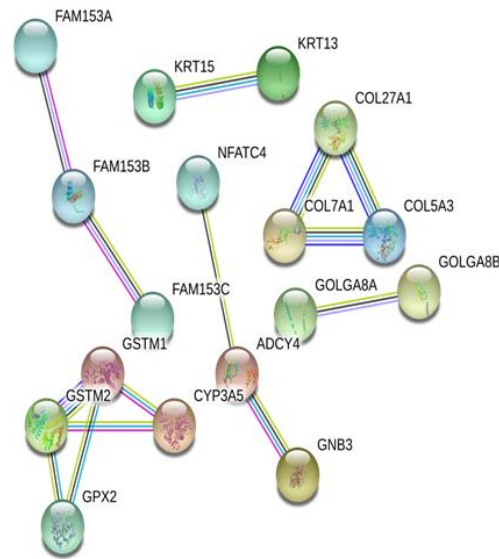
**Tabla 29. Los 5 principales términos de ontología de genes (GO) con el p-valor más bajo para los siete genes sobreexpresados presentes en la red PPI**

Categoría	Termino GO	Descripción	Conteo	p-valor	Genes
PB	GO:0097195	reflejo pilomotor	1	0.00218273	HPN
	GO:0097066	respuesta al estímulo de la hormona tiroidea	1	0.00218273	HPN
	GO:0034769	desmontaje de la membrana basal	1	0.00218273	HPN
	GO:2000611	regulación positiva de la generación de hormona tiroidea	1	0.00392822	HPN
	GO:2000347	regulación positiva de la proliferación de hepatocitos	1	0.00392822	HPN
FM	GO:0008111	actividad de la alfa-metilacil-CoA racemasa	1	0.00327409	AMACR
	GO:0070008	actividad exopeptidasa de tipo serina	1	0.00491027	HPN

	GO:0015269	actividad del canal de potasio activada por calcio	1	0.0114463	HPN
	GO:0004984	actividad del receptor olfativo	2	0.0130799	OR51E1,OR51E2
	GO:0005507	enlace de iones de cobre	1	0.022403	F5
CC	GO:0005886	membrana plasmática	4	0.0305175	OR51E1,OR51E2, HPN, F5
	GO:0005615	espacio extracelular	2	0.0387808	CRISP3, F5
	GO:0005782	matriz peroxisomal	1	0.038781	AMACR
	GO:0042581	gránulo específico	1	0.0388491	CRIPS3
	GO:0031093	luz de granulado alfa de plaquetas	1	0.0454999	F5
KEGG	Kegg:04610	Cascadas de coagulación y complemento	1	0.0176354	F5
	Kegg:04740	Transducción olfativa	2	0.00955577	OR51E2,OR51E1
	Kegg:04146	Peroxisoma	1	0.01545	AMACR
	Kegg:00120	Biosíntesis primaria de ácidos biliares	1	0.00653957	AMACR

Al evaluar los 50 primeros genes sub-expresados, se obtuvo una red PPI con 17 genes (figura 53). Se observó una red formada por GSTM1, GSTM2, CYP3A5 y GPX2, vinculados a procesos de desintoxicación celular, metabolismo de xenobiótico y proceso catabólico del nitrobenzeno. También destaca la red

formada entre COL27A1, COL7A1 y COL5A3, involucradas en la formación de trímeros de colágeno y componentes de matriz extracelular. También se identificó la red formada por NFATC4, ADCY4 y GNB3 vinculadas a la adicción a morfina, a la sinapsis GABAérgica y a la señalización endocannabinoide retrógrada. Además se generó una conexión entre dos genes, KRT15 y KRT13 asociados a la ruta de señalización de estrógeno. Por otra parte se generó la red con los genes FAM153A, FAM153B y FAM153C, pertenecientes a la familia FAM con función aún desconocida y la red GOLGA8A y GOLGA8B ambos asociados a la familia de proteínas que se encuentran en el aparato de Golgi de eucariotas.



**Figura 53. Red de interacción proteína-proteína (PPI) mediante STRING para genes subexpresados en Cpp. Se muestra la red resultante al ingresar los 50 genes más subexpresados en CP. Se obtuvieron 39 nodos, de los cuales solo 77 mostraron conexión, con un grado medio de nodo de 0.718, un coeficiente promedio de agrupación local de 0.368 y un p-valor de enriquecimiento de  $1.92 \times 10^{-6}$ .**

## 6.5. Predicción del daño de estructura y función

Se investigó el daño a nivel de la estructura y función de las proteínas codificadas por los genes más significativamente mutados en CPp. Se evaluaron los 5 primeros GSMs, que incluyeron a SPOP, TP53, FOXA1, MLL2 y MLL3 para CPp. De 31 mutaciones puntuales de SPOP reportadas en CPp, se predijo que 19 de ellas (61%) serían probablemente dañinas, en término de función proteica. Los residuos F102, F104 y W131 tuvieron las más altas frecuencias de alteraciones dañinas (Tabla 30). La mayoría de mutantes predichas dañinas (90%) mostraron tener un efecto energéticamente desestabilizador basándose en un cómputo teórico de la energía libre (Tabla 30). Sólo las mutaciones S119N y K134N mostraron tener un posible efecto neutral sobre la estabilidad energética de la proteína mutante.

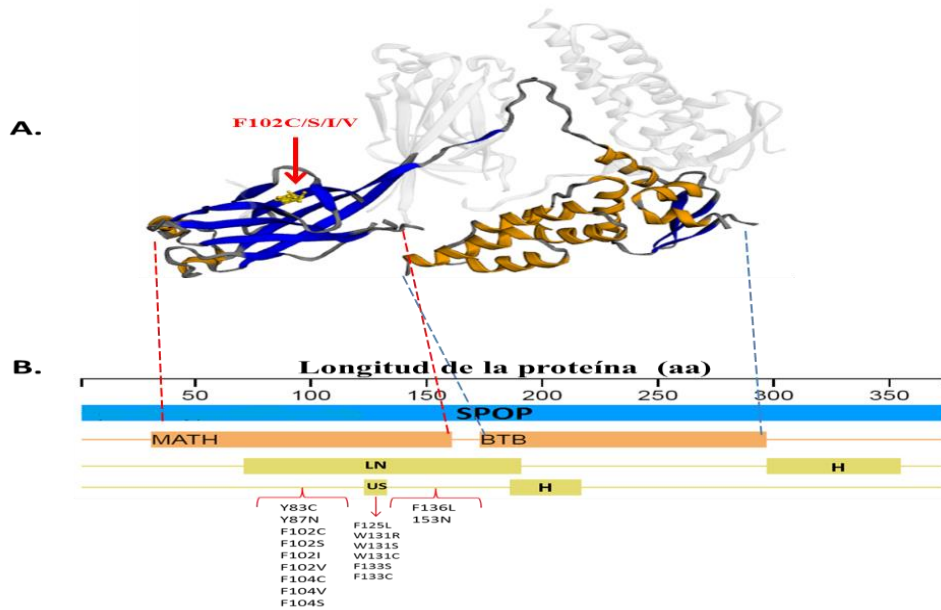
**Tabla 30. Predicción funcional y energética de las mutaciones puntuales probablemente dañinas en SPOP.**

Residuo Wild-Type	Posición	Residuo alterado	Predicción funcional (Polyphen)	Predicción de cambio de estabilidad (I-Mutant 3.0)	Valor de cambio de energía libre(CEL) (Kcal/mol)
Y	83	C	1	Desestabilizadora	-0.78
Y	87	N	0.995	Desestabilizadora	-1.06
F	102	C	1	Desestabilizadora	-1.74
F	102	S	1	Desestabilizadora	-1.81
F	102	I	0.998	Desestabilizadora	-0.94
F	102	V	0.998	Desestabilizadora	-1.77
F	104	C	1	Desestabilizadora	-1.69
F	104	V	0.993	Desestabilizadora	-1.78
F	104	S	1	Desestabilizadora	-1.81
S	119	N	1	Neutral	-0.13
F	125	L	1	Desestabilizadora	-0.84
W	131	R	1	Desestabilizadora	-1.27
W	131	S	0.999	Desestabilizadora	-1.8
W	131	C	1	Desestabilizadora	-1.71

F	133	S	0.995	Desestabilizadora	-1.87
F	133	C	1	Desestabilizadora	-1.84
K	134	N	1	Neutral	-0.24
F	136	L	0.995	Desestabilizadora	-1.32
D	153	N	0.997	Desestabilizadora	-0.82

La selección de mutantes probablemente dañinas se realizó con Biomuta v4.0. Para la predicción en el cambio de estabilidad de la proteína se empleó I-Mutant 3.0, donde el valor del cambio de energía libre (CEL) predice el cambio de estabilidad según lo siguiente:  $CEL < -0.5$ , cambio desestabilizador,  $CEL > 0.5$ , un incremento en la estabilización,  $-0.5 < CEL < 0.5$ , cambio neutral.

Todas las mutaciones que se predijeron dañinas en SPOP se albergaron en el dominio MATH, que está involucrado en la interacción con sustratos de proteína-ubiquitina ligasa, como H2AFY y BMI1. Asimismo, las 17 mutaciones desestabilizadoras se ubicaron en la región requerida para la localización nuclear de SPOP. Así mismo, seis mutaciones dañinas se encontraron en la zona de unión al sustrato (Figura 54)



**Figura 54. Localización de mutaciones “con pérdida de sentido” en SPOP. (A) Estructura tridimensional de los dominios proteicos MATH y BTB de SPOP, la flecha roja indica la posición F102 donde se ubicaron la mayor cantidad de mutaciones predichas dañinas y desestabilizadoras; se precisan los cambios aminoacídicos. La línea horizontal superior indica la longitud de la proteína dada en aminoácidos (aa). (B) Representación planar de los dominios y regiones de SPOP albergando mutaciones predichas dañinas y desestabilizadoras. En naranja los dominios MATH y BTB, en amarillo la región requerida para la localización nuclear (LN), regiones importantes para la Homodimerización (H) y la zona para Unión a Sustrato (US).**

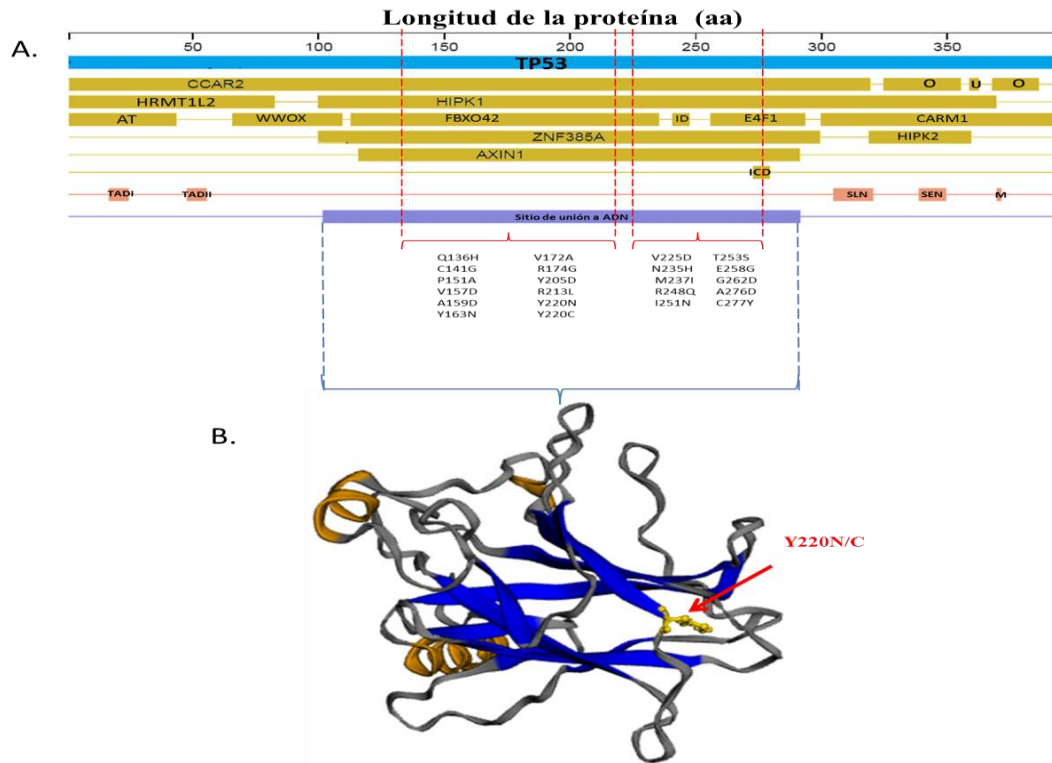
De las 43 mutaciones puntales en TP53 reportadas para CPp, 33 (77%) fueron predichas como probablemente dañinas y, de estas últimas, más de la mitad (67%) mostraron tener un efecto desestabilizador para la estructura proteica (Tabla 31).

**Tabla 31. Predicción funcional y energética de las mutaciones puntuales probablemente dañinas en TP53.**

Residuo Wild-Type	Posición	Residuo alterado	Predicción funcional (Polyphen)	Predicción de cambio de estabilidad	Valor de cambio de energía libre (Kcal/mol)
G	105	V	1	Neutral	-0.27
T	125	M	1	Neutral	-0.23
K	132	T	1	Neutral	-0.28
Q	136	H	1	Desestabilizadora	-0.77
C	141	G	0.998	Desestabilizadora	-0.77
P	151	A	0.98	Desestabilizadora	-1.26
P	152	L	0.982	Neutral	-0.37
V	157	D	1	Desestabilizadora	-1.42
A	159	D	0.984	Desestabilizadora	-0.55
Y	163	N	1	Desestabilizadora	-1.12
V	172	A	0.993	Desestabilizadora	-1.17
R	174	G	0.968	Desestabilizadora	-1.34
C	176	R	1	Neutral	-0.29
C	176	W	1	Neutral	-0.13
Y	205	D	0.999	Desestabilizadora	-1.41
R	213	L	1	Desestabilizadora	-0.69
Y	220	N	0.999	Desestabilizadora	-1.20
Y	220	C	1	Desestabilizadora	-1.02
V	225	D	0.958	Desestabilizadora	-1.48
N	235	H	1	Desestabilizadora	-0.55
M	237	I	1	Desestabilizadora	-0.84
C	238	W	1	Neutral	-0.28
R	248	Q	1	Desestabilizadora	-1.18
I	251	N	0.995	Desestabilizadora	-1.99
T	253	S	0.998	Desestabilizadora	-0.77
E	258	G	1	Desestabilizadora	-1.19
G	262	D	0.999	Desestabilizadora	-1.11
S	269	R	0.998	Neutral	-0.40
A	276	D	1	Desestabilizadora	-0.76
C	277	Y	1	Desestabilizadora	-0.51
P	278	L	1	Neutral	-0.37
D	281	Y	1	Neutral	-0.23
K	291	T	0.991	Neutral	-0.37

La selección de mutantes probablemente dañinas se realizó con Biomuta v4.0. Para la predicción en el cambio de estabilidad de la proteína se empleó I-Mutant 3.0, donde el valor del cambio de energía libre (CEL) predice el cambio de estabilidad según lo siguiente:  $CEL < -0.5$ , cambio desestabilizador,  $CEL > 0.5$ , un incremento en la estabilización,  $-0.5 < CEL < 0.5$ , cambio neutral.

Todas las mutaciones se albergaron en el sitio de unión al ADN que es también sitio de interacción intermolecular con otras proteínas como HIPK1, FBXO42, ZNF385A, CCAR2, entre otras (Figura 55). Una única mutación, R248Q, afectó la zona de interacción con el dominio 53BP2 SH3 (Figura 55).



**Figura 55. Localización de mutaciones “con pérdida de sentido” en TP53. (A) Representación planar de los dominios y regiones de TP53 albergando mutaciones predichas dañinas y desestabilizadoras. En amarillo, CCAR2, HRMT1L2, HIPK1, WWOX, FBXO42, HIPK2, ZNF385A, CARM1, AXIN1 y E4F1 indican las regiones en las cuales TP53 interactúa con dichas proteínas, región para la oligomerización (O), activación de la transcripción (AT), zona de interacción con el dominio 53BP2 SH3 (ID), zona de interacción con el DNA (ICD) y región de interacción con la proteína USP7 (U). En rosado los motivos de secuencia corta, TADI y TADII, señal de localización nuclear bipartita (SLN), señal de exportación nuclear (SEN), motivo KR-STA-K (M). En morado el sitio de unión a DNA. (B) Estructura tridimensional de la región de unión a DNA mostrando la posición Y220, que puede albergar dos variantes predichas dañinas (Y220N y Y220C). La línea horizontal superior indica la longitud de la proteína dada en aminoácidos (aa).**

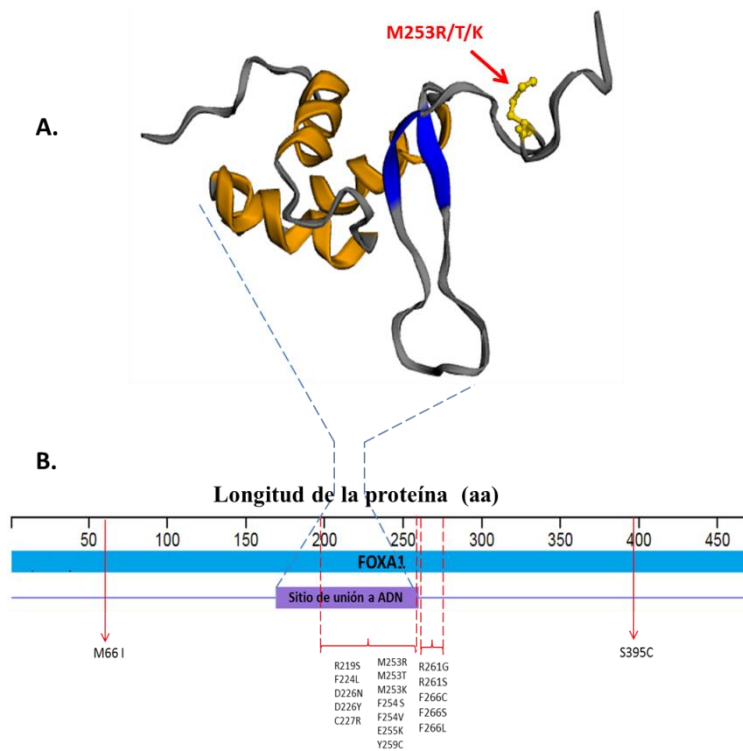
31 mutaciones puntuales fueron reportadas en FOXA1 para CPP, 26 de ellas (84%) fueron predichas como probablemente dañinas, de las cuales 19 (73%) fueron desestabilizadoras, seis neutrales y una incrementaba la estabilidad de la proteína (Tabla32). Los residuos M253 y F266 fueron los que tuvieron una mayor frecuencia de variantes dañinas (tres en cada uno).

**Tabla 32. Predicción funcional y energética de las mutaciones puntuales probablemente dañinas en FOXA1**

Residuo Wild-Type	Posición	Residuo alterado	Predicción funcional (Polyphen)	Predicción de cambio de estabilidad	Valor de cambio de energía libre (Kcal/mol)
M	66	I	0.963	Desestabilizadora	-0.80
S	217	A	0.976	Neutral	-0.48
R	219	S	1	Desestabilizadora	-1.17
F	224	L	1	Desestabilizadora	-1.61
D	226	N	0.978	Desestabilizadora	-1.60
D	226	Y	0.999	Desestabilizadora	-0.77
C	227	R	1	Desestabilizadora	-0.74
A	232	V	0.994	Neutral	-0.3
P	248	L	1	Neutral	-0.50
D	249	V	0.998	Incrementada	0.04
S	250	Y	1	Neutral	-0.07
M	253	R	0.998	Desestabilizadora	-1.05
M	253	T	0.998	Desestabilizadora	-1.31
M	253	K	0.996	Desestabilizadora	-1.72
F	254	S	1	Desestabilizadora	-2.02
F	254	V	0.999	Desestabilizadora	-1.61
E	255	K	1	Desestabilizadora	-0.87
Y	259	C	1	Desestabilizadora	-1.29
R	261	G	1	Desestabilizadora	-1.56
R	261	S	1	Desestabilizadora	-1.30
R	262	L	1	Neutral	-0.35
F	266	C	1	Desestabilizadora	-1.42
F	266	S	1	Desestabilizadora	-1.50
F	266	L	1	Desestabilizadora	-0.87
D	391	E	0.996	Neutral	-0.13
S	395	C	0.983	Desestabilizadora	-0.81

La selección de mutantes probablemente dañinas se realizó con Biomuta v4.0. Para la predicción en el cambio de estabilidad de la proteína se empleó I-Mutant 3.0, donde el valor del cambio de energía libre (CEL) predice el cambio de estabilidad según lo siguiente:  $CEL < -0.5$ , cambio desestabilizador,  $CEL > 0.5$ , un incremento en la estabilización,  $-0.5 < CEL < 0.5$ , cambio neutral.

De dichas variantes desestabilizantes en FOXA1, 12 se localizaron en el sitio de unión al ADN, siendo la posición M253 la que más mutaciones contuvo (M253R, M253T y M253K) (Figura 56). Otras siete variantes se localizaron en regiones no reconocidas como dominios o regiones funcionales de la proteína.



**Figura 56. Localización de mutaciones “con pérdida de sentido” en FOXA1. (A) Estructura tridimensional de la región de unión a DNA mostrando la posición M253, que puede albergar hasta tres variantes predichas dañinas (M253R, M253T y M253K). La línea horizontal superior indica la longitud de la proteína dada en aminoácidos (aa). (B) Representación planar de la región de unión a DNA de FOXA1 (en morado) albergando mutaciones predichas dañinas y desestabilizadoras.**

Para el gen MLL2, también conocido como KMT2D, se reportaron 28 mutaciones puntuales en Cpp, de las cuales 16 mutaciones (57%) fueron predichas como probablemente dañinas. Menos de la mitad (43%) de las mutaciones catalogadas como dañinas resultaron ser desestabilizantes (Tabla 33), siendo el residuo R3547 el único que contuvo dos variantes desestabilizadoras.

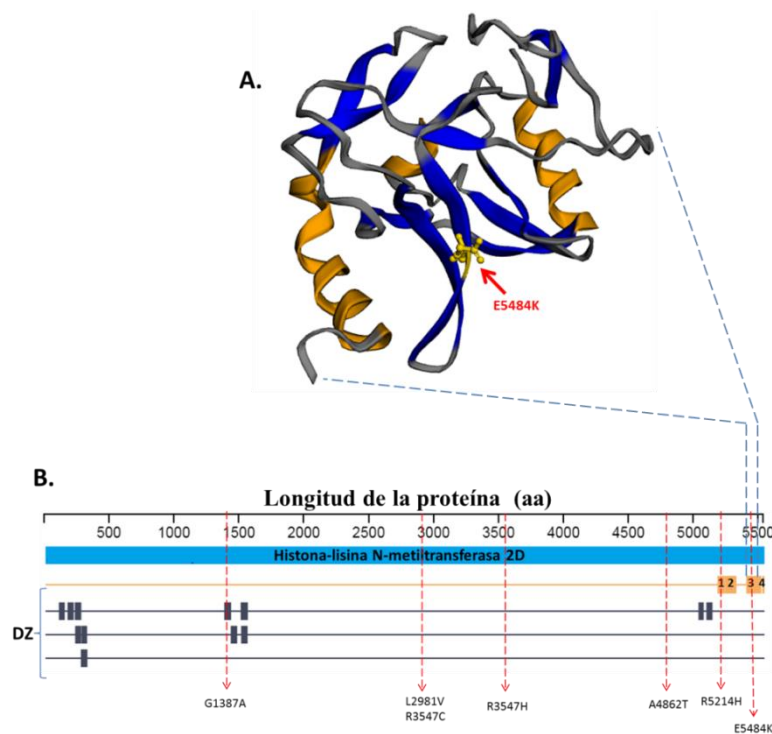
**Tabla 33. Predicción funcional y energética de las mutaciones puntuales probablemente dañinas en MLL2**

Residuo Wild-Type	Posición	Residuo alterado	Predicción funcional (Polyphen)	Predicción de cambio de estabilidad	Valor de cambio de energía libre (Kcal/mol)
D	1121	Y	0.998	neutral	-0.06
P	1232	L	0.998	neutral	-0.39
R	1290	W	0.999	neutral	-0.49
G	1384	V	1	neutral	-0.24
G	1387	A	1	Desestabilizadora	-0.64
N	1602	D	0.979	neutral	-0.08
D	2056	G	1	neutral	-0.38
R	2460	L	0.995	neutral	-0.33
L	2981	V	0.97	Desestabilizadora	-1.51
R	3547	C	1	Desestabilizadora	-0.88
R	3547	H	1	Desestabilizadora	-0.57
A	4862	T	0.97	Desestabilizadora	-0.60
C	5109	F	0.999	neutral	-0.21
R	5174	W	1	neutral	-0.40
R	5214	H	1	Desestabilizadora	-1.51
E	5484	K	1	Desestabilizadora	-1.23

La selección de mutantes probablemente dañinas se realizó con Biomuta v4.0. Para la predicción en el cambio de estabilidad de la proteína se empleó I-Mutant 3.0, donde el valor del cambio de energía libre (CEL) predice el cambio de estabilidad según lo siguiente:  $CEL < -0.5$ , cambio desestabilizador,  $CEL > 0.5$ , un incremento en la estabilización,  $-0.5 < CEL < 0.5$ , cambio neutral.

La variante de MLL2, G1387A, predicha dañina y desestabilizadora afectó una región que alberga dedos de Zinc, importantes para la asociación con ADN. Mientras que la mutante R5214H se ubicó en el dominio FYR N-terminal,

involucrado en la asociación con módulos que reconocen modificaciones de histonas de proteínas de varios eucariotas. La mutación E5484K se encontró en el dominio SET 8, que asociado con las proteínas relacionadas con la miotubularina modulan el control del crecimiento (Figura 57). Otras cuatro variantes se localizaron en regiones no reconocidas como dominios o regiones funcionales de la proteína.



**Figura 57. Localización de mutaciones “con pérdida de sentido” en MLL2. (A) Estructura tridimensional del dominio SET mostrando la mutación E5484K. La línea horizontal superior indica la longitud de la proteína dada en aminoácidos (aa). (B) Representación planar de las regiones y dominios de MLL2 albergando mutaciones predichas dañinas y desestabilizadoras. En anaranjado los dominios FYR N-terminal (1), FYR C-terminal (2), SET (3) y Post-SET (4). En color plomo las regiones que albergan dedos de Zinc (DZ), en particular la región PHD-tipo 4 que contiene la mutación G1387A.**

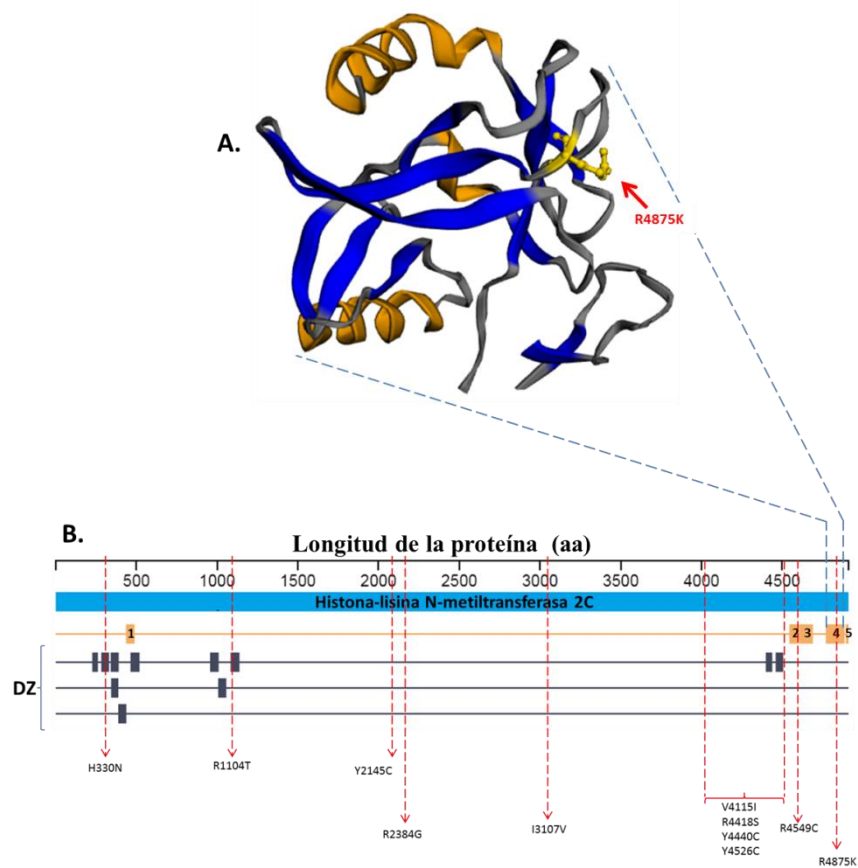
En el caso del gen MLL3 (KMT2C) se reportaron 33 mutaciones puntuales para Cpp, de las cuales 15 (45%) fueron identificadas como probablemente dañinas. La gran mayoría (73%) de variantes dañinas tuvieron un comportamiento desestabilizante para el producto proteico y solos 4 mutaciones resultaron con un efecto neutro (Tabla 34).

**Tabla 34. Predicción funcional y energética de las mutaciones puntuales probablemente dañinas en MLL3.**

Residuo Wild-Type	Posición	Residuo alterado	Predicción funcional (Polyphen)	Predicción de cambio de estabilidad	Valor de cambio de energía libre (Kcal/mol)
H	330	N	0.997	Desestabilizadora	-0.52
D	372	N	0.997	Neutral	-0.35
R	1104	T	0.999	Desestabilizadora	-0.66
Y	2145	C	0.998	Desestabilizadora	-1.18
R	2384	G	1	Desestabilizadora	-0.76
Q	2397	H	0.998	Neutral	-0.29
I	3107	V	0.988	Desestabilizadora	-0.75
V	4115	I	0.999	Desestabilizadora	-0.65
C	4198	W	1	Neutral	0.06
R	4418	S	0.997	Desestabilizadora	-0.83
C	4432	R	1	Neutral	-0.25
Y	4440	C	0.988	Desestabilizadora	-0.84
Y	4526	C	0.999	Desestabilizadora	-1.22
R	4549	C	1	Desestabilizadora	-1.07
R	4875	K	0.997	Desestabilizadora	-0.74

La selección de mutantes probablemente dañinas se realizó con Biomuta v4.0. Para la predicción en el cambio de estabilidad de la proteína se empleó I-Mutant 3.0, donde el valor del cambio de energía libre (CEL) predice el cambio de estabilidad según lo siguiente:  $CEL < -0.5$ , cambio desestabilizador,  $CEL > 0.5$ , un incremento en la estabilización,  $-0.5 < CEL < 0.5$ , cambio neutral.

Las variantes predichas dañinas y desestabilizadoras de KMT2C, H330N, R1104T y Y4526C, se ubicaron en regiones que albergan dedos de Zinc (Figura 58). La mutación R4549C se ubicó en el dominio FYR N-terminal, mientras que la R4875K en el dominio SET (Figura 58). Otras seis variantes se localizaron en regiones no reconocidas como dominios o regiones funcionales de la proteína.



**Figura 58. Localización de mutaciones “con pérdida de sentido” en MLL3. (A) Estructura tridimensional del dominio SET. La línea horizontal superior indica la longitud de la proteína dada en aminoácidos (aa). (B) Representación planar de las regiones y dominios de MLL3 albergando mutaciones predichas dañinas y desestabilizadoras. En anaranjado los dominios DHHC (1), FYR N-terminal (2), FYR C-terminal (3), SET (4) y Post-SET (5). En color plomo las regiones que albergan dedos de Zinc (DZ), en particular las regiones PHD-tipo1 y PHD-tipo7 que contienen las mutaciones H330N y R1104T respectivamente.**

## 6.6. Determinación de la frecuencia de las mutaciones conductoras en subgrupos divididos por fenotipo.

La evaluación de la frecuencia de mutaciones conductoras en CP se hizo a través de cBioPortal en la cohorte TCGA. Se identificaron 16 mutaciones conductoras caracterizadas en CP obtenidas de “Cancer Genome Interpreter” (Tabla 35) (50). Las mutaciones conductoras comprometieron ocho genes incluyendo KLF6, AR y EPHB2, con al menos tres mutaciones cada uno. La mayoría de mutaciones conductoras del CP (75%) fueron del tipo missense, mientras que solo el 15% fueron del tipo trunco. El cromosoma 10 alberga la mayor cantidad de mutaciones conductoras del CP.

**Tabla 35. Mutaciones conductoras en CP extraídas de CGI**

Cromosoma	Gen	Mutación missense
14	AKT1	E17K L702H
x	AR	F877L T878A
22	CHEK2	E239*
7	EGFR	G735S A279S
1	EPHB2	D679N Q723* L169P S137*
10	KLF6	A123D S116P W64R
7	MAD1L1	R59C
10	PTEN	Y188*

De dichas variantes, sólo seis se encontraron presentes en la población de estudio de CP TCGA (Tabla 36). Dos mutaciones conductoras se observaron en la cohorte de Cpp. Una de ellas fue la mutación E17K del gen AKT1, asociado a la regulación de la maquinaria apoptótica, presente en una frecuencia de 0.2% en la población. Mientras que la mutación S137\* del gen KLF6, un activador transcripcional, estuvo presente en el 0.1% en los individuos. En CP metastásico se identificaron cuatro mutaciones conductoras incluyendo L702H, T878A y F877L del gen AR, que codifica para el receptor de andrógeno, todas ellas ausentes en Cpp, y E17K de AKT1. Esta última variante estuvo presente tanto en Cpp como metastásico siendo más frecuente en Cpm.

**Tabla 36. Comparación de mutaciones conductoras para Ca. Próstata primario y metastásico presentes en la cohorte TCGA.**

Tipo de tumor	Gen	Mutación	Frecuencia
Primario	AKT1	E17K	0.23%
	KLF6	S137*	0.06%
Metastásico	AR	L702H	3.43%
		T878A	3.05%
		F877L	0.13%
		AKT1	E17K

## 6.7. SNPs de riesgo a CP

A través de GWAS y cBioPortal se evaluaron los SNPs de riesgo a Cpp y metastásico. Para el estadio primario de la enfermedad se identificaron 9 SNPs del tipo pérdida de sentido (Tabla 37). Sin embargo, ninguno de ellos estuvo presente en la cohorte TCGA evaluada en cBioPortal. Para Cpm no se observó ningún SNPs del tipo pérdida de sentido, siendo todos del tipo variante de intrón los cuales no fueron analizados en el presente estudio.

**Tabla 37. SNPs de riesgo a Cpp del tipo pérdida de sentido.**

Gen	código del SNP	Cambio de aminoácido	Cambio de nucleótido	Referencia
SH2B3	rs3184504	W262R	TGG/AGG	(51)
KLK3	rs17632542	I179T	ATT/ACT	(52)
CCHCR1	rs130067	E275D	GAA/GAC	(53)
MLPH	rs2292884	H347R	CAT/CGT	
HOXB13	rs138213197	G84E	GGA/GAA	
INCENP	rs2277283	M506T	ATG/ACG	(54)
ATM	rs1800057	P1054H	CCT/CAT	
CDKN1B	rs2066827	V109D	GTC/GAC	
RFX7	rs33984059	L1256P	CTG/CCG	

## 7. Discusión

EL CP es un problema de salud pública a nivel mundial y en los últimos años los esfuerzos por mejorar la comprensión de esta enfermedad se han centrado en la identificación de marcadores diagnósticos y pronósticos, así como en el reconocimiento de dianas terapéuticas que permitan hacer frente a esta neoplasia. Las tecnologías de alto rendimiento han puesto a disposición una gran cantidad de datos ómicos que permiten evaluar los patrones de alteraciones genéticas que albergan los tumores. Toda esta información en gran parte está disponible públicamente gracias a proyectos genomas en cáncer como TCGA, que facultan a los investigadores contar con datos no solo genómicos, sino también clínicos que permiten una evaluación integral del perfil de la enfermedad (10,55,56). Para el CP y otras neoplasias se han conducido estudios secundarios utilizando datos de cohortes de TCGA, donde se describen tanto el panorama mutacional así como la correlación de algunas variantes genotípicas con el fenotipo (57,58). Así también algunos investigadores han empleado TCGA para explorar posibles candidatos a marcadores para el CP (59) y otros para entender mejor la oncogénesis de esta neoplasia (47). TCGA también ha ayudado a identificar marcadores pronósticos para otros tipos de neoplasia como el cáncer de ovario y el de pulmón (11,12).

En el presente trabajo de investigación, mediante el análisis de datos genómicos e información clínica, se identificaron alteraciones genéticas frecuentemente presentes en CP y se establecieron asociaciones de dichos caracteres con el fenotipo de la enfermedad incluyendo el pronóstico. El presente estudio detalla el

panorama de mutaciones genéticas del CP, algunas de las cuales podrían en un futuro ser validadas como marcadores de esta enfermedad para beneficio de los pacientes. Se han identificado genes cuyas alteraciones genéticas, incluyendo mutaciones puntuales, ANCs y expresión diferencial en Cpp, están asociadas al estadio patológico T y N, al puntaje de Gleason, al PSA y a variables de supervivencia como la SG y SLE. En tanto que, para CPm se han identificado principalmente mutaciones puntuales asociadas a la SG.

La población del presente estudio se caracterizó por ser pacientes con Cpp, con una edad promedio al momento del diagnóstico de 61 años, mientras que para el grupo de pacientes con CPm fue de 67 años, lo cual está en concordancia con lo previamente reportado por Armenia en relación a la diferencia de edades entre pacientes con Cpp y CPm (60). La raza blanca fue principalmente predominante en la población de estudios albergados en cBioPortal, siendo una minoría la raza negra y asiática. Ello podría representar un sesgo importante en estudios genómicos donde la variabilidad genética ligada a la raza no se ve representada por los sujetos enrolados en estudios del cáncer a nivel mundial, lo que finalmente se traduce en diferencias perjudiciales al momento de la toma de decisiones clínicas en poblaciones menos favorecidas (61). La SG de entre 0-5 años y la SLE de entre 0-3 años fueron predominantes en los individuos enrolados, lo que refleja que este tipo de neoplasia, en concordancia con lo reportado por Globocan 2018, se sigue ubicando como una de las principales causantes de muertes en hombres a nivel mundial. La escala de Gleason de 9

estuvo mayoritariamente presente en la población, lo que indica que la mayoría de los casos aún son detectados en una etapa avanzada (62)

TP53 y SPOP fueron los únicos dos genes significativamente mutados comunes a CPp y CPm, sin embargo SPOP tuvo una frecuencia de mutación en el estadio temprano de la enfermedad mayor que en la fase avanzada lo que concuerda con hallazgos previos (60,63,64). Lo contrario sucedió con TP53, estando presente en una mayor frecuencia en CPm. El análisis de ontología de GSMs confirmó que, tanto en CPp como en CPm, el proceso de regulación de la transcripción está enriquecido. Sin embargo en CP primario adicionalmente se han determinado genes significativamente mutados implicados en proliferación y supervivencia celular incluyendo a PTEN, PIK3CA, CDKN1B y AKT1. En tanto que en CP metastásico sobresalen TP53 y BRAC2, ligados a procesos de respuesta al daño del ADN, tal como ha sido previamente reportado en CPm (60,65). Adicionalmente en CP primario se han identificado que los reguladores epigenéticos KMT2D y KMT2C poseen altas frecuencias de mutación, lo cual coincide con hallazgos anteriores (34,66). Se sabe que KMT2D modula la expresión de genes vinculados a la vía p53, a la señalización de colestasis y a la señalización mediada por AMPc (67).

Trabajos previos han identificado múltiples vías alteradas recurrentemente en CP, incluyendo la señalización de andrógenos, reparación de ADN y las rutas de señalización de PI3K y AKT) (34,68). En el presente trabajo se identificaron genes significativamente mutados en CPp asociados a regulación de la

transcripción desde el promotor de la ARN polimerasa II, en la regulación negativa del proceso apoptótico y en vías de señalización intracelular como la mediada por el receptor del factor de crecimiento de fibroblastos, la ruta de señalización por fosfatidilinositol y la señalización en cascada de la proteína quinasa B. En contraste, para el estadio metastásico, los genes significativamente mutados se involucraron mayoritariamente a la regulación de la transcripción, mecanismos de reparación al daño del ADN y en la regulación de la actividad de la fosfoproteína fosfatasa. Estos hallazgos pueden sugerir que en el estadio temprano de la enfermedad el tumor potencie vías que le permitan crecer y proliferar, mientras que en etapas avanzadas se esperaría que los mecanismos de invasión estén más activos y le permitan a las células exacerbar el cáncer.

Más allá de una mera descripción de las mutaciones puntuales, es importante reconocer la asociación de dicha huella genómica con la manifestación clínica de la enfermedad. Aplicando pruebas estadísticas de correlación, se estableció la asociación de ciertos cambios en el genoma con los caracteres fenotípicos de los sujetos. Se identificaron 3 genes significativamente mutados en CPr que al encontrarse mutados se correlacionaban con al menos a una variable clínica. Mutaciones en el gen ZFP36L2, que regula la respuesta a factores de crecimiento, se asociaron a una menor SG de los pacientes. Anteriormente se ha reportado que este gen promueve la agresividad de las células cancerosas en el adenocarcinoma ductal pancreático (69). Han et al también reportaron que ZFP36L2 es un potencial indicador de pronóstico del cáncer de vejiga muscular invasivo (70). Por otra parte, mutaciones puntuales del gen supresor de tumores

TP53, se correlacionaron tanto a categorías altas del estadio patológico T y N de la enfermedad, así mismo este gen fue reconocido como un GFM en CPp y mutaciones en él tuvieron que ver con una menor SLE. Tuna et al. describe que algunas mutaciones en TP53 están asociadas a una SG y SLE desfavorable en pacientes con cáncer de ovario seroso de alto grado (71). En particular se remarcan las mutaciones en los residuos G266, Y163C y R282, todas vinculadas al mal pronóstico, las cuales se localizan en residuos también dañados en CPp identificados en esta tesis. Por otro lado, mutaciones en ZMYM3, un componente del complejo multiproteico histona desacetilasa, el cual es un silenciador génico, estuvieron asociadas con un mayor puntaje de Gleason. Leung et al. describió que ZMYM3 regularía la localización de BRCA1 en el ADN dañado para promover su reparación (72), y esto reflejaría la implicancia de este gen en procesos de mantenimiento del material genético previo a acumulaciones dañinas que promueven el cáncer. Sin embargo, aún no se ha establecido si su rol en la neoplasia de próstata se limita solo a mecanismos de reparación.

En la evaluación de los genes frecuentemente mutados en CPp, se identificaron dos genes asociados a la SG y seis genes asociados a la SLE de los pacientes. Se encontró que pacientes con mutaciones en el gen LRP1B, proteína 1B relacionada con el receptor de lipoproteínas de baja densidad presentaron una menor SG en comparación a los pacientes que no presentaron mutaciones en dicho gene. Se sabe que LRP1B promueve la proliferación celular (73) y su estado mutacional ha mostrado ser un marcador de buena respuesta a inmunoterapia en pacientes con melanoma y cáncer de pulmón de células no pequeñas (74). Se postula que

LRP1B sería un importante regulador de la respuesta inflamatoria y de la progresión en cáncer y que sus altas frecuencias de mutaciones en varias neoplasias estarían vinculadas tanto al tamaño del gen (16 kbp) como a su ubicación cerca del sitio frágil FRA2F (75). CSMD3, la proteína 3 que contiene dominios CUB y sushi, fue otro gen cuyas mutaciones se asociaron a un menor tiempo de supervivencia, además de estar asociado a una menor SLE en pacientes con CPp. La familia CSMD alberga estructuras parecidas en la cual cada una se caracteriza por contener catorce dominios CUB, cada dominio separado del otro por un dominio de sushi, seguido de una matriz compacta de dominios de sushi, un dominio transmembrana y una cola citoplasmática corta, en particular la proteína CSMD1 se perfila como un supresor de tumores (73). Mutaciones en CSMD3 se asociaron con un mejor pronóstico en pacientes con carcinoma de células escamosas esofágicas, lo cual es opuesto a los hallazgos de la presente investigación para CP (74). No sólo mutaciones en CSMD3 estarían asociadas con el fenotipo del cáncer sino también alteraciones en su expresión. Así, la expresión de CSMD3 está correlacionada con la SG y la SLE en neuroblastoma (75) así como con la invasión linfática en cáncer colorrectal (76).

En CPm se identificaron tres genes asociados a la SG que incluyeron TMPRSS2, KMT2C y ZFHX3. Los pacientes con mutaciones puntuales en cualquiera de dichos genes presentaron una SG mayor en comparación a los pacientes que tuvieron el gen sin mutar. Contrariamente a lo encontrado en nuestro trabajo, un estudio previo encontró que mutaciones en KMT2C, se asoció a un peor pronóstico en cáncer de pulmón temprano (77). Por otro lado, el gen ZFHX3 tuvo

una alta frecuencia de mutación en un estudio de cáncer endometrial, encontrándose que mutaciones en dicho gen se asociaron a tumores de alto grado, a una invasión más frecuente del espacio linfovascular y a una SG y SLE desfavorables (78).

Otra de las alteraciones recurrentes en el CP son las ANCs, donde las deleciones de segmentos cromosómicos son encontradas en la etapa inicial del tumor, mientras que las amplificaciones son predominantes a medida que la enfermedad se va desarrollando (79). Los cromosomas 6, 7, 8, 10, 13 y 17 son los más frecuentemente alterados en CP, siendo los brazos 8p, 1q, 10q, 13q, 16q, 5q, 6q y 17p las regiones que albergan mayor número deleciones (79,80), lo que está en concordancia con nuestros hallazgos como más adelante se detallan. En contraste, se ha reportado que los brazos 8q y Xq muestran principalmente amplificaciones y los brazos 7p y 7q pueden mostrar tanto deleciones como amplificaciones (80). En el presente estudio, encontramos que el supresor de tumor PTEN, ubicado en el brazo 10q, fue el gen más frecuentemente delecionado, tanto en CPp como CPm, lo cual conduciría a la pérdida parcial o total de la función proteica. Además, PTEN fue el único gen cuya deleción se correlacionó a un estadio patológico T mayor y a puntajes de Gleason altos. Las ANCs afectan de forma distinta a regiones cromosómicas, lo cual se evidenció en este estudio donde las deleciones ocurrieron más frecuentemente en el brazo cromosómico 6p, en CPp, y en brazo 10q en CPm, lo que está en concordancia con lo reportado para esta neoplasia (79).

Otro de los genes frecuentemente delecionados en CPp en la cohorte en estudio fue MAP3K7, ubicado en el brazo cromosómico 6q. Se ha reportado previamente que deleciones en MPA3K7 se asocian al CP de alto riesgo (81) y a la agresividad del tumor en pacientes con CP ERG-negativo (82). MPA3K7 cumpliría un rol de supresor de tumor en CP y con la deleción conjunta de CHD1 promovería la agresividad de esta neoplasia (83). CHD1, que codifica una proteína asociada a histona que es requerida para la reparación de la rotura del ADN de doble cadena, también fue un gen frecuentemente delecionado hallado en este trabajo. LRP1B también estuvo frecuentemente delecionado en la cohorte evaluada, que fue además identificado como un GSM asociado a la SG como se mencionó anteriormente.

La deleción de los genes GTF2F2, TPT1 y GPALPP1 estuvo asociada con un mayor estadio patológico T. GTF2F2 es un factor de transcripción que se une a la ARN polimerasa II, mientras que TPT1 participa en la fijación de calcio y la estabilización de microtúbulos. La pérdida de estos genes podría promover la supervivencia de las células tumorales o potenciar mecanismos de invasión. Chen et al. evaluó la regulación de TPT1 por p53, donde este último sobrerregulaba a TPT1, lo que resultaba en una reducción de estrés oxidativo, de la apoptosis y promoción de la supervivencia celular en respuesta a H<sub>2</sub>O<sub>2</sub> (84). En el caso de los genes GTF2F2 y GPALPP1 no se ha reportado algún estudio previo que haya vinculado algunos de estos genes con el CP.

En el grupo de genes frecuentemente amplificados en CP, NDRG1 fue el gen más frecuentemente amplificado en CPp, el cual también estuvo amplificado en CPM pero en una menor frecuencia. Se ha visto que la sobreexpresión de NDRG1 en PC suprime el fenotipo metastásico y su expresión se correlaciona con una menor progresión de la enfermedad (85). Además, la sobreexpresión de este gen modula proteínas y vías importantes de señalización que incluyen las mediadas por TGF $\beta$ , WNT y PI3K/AKT (86). En el CPM el gen más frecuentemente amplificado fue el receptor de andrógenos (AR) ubicado en el brazo Xq. El AR media una ruta de señalización central en la función normal de la próstata y se conoce que la actividad constitutiva de esta proteína conduce la proliferación celular, el crecimiento y la pérdida de apoptosis lo que resulta en la tumorigénesis (87). Por ello, la amplificación de AR sería uno de los mecanismos promotores de la carcinogénesis de próstata. La amplificación de los genes CASC8, MYC, POU5F1B, CCAT1, PCAT1 y PCAT2 estuvo asociada a una menor SG, a categorías altas del estadio patológico T y N, a una mayor escala de Gleason y a un PSA alto. Liu et al. mostró la asociación de CASC8, junto a otros 7 lncRNAs, con la SG de pacientes con cáncer de esófago (88). Un reciente trabajo conducido por Hongyan et al. mostro que la sobreexpresión de MYC estuvo asociado a altos puntajes de Gleason y correlacionado a un pronóstico desfavorable en paciente chinos con CP (89). MYC es un factor de transcripción que activa la transcripción de genes asociados al crecimiento, la sobreexpresión de este gen se ha asociado con alteraciones genéticas como amplificaciones y translocaciones en varias enfermedades neoplásicas (90). La amplificación de MYC ha sido reportada como un marcador potencial para la progresión de la enfermedad y el

pronóstico en CP (91), sin embargo la correlación de la sobreexpresión y/o amplificación de este gen con variables clínico-patológicas sigue siendo controversial (92–94). POU5F1B, también conocido como OCT4-pg, es uno de los pseudogenes de OCT4 y tiene un 95% de homología con OCT4, que a su vez juega un papel importante en el desarrollo y la reprogramación de células somáticas en células madre pluripotentes (95). Estudios previos han reportado a POU5F1B como un gen de susceptibilidad en cáncer colorrectal, en cáncer de vejiga y en leucemia linfocítica crónica (96–98). También se ha informado que promueve la proliferación, la tumorigenicidad y la angiogénesis en el cáncer gástrico en sinergia con MYC (99).

Un estudio llevado a cabo por Yaozhen et al. mostró que la sobreexpresión de POU5F1B promovió la proliferación del hepatocarcinoma celular, además encontraron que la expresión de POU5F1B se correlacionó positivamente con la activación de la vía AKT (95), vía de transducción de señales que promueve la supervivencia y el crecimiento en respuesta señales extracelulares. CCAT es un gen que produce un ARN largo no codificante asociado a la formación de tumores, cuya sobreexpresión se hayo que también estuvo asociado a un pronóstico desfavorable. Algunos estudios sugieren que CCAT1 fomenta la proliferación del CP actuando como un andamio para DDX5 y para el complejo transcripcional del AR, facilitando la expresión de genes regulados por el AR y estimulando así la progresión de la enfermedad (100). Contrariamente a lo encontrado en el presente trabajo, Chen et al. han señalado que la subexpresión de CCAT1 estaría asociada a una menor proliferación celular del CP en una línea

celular con alto potencial metastásico ( Células PC-3) (101). PCAT1 y PCAT2 son también genes productores de ARN no codificantes (lncRNA) vinculados al CP. La expresión de PCAT1 se vincula a la progresión de CP resistente a castración. El lncRNA-PCAT1 desorganizaría el complejo PHLPP/FKBP51/IKK $\alpha$ , proteínas reguladoras en la activación de la rutas de señalización AKT y NF- $\kappa$ B que contribuiría a la progresión de la neoplasia.

La variación de la expresión génica en Cpp también fue evaluada en el presente estudio, encontrándose que PCA3 fue el gen más sobreexpresado en la cohorte TCGA. Se ha reportado que PCA3 regula la supervivencia celular del CP a través de la modulación de la señalización del receptor de andrógenos, además regulando la expresión de varios genes sensibles al andrógeno y asociados con el cáncer, incluidos los marcadores de transición epitelial-mesenquimal (EMT) y genes reguladores de la expresión y señalización celular (102). Se encontró que la expresión de PCA3 se correlaciono negativamente al valor de PSA y se asoció a puntajes altos de Gleason y a los estadios patológicos T y N mayores. Se ha descrito que la expresión de PCA3 se asocia con valores de la escala de Gleason  $\geq 7$  así como a un estadio patológico T3 y un volumen del tumor  $\geq 0.5$ mL (103). De hecho, la prueba de PCA3 en orina ha sido aprobada por la FDA y es utilizada para el diagnóstico de CP. Sin embargo, su empleabilidad como prueba de primera línea o para detectar un PC de alto grado sigue siendo controvertida ya que se ha reportado una baja sensibilidad en biopsia, donde la puntuación de la prueba de PCA3 no detecta el 36% de tumores avanzados en hombres con valores bajos de PSA (104). GJB1 y ABCC4 fueron otros dos genes sobreexpresados y

asociados a las cuatro variables clínicas al igual que PCA3. GJB1 es un tipo de conexina de unión gap, conocida también como Cx32. Se sabe que las conexinas y sus canales de membrana celular aportan a la regulación de la proliferación celular y su desregulación está asociada a la carcinogénesis (105), por ejemplo algunas tienen funciones compartimentales en las glándulas mamarias y están asociadas al pronóstico en cáncer de mama (106). En un estudio previo se ha reportado que TLR9, uno de los receptores tipo Toll (TLR) que juegan un papel importante en la tumorigénesis y el progreso del CP, regula la invasión y metástasis del CP al alterar la expresión de ciertos genes, entre ellos GJB1 (107). ABCC4 es una proteína asociada a resistencia a diversos fármacos, anteriormente se ha descrito que la sobreexpresión de este gen reduce la eficacia del docetaxel, un agente quimioterapéutico empleado en el tratamiento del CP, en células tumorales in vitro (108). He et al. también encontraron que altos niveles de expresión de ABCC4 se asociaron con una SG desfavorable en pacientes con CP (58).

Por otra parte se identificó la sobreexpresión de AMACR en CPp, una enzima importante en la inversión quiral en el metabolismo de ácidos grasos de cadena ramificada y que también modula la entrada de ácidos grasos de cadena ramificada en la  $\beta$ -oxidación en peroxisomas y mitocondrias. La sobreexpresión de este gen, junto a la sobreexpresión de NKX3-1, GOLM1 y TRPM8, ha sido reportada como un potencial marcador para CP de alto riesgo, teniendo una asociación con la extensión extracapsular, el estadio tumoral y la invasión seminal vesicular (109). Además se ha mostrado que la supresión de la expresión de AMACR por siRNA (ARN de silenciamiento) reduce significativamente la

proliferación de una línea celular de cáncer de próstata sensible a andrógenos (110). También se identificó que la sobreexpresión de TMEFF2, que codifica a un miembro de la familia de tomoregulinas de proteínas transmembrana, estuvo asociada al estadio patológico T y a la escala de Gleason. Contrariamente a nuestros hallazgos, Georgescu et al. indica que la subexpresión de TMEFF2 se asoció a una menor SLE, teniendo un rol en la regulación del ciclo celular, aunque este patrón de expresión podría servir como marcador solo para una subclase de PC (111).

Por otra parte, Gao et al. refieren que una alta expresión de TMEFF2 se asocia a un pronóstico desfavorable en pacientes con cáncer endometrial, así mismo, indican que la subexpresión de este gen redujo la activación de las vías de señalización MAPK y PI3K y reprimió la transición epitelial-mesenquimal (112). PODXL2, que desempeña un papel como ligando para selectinas vasculares mediando la rodadura de los leucocitos sobre las superficies vasculares, fue otro gene cuya sobreexpresión se asoció con un estadio patológico T avanzado y con la escala de Gleason. En concordancia con lo hallado para PODXL2, Sizemore et al. destacan que la sobreexpresión de podocalyxina se vincula a la migración e invasión celular en CP y cáncer de mama. Se postula que esta sobreexpresión además activaría las vías MAPK y PI3K y que las podocalyxina formaría un complejo con ezrin, un mediador de metástasis (113).

También se estudiaron los genes subexpresados en CP. La ontología predicha mostró que la transducción de señales y procesos de regulación de la proliferación, adhesión y diferenciación celular estuvieron enriquecidos en el

grupo de genes subexpresados en Cpp. La adhesión celular, que mantiene a las células normales congregadas formando los tejidos, sufre alteraciones en cáncer y conduce a la disociación de uniones célula-célula y célula-matriz extracelular, permitiendo a las células cancerosas viajar a otras localizaciones. En tanto que la diferenciación celular, que permite la especialización de las células normales, está postergada en células cancerosas dando privilegio a la proliferación. El gen ADAM33, codante para una enzima metaloendopeptidasa, resultó el más subexpresado en Cpp y mostró estar correlacionado negativamente al PSA. Manica et al, reportó que los pacientes con cáncer de mama que tienen subexpresión de ADAM33 muestran una SG menor y una supervivencia sin metástasis (114).

También se observó que GABRE, receptor de ácido gamma-aminobutírico, estuvo subexpresados en el Cpp, un estudio previo a identificado un locus que alberga GABRE~miR-452~miR-224 se encuentra subregulado e hipermetilado en CP y asociado a la recurrencia bioquímica después de una prostatectomía radical (115). También se hallaron pseudogenes sub-expresados, tales como RP11-394O4.5, MEG3 y ACTA2-AS1, siendo todos ARN no codificantes de cadena larga. El gen HIF3A fue el único gen subexpresados que se asoció a cuatro variables clínicas incluyendo un PSA alto, puntajes de Gleason altos y estadios patológicos T y N mayores. Contrariamente a lo encontrado en este estudio, se ha informado que la sobreexpresión de HIF3A promueve la proliferación celular en cáncer de colorrectal activando la vía de señalización JAK-STAT3 (116), ruta involucrada en procesos como la división celular, la muerte celular y la formación de tumores e inmunidad. Así también la sobreexpresión de HIF3A promueve la invasión y

metástasis en cáncer pancreático mediante la regulación de la vía de señalización HIF3 $\alpha$  / RhoC-ROCK (117). La subexpresión del gen GSTM2 estuvo asociado un estadio T mayor así como a una PSA y puntaje de Gleason altos. Se ha descrito que GSTM2 se encuentra hipermetilado en los estadios tempranos del CP (118), así mismo un estudio ha reportado que la hipermetilación de este gen se ha correlacionado a la disminución de su transcripción, además de estar vinculado a la proliferación celular y asociado a la SG en pacientes con cáncer colorrectal (119).

También encontramos que la subexpresión de CFD y CYP3A5 se asociaron a los estadios patológicos T y N y a la escala de Gleason. CFD es un proteasa que cataliza la escisión del factor B, un paso limitante de la vía alternativa de activación del complemento, que forma parte de la respuesta inmunitaria. Loveridge et al. señala que la sobreexpresión de BRF1, un factor limitante para la transcripción mediada por la ARN polimerasa III, promueve la tumorigénesis en CP asociada a la activación de la vía del complemento, afectando los cambios en la secreción de CDF en ensayos in vitro y en los niveles reducidos de CDF y C7 in vivo (120). CYP3A5 es una proteína que pertenece a la superfamilia de enzimas del citocromo P450. Las proteínas de dicha familia estas asociadas al metabolismo de fármacos y regulan la síntesis de colesterol, esteroides. Contrariamente a la asociación encontrada en nuestro trabajo, Mitra et al. refiere que la expresión reducida de CYP3A5 reduciría la localización nuclear del AR y subsecuentemente induciría la inhibición del crecimiento celular en células de la próstata (121). También se identificaron dos golginas, GOLGA8A y GOLGA8B,

cuyas subexpresiones se asociaron a algunos fenotipos clínicos. Las golginas integran una familia de proteínas que se ubican en el aparato de Golgi. La subexpresión de GOLGA8A estuvo asociada a categorías altas de los estadios patológicos T y N y a valores bajos de PSA. Mientras que la subexpresión de GOLGA8B se asoció a puntajes de Gleason altos y valores bajos de PSA. En contraste a lo encontrado en nuestra evaluación para CP, Wang et al. señalan que la subexpresión de GOLGA8B se asociaría a una mayor SG en pacientes con carcinoma renal (122). La subexpresión LCAT, una enzima que permite la esterificación de colesterol extracelular necesaria para el transporte del mismo, se asoció solamente a valores altos de PSA. Un estudio previo ha reportado que los niveles de reducidos de LCAT junto a ACAT, otra enzima asociada al metabolismo del colesterol, estarían asociados al proceso de la proliferación celular (123).

Finalmente se identificaron a los genes FAM153A y FAM153B, miembros de la familia con similitud de secuencia 153, cuyas subexpresiones se asociaron a valores altos de PSA; adicionalmente la subexpresión de FAM153A se asoció a puntajes altos de Gleason. FAM153B interactúa con PLOD3 catalizando la hidroxilación de residuos de lisina en péptidos de tipo colágeno, FAM153A tendría una función similar. En los genes subexpresados vinculados a variables pronosticas, se destaca que la subexpresión de EGFL8 se asoció a una menor SLE en paciente con CPp. En correspondencia con lo encontrado en nuestro análisis, Wu et al. previamente reportó que la subexpresión de EGFL8 se asoció a con un

peor pronóstico para la SG y SLE en paciente con cáncer colorrectal, así mismo se asoció al potencial metastásico de dicha neoplasia (124).

El interactoma de los 50 genes más sobreexpresados de Cpp mostró una red de contactos en común entre las proteínas AMACR, CRISP3, OR51E1, OR51E2, GOLM1, F5 y HPN (Figura 52). Ello sugiere que su interacción intermolecular es importante en algunos procesos de la carcinogénesis prostática. Dentro de dichos procesos biológicos enriquecidos están la regulación de la respuesta a la hormona tiroidea, el reflejo pilomotor, desmontaje de la membrana y regulación en la proliferación de hepatocitos, todos ellos vinculados al gen HPN, hepsina. Este es uno de los genes más frecuentemente sobreexpresados en CP (125). En cuanto a las funciones biológicas enriquecidas, destacan las actividades catalíticas del tipo racemasa vinculada al gen AMACR cuya sobreexpresión ha sido vinculada a la proliferación e invasión en CP (94). La actividad exopeptidasa así como la actividad de canales de potasio asociadas al gen HPN y receptores olfativos donde se involucran los genes OR51E1 y OR51E2.

Se conoce que la carcinogénesis es un proceso dirigido por genes y mutaciones conductoras o drivers. Típicamente cada neoplasia tiene un grupo de genes y mutaciones conductoras que fomentan el desarrollo del cáncer. En CP observamos que la mutación driver E17K del gen AKT1, vinculado a la regulación de la maquinaria apoptótica, tuvo una frecuencia de aparición en la población de pacientes con Cpm mayor que en Cpp, esto en concordancia con lo reportado por Ritch et al. (126). Este hallazgo apoya el hecho de que ciertas mutaciones

conductoras aparecen durante toda la historia evolutiva del cáncer, no sólo en la etapa de promoción sino también de invasión y metástasis. Contrario a ello, existen otras mutaciones drivers críticas en la metástasis y ausentes en tumores primarios. Así, en CPM se identificaron cuatro mutaciones conductoras incluyendo L702H, T878A y F877L del gen AR, que codifica para el receptor de andrógeno, todas ellas ausentes en CPP. Por otro lado, la mutación driver S137\* del gen KLF6, un activador transcripcional, se presentó en el 0.1% de los individuos con CPP. Se sabe que la pérdida de función de KLF6 dada por mutaciones somáticas y por variantes de empalme otorgan ventaja de supervivencia a las células tumorales en condiciones de castración, lo que causaría la resistencia del tumor a la terapia hormonal (127).

Si bien sólo el 5% de casos de cáncer a nivel mundial son hereditarios, es posible que algunos casos enrolados en los proyectos TCGA sean cáncer hereditario. Según la evidencia colectada en el presente estudio, basada en la frecuencia de SNPs de tipo missense de riesgo a CP, encontramos que ningún sujeto de los estudios para CP albergados en cBioPortal poseía dichos caracteres de predisposición a cáncer. Debido a que sólo se consideraron SNPs de tipo missense, no se puede concluir que otros SNPs (i.e. truncos, marco de lectura, etc.) estuvieran ausentes en la cohorte de estudio.

Para establecer los cambios a nivel de la estructura, estabilidad y función proteica causados por las variaciones de aminoácidos, se investigaron las mutaciones de pérdida de sentido en los 5 primeros GSM en CPP. Teniendo en cuenta la

predicción funcional de Biomuta se identificaron 19 mutaciones predichas dañinas en SPOP, de las cuales 17 tendrían un efecto desestabilizador en términos energéticos, además que afectarían la localización nuclear de SPOP y comprometen principalmente al dominio MATH, que media la interacción con sustratos de proteína-ubiquitina ligasa. Se identificaron las mutaciones Y87N y F133S/C que interesantemente coinciden residuos reportados anteriormente (Y87A y F133A) que inducirían a una fuerte reducción de la afinidad por sustratos (128). Es posible que los cambios en esas posiciones conduzcan a efectos similares y tenga una implicancia importante en la actividad enzimática de la proteína. Las mutaciones de SPOP F102C/S/V, F104C/V/S, F125L, W131R y F133S/C se han descrito previamente como hotspots recurrentes en varios tipos de cáncer (129–131).

En TP53, en primer lugar se hallaron 33 mutaciones probablemente predichas dañinas, de las cuales 22 fueron desestabilizantes para la proteína, pudiendo total alterar la unión al ADN que es también sitio de interacción intermolecular con otras proteínas target. Una única mutación, R248Q, afectó la zona de interacción con el dominio 53BP2 SH3. Una mutación similar pero de cambio en el aminoácido, R258S, afecta la inducción de la degradación de SNAIL que a su vez afectaría la invasión tumoral y metástasis (132).

En FOXA1 se hallaron 26 variantes predichas probablemente dañinas, de las cuales 19 afectaron la estabilidad del producto proteico, la mayoría de ellas afectarían la unión al ADN. Las mutaciones D226N, M253T/K, R261G y

F266L/S han sido identificadas como hotspots recurrentes en varias neoplasias como Cp, cáncer de mama, de vejiga, edometrial y de pulmón (129–131). Para el gen KMT2D, se reportaron 16 mutaciones predichas como probablemente dañinas, de las cuales 7 resultaron ser desestabilizantes, algunas afectarían el dominio dedos de Zinc, importante para la asociación con el ADN (133). Otras mutantes de KMT2D alterarían la función del dominio FYR N-terminal, vinculado a la identificación de modificaciones de histonas (134). La actividad metiltransferasa H3K4 de KMT2D también se vería comprometida con al menos una mutación localizada en el dominio SET (135). Finalmente, la mayoría de mutaciones de KMT2C fueron probablemente dañinas, la mayor parte de ellas desestabilizarían la proteína afectarían la funcionalidad de dominios críticos como dedos de Zinc, FYR N-terminal y SET.

La evidencia presentada en este trabajo de tesis ha sido conseguida a través de la ejecución de métodos de análisis computacional, que involucraron el uso de datos clínicos y genómicos provenientes de individuos enrolados en proyectos de cáncer internacionales. Las herramientas de análisis bioinformáticas permitieron describir el panorama de las alteraciones genéticas en CP, así como identificar algunas variantes asociadas a variables clínico-patológicas y de supervivencia de la enfermedad, con el objetivo de reconocer potenciales marcadores de pronósticos a futuro en beneficio de los pacientes de CP. En esta tesis se han reconocido variantes asociadas factores de manejo clínico no reportadas antes para CP como mutaciones en el gen ZFP36L2 asociado a una menor SG o la sobreexpresión de HIF3A asociado a pronostico desfavorable. También se han

confirmado hallazgos previos de variantes genéticas que serían potenciales marcadores para esta neoplasia, como la delección del gen MAP3K7 se vincularía a peores pronósticos de la enfermedad o la sobreexpresión de AMACR relacionado a un CP de alto riesgo. El análisis secundario de datos genómicos como el desarrollado en este estudio muestra el potencial conocimiento que se alberga en la minería de datos ómicos no sólo para comprender mejor la biología del CP sino la de otros tipos de neoplasias.

## 8. Conclusiones

De forma integral se puede concluir que el empleo de herramientas bioinformáticas y la minería de datos permitieron reconocer algunas variantes genéticas vinculadas a variables fenotípicas y de supervivencia de pacientes con CP a través de un análisis computacional del perfil molecular de la enfermedad.

De forma particular, se puede mencionar las siguientes conclusiones específicas:

-Se identificaron 35 genes significativamente mutados (GSM) en CPp y 79 genes significativamente mutados en CPM. De los GSM en CPp, ZFP36L2 se asoció a la sobrevida global, ZMYM3 se asoció a la escala de Gleason mientras que TP53 se vinculó al estadio patológico T, N y escala de Gleason. Dentro de los genes frecuentemente mutados (GFM) en CPp, se reconocieron dos asociados a la sobrevida global y seis a la sobrevida libre de enfermedad, mientras que para CPM se identificaron tres GFM asociados a la sobrevida global.

-Se reconocieron 7 genes frecuentemente deletados y 14 genes frecuentemente amplificados comunes a CPp y CPM. En CPp la delección de PTEN estuvo correlacionado al estadio patológico T y a la escala de Gleason, mientras que la delección de los genes GTF2F2, TPT1 y GPALPP1 se asoció solo al estadio patológico T. La amplificación de los genes CASC8, MYC, POU5F1B, CCAT1, PCAT1 y PCAT2 en CPp se asociaron a la SG y a las variables clínicas PSA, escala de Gleason, estadio patológico T,N, mientras que la amplificación de los genes ZBTB10, ZNF704, RN7SL107P y FABP4 se asociaron a la SG y al estadio patológico T.

-Se identificaron 755 genes subexpresados y 227 genes sobreexpresados en Cpp. Se reconocieron 10 GDE con subexpresión y 16 GDE con sobreexpresión que se asociaron al menos a alguna característica clínica de la neoplasia. Particularmente la subexpresión de HIF3A y la sobreexpresión de PCA3, GJB1 y ABCC4 se asociaron a 4 parámetros fenotípicos incluyendo PSA, escala de Gleason, estadio patológico T y N.

-Se reportaron algunas mutaciones predichas dañinas en los principales genes significativamente mutados en Cpp. Por ejemplo, en SPOP se identificaron 17 mutaciones con potencial desestabilizador que afectarían el dominio MATH con función biológica relevante.

-Se establecieron diferencias en la prevalencia de mutaciones conductoras para Cpp y Cpm. Se observó que la mutación driver E17K del gen AKT1 tuvo una frecuencia mayor en pacientes con Cpm con respecto a pacientes con Cpp. Por otro lado, solamente en Cpm se hallaron cuatro mutaciones conductoras siendo L702H, T878A y F877L del gen AR.

-No se evidenció ningún SNP de tipo missense de riesgo a CP en la población de estudio.

## 9. Limitaciones

- La ausencia de información clínica y/o genómica en pacientes con CPm restringió la evaluación de la asociación entre las ANCs y GDE con variables clínico-patológicas para ese estadio de la enfermedad.
- La disponibilidad de la información clínica y/o antecedentes de los pacientes con CP, accesibles desde el recurso cBbioPortal, limitó el análisis de otras variables relevantes como la raza o etnia del individuo.
- La disponibilidad de la data de genes significativamente con alteraciones en el número de copias y el detalle de las regiones afectadas desde el recurso Firebrowse evidencio una limitación para el análisis de esa variante genética. Mientras que la disponibilidad de la data de atributos clínicos desde el portal de XENA limitó el análisis de correlación con la expresión génica solo a aquellos casos con información completa.
- La evaluación de deleciones y amplificaciones comunes a CPp y CPm se vio limitada por la falta de data precisa de las zonas cromosómicas involucradas de los genes evaluados a través de cBioportal .

## **10. Recomendaciones**

-En un estudio posterior, se recomienda analizar la información genómica y clínica públicamente disponibles de otras cohortes a fin de evaluar las alteraciones genéticas reportadas en la presente tesis y comparar los patrones de asociación descritos. Además se debe considerar proyectos que cuenten con información consistente de los individuos, teniendo en cuenta variables como la raza, estatus de ETS, estadio metastásico del tumor, etc.

-Para un entendimiento más integral, es recomendable incorporar a futuros estudios del genoma del CP otros componentes como la fusión de genes, los mecanismos epigenéticos y los miRNAs que podrían tener un rol importante en la carcinogénesis y metástasis de la enfermedad.

-En un análisis posterior, es recomendable incluir otros tipos de SNPs, diferentes al de missense, al evaluar la frecuencia de polimorfismos de riesgo a CP en diferentes estadios de la enfermedad.

-En trabajos futuros, sería recomendable ahondar en la comprensión del rol biológico de las variantes genéticas identificadas como posibles biomarcadores de pronóstico. Con el fin de tener una mejor visión del potencial práctico que podrían tener en el manejo clínico de la enfermedad.

## 11. Referencias Bibliográficas

1. Globocan 2012: Estimated Cancer Incidence, Mortality and Prevalence Worldwide in 2012. 2012.
2. American Cancer Society. ¿Qué es el cáncer de próstata? [Internet]. [citado 9 de marzo de 2018]. Disponible en: <https://www.cancer.org/es/cancer/cancer-de-prostata/acerca/que-es-cancer-de-prostata.html>
3. Grizzle WE, Kittles RA, Rais-Bahrami S, Shah E, Adams GW, DeGuenther MS, et al. Self-Identified African Americans and prostate cancer risk: West African genetic ancestry is associated with prostate cancer diagnosis and with higher Gleason sum on biopsy. *Cancer Med.* 2019;
4. Walz J, Chun FK-H, Klein EA, Reuther A, Saad F, Graefen M, et al. Nomogram Predicting the Probability of Early Recurrence After Radical Prostatectomy for Prostate Cancer. *J Urol.* 2009;181(2):601-8.
5. Wolters T, Van Der Kwast TH, Vissers CJ, Bangma CH, Roobol M, Schröder FH, et al. False-negative prostate needle biopsies: Frequency, histopathologic features, and follow-up. *Am J Surg Pathol.* 2010;34(1):35-43.
6. Dias A, Kote-Jarai Z, Mikropoulos C, Eeles R. Prostate cancer germline variations and implications for screening and treatment. *Cold Spring Harb Perspect Med.* 2018;8(9):a030379.
7. Pritchard CC, Mateo J, Walsh MF, De Sarkar N, Abida W, Beltran H, et al. Inherited DNA-repair gene mutations in men with metastatic prostate cancer. *N Engl J Med.* 2016;375(5):443-53.
8. Kontos CK, Adamopoulos PG, Scorilas A. Prognostic and predictive biomarkers in prostate cancer. *Expert Rev Mol Diagn.* 2015;15(12):1567-76.
9. Cheng PF, Dummer R, Levesque MP. Data mining The Cancer Genome Atlas in the era of precision cancer medicine. *Swiss Med Wkly.* 2015;145(September):1-5.

10. Tomczak K, Czerwińska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): An immeasurable source of knowledge. *Wspolczesna Onkol.* 2015;1A:A68-77.
11. Hammerman PS, Voet D, Lawrence MS, Voet D, Jing R, Cibulskis K, et al. Comprehensive genomic characterization of squamous cell lung cancers. *Nature.* 2012;489(7417):519-25.
12. Shih IM, Nakayama K, Wu G, Nakayama N, Zhang J, Wang TL. Amplification of the ch19p13.2 NACC1 locus in ovarian high-grade serous carcinoma. *Mod Pathol.* 2011;24(5):638-45.
13. Haas GP, Delongchamps N, Brawley OW, Wang CY, de la Roza G. The worldwide epidemiology of prostate cancer: perspectives from autopsy studies. *Can J Urol.* 2008;15(1):3866-71.
14. Shoag J, Barbieri C. Clinical variability and molecular heterogeneity in prostate cancer. *Asian J Androl.* 2016;18(4):543.
15. Siegel R. Cáncer Statistics. *Ca Cáncer J.* 2017;67(1):7-30.
16. Tryggvadóttir L, Vidarsdóttir L, Thorgeirsson T, Jonasson JG, Ólafsdóttir EJ, Ólafsdóttir GH, et al. Prostate cancer progression and survival in BRCA2 mutation carriers. *J Natl Cancer Inst.* 2007;99(12):929-35.
17. Ball MW, Gorin MA, Drake CG, Hammers HJ, Allaf ME. The Landscape of Whole-genome Alterations and Pathologic Features in Genitourinary Malignancies: An Analysis of the Cancer Genome Atlas. *Eur Urol Focus.* 2017;1-6.
18. Dominguez-Valentin M, Joost P, Therkildsen C, Jonsson M, Rambech E, Nilbert M. Frequent mismatch-repair defects link prostate cancer to Lynch syndrome. *BMC Urol.* 2016;16(1):15.
19. Kohaar I, Petrovics G, Srivastava S. Molecular Sciences A Rich Array of Prostate Cancer Molecular Biomarkers: Opportunities and Challenges.
20. Laitinen VH, Akinrinade O, Rantapero T, Tammela TLJ, Wahlfors T, Schleutker J. Germline copy number variation analysis in Finnish families with hereditary prostate

cancer. *Prostate*. 2016;76(3):316-24.

21. Jinga V, Csiki IE, Manolescu A, Iordache P, Mates IN, Radavoi D, et al. Replication study of 34 common SNPs associated with prostate cancer in the Romanian population. *J Cell Mol Med*. 2016;20(4):594-600.

22. Chevli KK, Duff M, Walter P, Yu C, Capuder B, Elshafei A, et al. Urinary PCA3 as a predictor of prostate cancer in a cohort of 3,073 men undergoing initial prostate biopsy. *J Urol*. junio de 2014;191(6):1743-8.

23. Cáncer de próstata: Factores de riesgo y prevención | Cancer.Net [Internet]. Disponible en: <https://www.cancer.net/es/tipos-de-cáncer/cáncer-de-próstata/factores-de-riesgo-y-prevención>

24. Tratamiento inicial del cáncer de próstata según la etapa [Internet]. [citado 17 de octubre de 2019]. Disponible en: <https://www.cancer.org/es/cancer/cancer-de-prostata/tratamiento/por-etapa.html>

25. Nelson, D., and Cox M. *Lehninger principles of biochemistry*. Biochem Mol Biol Educ. 2005;33.

26. Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. *Nat Rev Genet*. 2006;7(2):85-97.

27. Fredman D, White SJ, Potter S, Eichler EE, Den Dunnen JT, Brookes AJ. Complex SNP-related sequence variation in segmental genome duplications. *Nat Genet*. 2004;36(8):861-6.

28. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. *Nature*. 2006;444(7118):444-54.

29. Zack TI, Schumacher SE, Carter SL, Cherniack AD, Saksena G, Tabak B, et al. Pan-cancer patterns of somatic copy number alteration. *Nat Genet*. 2013;45(10):1134-40.

30. Heitzer E, Ulz P, Geigl JB, Speicher MR. Non-invasive detection of genome-wide somatic copy number alterations by liquid biopsies. *Mol Oncol*. 2016;10(3):494-502.

31. Hieronymus H, Murali R, Tin A, Yadav K, Abida W, Moller H, et al. Tumor copy

number alteration burden is a pan-cancer prognostic factor associated with recurrence and death. *Elife*. 2018;7.

32. Williams JL, Greer PA, Squire JA. Recurrent copy number alterations in prostate cancer: An in silico meta-analysis of publicly available genomic data. *Cancer Genet*. 2014;207(10-12):474-88.

33. Visakorpi T, Kallioniemi AH, Syvänen AC, Hyytinen ER, Karhu R, Tammela T, et al. Genetic changes in primary and recurrent prostate cancer by comparative genomic hybridization. *Cancer Res*. 1995;55(2):342-7.

34. Abeshouse A, Ahn J, Akbani R, Ally A, Amin S, Andry CD, et al. The Molecular Taxonomy of Primary Prostate Cancer. *Cell*. 2015;163(4):1011-25.

35. Abeshouse A, Ahn J, Akbani R, Ally A, Amin S, Andry CD, et al. The Molecular Taxonomy of Primary Prostate Cancer. *Cell*. noviembre de 2015;163(4):1011-25.

36. Fan S, Liang Z, Gao Z, Pan Z, Han S, Liu X, et al. Identification of the key genes and pathways in prostate cancer. *Oncol Lett*. 2018;16(5):6663-9.

37. Rakul SA, Kamilova TA, Golota AS, Shcherbak SG. Prognostic and predictive biomarkers of prostate cancer. *Onkourologiya*. 2017;13(4):111-21.

38. Esfahani M, Ataei N, Panjehpour M. Biomarkers for evaluation of prostate cancer prognosis. Vol. 16, *Asian Pacific Journal of Cancer Prevention*. 2015. p. 2601-11.

39. Durand X, Xylinas E, Radulescu C, Haus-Cheymol R, Moutereau S, Ploussard G, et al. The value of urinary prostate cancer gene 3 (PCA3) scores in predicting pathological features at radical prostatectomy. *BJU Int*. 2012;110(1):43-9.

40. Ayday E, Raisaro JL, Hubaux JP, Rougemont J. Protecting and evaluating genomic privacy in medical tests and personalized medicine. En: *Proceedings of the ACM Conference on Computer and Communications Security*. 2013. p. 95-106.

41. Schirmer MA, Lüske CM, Roppel S, Schaudinn A, Zimmer C, Pflüger R, et al. Relevance of Sp Binding Site Polymorphism in WWOX for Treatment Outcome in Pancreatic Cancer. *J Natl Cancer Inst*. 2016;108(5):djv387.

42. Xu M, Gao Y, Yu T, Wang J, Cheng L, Cheng L, et al. Functional promoter rs2295080 T>G variant in MTOR gene is associated with risk of colorectal cancer in a Chinese population. *Biomed Pharmacother.* 2015;70(C):28-32.
43. Shao N, Li JM, Xu B, Wang Y, Lu XJ, Feng NH. Role of the functional variant (-652T>G) in the XRCC4 promoter in prostate cancer. *Mol Biol Rep.* 2014;41(11):7463-70.
44. Li L, Sun R, Liang Y, Pan X, Li Z, Bai P, et al. Association between polymorphisms in long non-coding RNA PRNCR1 in 8q24 and risk of colorectal cancer. *J Exp Clin Cancer Res.* 2013;32(1):104.
45. Kral M, Rosinska V, Student V, Grepl M, Hrabec M, Bouchal J. Genetic determinants of prostate cancer: a review. *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub.* 2011;155(1):3-9.
46. The Cancer Genome Atlas - TCGA [Internet]. [citado 8 de marzo de 2018]. Disponible en: <https://cancergenome.nih.gov/>
47. Zhang H, Shen T, Zhang Z, Li Y, Pan Z. Expression of KIF18A Is Associated with Increased Tumor Stage and Cell Proliferation in Prostate Cancer. *Med Sci Monit.* 2019;25:6418-28.
48. Stratton MR, Campbell PJ, Futreal PA. The cancer genome.
49. Moráňová L, Bartošík M. Long Non-Coding RNAs - Current Methods of Detection and Clinical Applications. *Klin Onkol.* 2019;32(Supplementum 3):65-71.
50. Tamborero D, Rubio-Perez C, Deu-Pons J, Schroeder MP, Vivancos A, Rovira A, et al. Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. *Genome Med* [Internet]. 2018;10(1):25. Disponible en: <https://www.cancergenomeinterpreter.org>
51. Fehring G, Kraft P, Pharoah PD, Eeles RA, Chatterjee N, Schumacher FR, et al. Cross-Cancer Genome-Wide Analysis of Lung, Ovary, Breast, Prostate, and Colorectal Cancer Reveals Novel Pleiotropic Associations. *Cancer Res.* 2016;76(17):5103-14.

52. Lange EM, Johnson AM, Wang Y, Zuhlke KA, Lu Y, Ribado J V., et al. Genome-Wide Association Scan for Variants Associated with Early-Onset Prostate Cancer. Ewart D, editor. PLoS One [Internet]. 2014;9(4):e93436. Disponible en: <http://dx.plos.org/10.1371/journal.pone.0093436>
53. Kote-Jarai Z, Olama AA Al, Giles GG, Severi G, Schleutker J, Weischer M, et al. Seven prostate cancer susceptibility loci identified by a multi-stage genome-wide association study. *Nat Genet.* 2011;43(8):785-91.
54. Schumacher FR, Al Olama AA, Berndt SI, Benlloch S, Ahmed M, Saunders EJ, et al. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nat Genet.* 2018;50(7):928-36.
55. Cheng PF, Dummer R, Levesque MP. Data mining The Cancer Genome Atlas in the era of precision cancer medicine. Vol. 145, *Swiss Medical Weekly.* 2015.
56. Lee HJ, Palm J, Grimes SM, Ji HP. The Cancer Genome Atlas Clinical Explorer: A web and mobile interface for identifying clinical-genomic driver associations. *Genome Med.* 2015;7(1):1-14.
57. Ball MW, Gorin MA, Drake CG, Hammers HJ, Allaf ME. The Landscape of Whole-genome Alterations and Pathologic Features in Genitourinary Malignancies: An Analysis of the Cancer Genome Atlas. *Eur Urol Focus.* 2017;3(6):584-9.
58. He Z, Duan X, Zeng G. Identification of potential biomarkers and pivotal biological pathways for prostate cancer using bioinformatics analysis methods. *PeerJ.* 2019;2019(10).
59. Yoshie H, Sedukhina AS, Minagawa K, Oda K, Ohnuma S, Yanagisawa N, et al. A bioinformatics-to-clinic sequential approach to analysis of prostate cancer biomarkers using TCGA datasets and clinical samples: A new method for precision oncology? *Oncotarget.* 2017;8(59):99601-11.
60. Armenia J, Wankowicz SAM, Liu D, Gao J, Kundra R, Reznik E, et al. The long tail of oncogenic drivers in prostate cancer. *Nat Genet.* 2018;50(5):645-51.

61. Krishna S, Fan Y, Jarosek S, Adejoro O, Chamie K, Konety B. Racial Disparities in Active Surveillance for Prostate Cancer. *J Urol*. 2017;197(2):342-9.
62. Andrén O, Fall K, Franzén L, Andersson SO, Johansson JE, Rubin MA. How well does the gleason score predict prostate cancer death? A 20-year followup of a population based cohort in Sweden. *J Urol*. 2006;175(4):1337-40.
63. Barbieri CE, Baca SC, Lawrence MS, Demichelis F, Blattner M, Theurillat JP, et al. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nat Genet*. junio de 2012;44(6):685-9.
64. Baca SC, Prandi D, Lawrence MS, Mosquera JM, Romanel A, Drier Y, et al. Punctuated evolution of prostate cancer genomes. *Cell*. 2013;153(3):666-77.
65. Kumar A, Coleman I, Morrissey C, Zhang X, True LD, Gulati R, et al. Substantial interindividual and limited intraindividual genomic diversity among tumors from men with metastatic prostate cancer. *Nat Med*. 2016;22(4):369-78.
66. Lv S, Ji L, Chen B, Liu S, Lei C, Liu X, et al. Histone methyltransferase KMT2D sustains prostate carcinogenesis and metastasis via epigenetically activating LIFR and KLF4. *Oncogene*. 2018;37(10):1354-68.
67. Guo C, Chang CC, Wortham M, Chen LH, Kernagis DN, Qin X, et al. Global identification of MLL2-targeted loci reveals MLL2's role in diverse signaling pathways. *Proc Natl Acad Sci U S A*. 2012;109(43):17603-8.
68. Robinson D, Van Allen EM, Wu YM, Schultz N, Lonigro RJ, Mosquera JM, et al. Integrative clinical genomics of advanced prostate cancer. *Cell*. 2015;161(5):1215-28.
69. Yonemori K, Seki N, Kurahara H, Osako Y, Idichi T, Arai T, et al. ZFP36L2 promotes cancer cell aggressiveness and is regulated by antitumor microRNA-375 in pancreatic ductal adenocarcinoma. *Cancer Sci*. 2017;108(1):124-35.
70. Han Y, Zheng Q, Tian Y, Ji Z, Ye H. Identification of a nine-gene panel as a prognostic indicator for recurrence with muscle-invasive bladder cancer. *J Surg Oncol*. 15 de junio de 2019;119(8):1145-54.

71. Tuna M, Ju Z, Yoshihara K, Amos CI, Tanyi JL, Mills GB. Clinical relevance of TP53 hotspot mutations in high-grade serous ovarian cancers. *Br J Cancer*. 2019;
72. Leung JWC, Makharashvili N, Agarwal P, Chiu LY, Pourpre R, Cammarata MB, et al. ZMYM3 regulates BRCA1 localization at damaged chromatin to promote DNA repair. *Genes Dev*. 2017;31(3):260-74.
73. Tang MR, Wang YX, Guo S, Han SY, Wang D. CSMD1 exhibits antitumor activity in A375 melanoma cells through activation of the Smad pathway. *Apoptosis*. 2012;17(9):927-37.
74. Deng J, Chen H, Zhou D, Zhang J, Chen Y, Liu Q, et al. Comparative genomic analysis of esophageal squamous cell carcinoma between Asian and Caucasian patient populations. *Nat Commun*. 2017;8(1).
75. Khan FH, Pandian V, Ramraj S, Natarajan M, Aravindan S, Herman TS, et al. Acquired genetic alterations in tumor cells dictate the development of high-risk neuroblastoma and clinical outcomes. *BMC Cancer*. 2015;15(1).
76. Zhang R, Song C. Loss of CSMD1 or 2 may contribute to the poor prognosis of colorectal cancer patients. *Tumor Biol*. 2014;35(5):4419-23.
77. Qian J, Zhao S, Zou Y, Rahman SMJ, Senosain M-F, Stricker T, et al. Genomic Underpinnings of Tumor Behavior in in situ and Early Lung Adenocarcinoma. *Am J Respir Crit Care Med*. 2019;rccm.201902-0294OC.
78. Walker CJ, Miranda MA, O'Hern MJ, McElroy JP, Coombes KR, Bundschuh R, et al. Patterns of CTCF and ZFH3 mutation and associated outcomes in endometrial cancer. *J Natl Cancer Inst*. 2015;107(11).
79. Gandhi J, Afridi A, Vatsia S, Joshi G, Joshi G, Kaplan SA, et al. The molecular biology of prostate cancer: Current understanding and clinical implications. Vol. 21, *Prostate Cancer and Prostatic Diseases*. Nature Publishing Group; 2018. p. 22-36.
80. Schulz WA. Molecular biology of prostate cancer. *Mol Hum Reprod*. 2003;9(8):437-48.

81. Liu W, Chang BL, Cramer S, Koty PP, Li T, Sun J, et al. Deletion of a small consensus region at 6q15, including the MAP3K7 gene, is significantly associated with high-grade prostate cancers. *Clin Cancer Res.* 2007;13(17):5028-33.
82. Kluth M, Hesse J, Heidl A, Krohn A, Steurer S, Sirma H, et al. Genomic deletion of MAP3K7 at 6q12-22 is associated with early PSA recurrence in prostate cancer and absence of TMPRSS2:ERG fusions. *Mod Pathol.* 2013;26(7):975-83.
83. Rodrigues LU, Rider L, Nieto C, Romero L, Karimpour-Fard A, Loda M, et al. Coordinate loss of MAP3K7 and CHD1 promotes aggressive prostate cancer. *Cancer Res.* 2015;75(6):1021-34.
84. Chen W, Wang H, Tao S, Zheng Y, Wu W, Lian F, et al. Tumor protein translationally controlled 1 is a p53 target gene that promotes cell survival. *Cell Cycle.* 2013;12(14):2321-8.
85. Sharma A, Mendonca J, Ying J, Kim HS, Verdone JE, Zarif JC, et al. The prostate metastasis suppressor gene NDRG1 differentially regulates cell motility and invasion. *Mol Oncol.* 2017;11(6):655-69.
86. Dixon KM, Lui GYL, Kovacevic Z, Zhang D, Yao M, Chen Z, et al. Dp44mT targets the AKT, TGF- $\beta$  and ERK pathways via the metastasis suppressor NDRG1 in normal prostate epithelial cells and prostate cancer cells. *Br J Cancer.* 2013;108(2):409-19.
87. Attard G, Parker C, Eeles RA, Schröder F, Tomlins SA, Tannock I, et al. Prostate cancer. Vol. 387, *The Lancet.* Lancet Publishing Group; 2016. p. 70-82.
88. Liu H, Zhang Q, Lou Q, Zhang X, Cui Y, Wang P, et al. Differential Analysis of lncRNA, miRNA and mRNA Expression Profiles and the Prognostic Value of lncRNA in Esophageal Cancer. *Pathol Oncol Res.* 2019;
89. Dong H, Hu J, Wang L, Qi M, Lu N, Tan X, et al. SOX4 is activated by C-MYC in prostate cancer. *Med Oncol.* 2019;36(11).
90. Nesbit CE, Tersak JM, Prochownik E V. MYC oncogenes and human neoplastic disease. Vol. 18, *Oncogene.* 1999. p. 3004-16.

91. Fromont G, Godet J, Peyret A, Irani J, Celhay O, Rozet F, et al. 8q24 amplification is associated with Myc expression and prostate cancer progression and is an independent predictor of recurrence after radical prostatectomy. *Hum Pathol.* 2013;44(8):1617-23.
92. Gurel B, Iwata T, Koh CM, Jenkins RB, Lan F, Van Dang C, et al. Nuclear MYC protein overexpression is an early alteration in human prostate carcinogenesis. *Mod Pathol.* 2008;21(9):1156-67.
93. Yang G, Timme TL, Frolov A, Wheeler TM, Thompson TC. Combined c-Myc and caveolin-1 expression in human prostate carcinoma predicts prostate carcinoma progression. *Cancer.* 2005;103(6):1186-94.
94. Antonarakis ES, Keizman D, Zhang Z, Gurel B, Lotan TL, Hicks JL, et al. An immunohistochemical signature comprising PTEN, MYC, and Ki67 predicts progression in prostate cancer patients receiving adjuvant docetaxel after prostatectomy. *Cancer.* 2012;118(24):6063-71.
95. Pan Y, Zhan L, Chen L, Zhang H, Sun C, Xing C. POU5F1B promotes hepatocellular carcinoma proliferation by activating AKT. *Biomed Pharmacother.* 2018;100:374-80.
96. Zanke BW, Greenwood CMT, Rangrej J, Kustra R, Tenesa A, Farrington SM, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat Genet.* 2007;39(8):989-94.
97. Rafnar T, Sulem P, Thorleifsson G, Vermeulen SH, Helgason H, Saemundsdottir J, et al. Genome-wide association study yields variants at 20p12.2 that associate with urinary bladder cancer. *Hum Mol Genet.* octubre de 2014;23(20):5545-57.
98. Speedy HE, Di Bernardo MC, Sava GP, Dyer MJS, Holroyd A, Wang Y, et al. A genome-wide association study identifies multiple susceptibility loci for chronic lymphocytic leukemia. *Nat Genet.* 2014;46(1):56-60.
99. Hayashi H, Arao T, Togashi Y, Kato H, Fujita Y, De Velasco MA, et al. The OCT4 pseudogene POU5F1B is amplified and promotes an aggressive phenotype in gastric cancer. *Oncogene.* 2015;34(2):199-208.

100. You Z, Liu C, Wang C, Ling Z, Wang Y, Wang Y, et al. LncRNA CCAT1 promotes prostate cancer cell proliferation by interacting with DDX5 and miR-28-5p. *Mol Cancer Ther.* 2019;
101. Chen H, He Y, Hou YS, Chen DQ, He SL, Cao YF, et al. Long non-coding RNA CCAT1 promotes the migration and invasion of prostate cancer PC-3 cells. *Eur Rev Med Pharmacol Sci.* 2018;22(10):2991-6.
102. Emília A, Lemos G, Da A, Matos R, Ferreira LB, Rodrigues E, et al. The long non-coding RNA PCA3: an update of its functions and clinical applications as a biomarker in prostate cancer.
103. Ferro M, Lucarelli G, Bruzzese D, Perdonà S, Mazzarella C, Perruolo G, et al. Improving the prediction of pathologic outcomes in patients undergoing radical prostatectomy: The value of prostate cancer antigen 3 (PCA3), prostate health index (Phi) and sarcosine. *Anticancer Res.* 2015;35(2):1017-23.
104. Roobol MJ, Schröder FH, Van Leeuwen P, Wolters T, Van Den Bergh RCN, Van Leenders GJLH, et al. Performance of the prostate cancer antigen 3 (PCA3) gene and prostate-specific antigen in prescreened men: Exploring the value of PCA3 for a first-line diagnostic test. *Eur Urol.* 2010;58(4):475-81.
105. Czyż J. The stage-specific function of gap junctions during tumourigenesis. *Cell Mol Biol Lett.* 2008;13(1):92-102.
106. Teleki I, Szasz AM, Maros ME, Gyorffy B, Kulka J, Meggyeshazi N, et al. Correlations of differentially expressed gap junction connexins cx26, cx30, cx32, cx43 and cx46 with breast cancer progression and prognosis. *PLoS One.* 2014;9(11).
107. Luo Y, Jiang Q-W, Wu J-Y, Qiu J-G, Zhang W-J, Mei X-L, et al. Regulation of migration and invasion by Toll-like receptor-9 signaling network in prostate cancer. Vol. 6.
108. Oprea-Lager DE, Bijnsdorp I V., Van Moorselaar RJA, Van Den Eertwegh AJM, Hoekstra OS, Geldof AA. ABCC4 Decreases docetaxel and not cabazitaxel efficacy in

- prostate cancer cells in vitro. *Anticancer Res.* 2013;33(2):387-92.
109. Souza MF De, Kuasne H, Barros-Filho MDC, Cilião HL, Marchi FA, Fuganti PE, et al. Circulating mRNA signature as a marker for high-risk prostate cancer. *Carcinogenesis.* 2019;
110. Zha S, Ferdinandusse S, Denis S, Wanders RJ, Ewing CM, Luo J, et al.  $\alpha$ -Methylacyl-CoA Racemase as an Androgen-Independent Growth Modifier in Prostate Cancer. *Cancer Res.* 2003;63(21):7365-76.
111. Georgescu C, Corbin JM, Thibivilliers S, Webb ZD, Zhao YD, Koster J, et al. A TMEFF2-regulated cell cycle derived gene signature is prognostic of recurrence risk in prostate cancer. *BMC Cancer.* 2019;19(1).
112. Gao L, Nie X, Zheng M, Li X, Guo Q, Liu J, et al. TMEFF2 is a novel prognosis signature and target for endometrial carcinoma. *Life Sci.* 2019;116910.
113. Sizemore S, Cicek M, Sizemore N, Kwok PN, Casey G. Podocalyxin increases the aggressive phenotype of breast and prostate cancer cells in vitro through its interaction with ezrin. *Cancer Res.* 2007;67(13):6183-91.
114. Manica GCM, Ribeiro CF, Oliveira MASD, Pereira IT, Chequin A, Ramos EAS, et al. Down regulation of ADAM33 as a predictive biomarker of aggressive breast cancer. *Sci Rep.* 2017;7.
115. Kristensen H, Haldrup C, Strand S, Mundbjerg K, Mortensen MM, Thorsen K, et al. Hypermethylation of the GABRE $\beta$ miR-452miR-224 promoter in prostate cancer predicts biochemical recurrence after radical prostatectomy. *Clin Cancer Res.* 2014;20(8):2169-81.
116. Xue X, Jungles K, Onder G, Samhoun J, Gyorffy B, Hardiman KM. HIF-3 $\alpha$ 1 promotes colorectal tumor cell growth by activation of JAK-STAT3 signaling. *Oncotarget.* 2016;7(10):11567-79.
117. Zhou X, Guo X, Chen M, Xie C, Jiang J. HIF-3 $\alpha$  promotes metastatic phenotypes in pancreatic cancer by transcriptional regulation of the RhoC-ROCK1 signaling pathway.

Mol Cancer Res. 2018;16(1):124-34.

118. Ashour N, Angulo JC, Andrés G, Alelú R, González-Corpas A, Toledo M V., et al. A DNA hypermethylation profile reveals new potential biomarkers for prostate cancer diagnosis and prognosis. *Prostate*. 2014;74(12):1171-82.

119. Sun X, Chen D, Jin Z, Chen T, Lin A, Jin H, et al. Genome-wide methylation and expression profiling identify methylation-associated genes in colorectal cancer. *Epigenomics*. 2019;epi-2019-0133.

120. Loveridge CJ, Slater S, Campbell KJ, Nam NA, Knight J, Ahmad I, et al. BRF1 accelerates prostate tumorigenesis and perturbs immune infiltration. *Oncogene*. 2019;

121. Mitra R, Goodman OB. CYP3A5 regulates prostate cancer cell growth by facilitating nuclear translocation of AR. *Prostate*. 2015;75(5):527-38.

122. Wang J, Zhang C, He W, Gou X. Construction and comprehensive analysis of dysregulated long non-coding RNA-associated competing endogenous RNA network in clear cell renal cell carcinoma. *J Cell Biochem*. octubre de 2018;

123. Rao KN, Kottapally S, Eskander ED, Shinozuka H, Dessi S, Pani P. Acinar cell carcinoma of rat pancreas: Regulation of cholesterol esterification. *Br J Cancer*. 1986;54(2):305-10.

124. Wu F, Shirahata A, Sakuraba K, Kitamura Y, Goto T, Saito M, et al. Down-regulation of EGFL8: a novel prognostic biomarker for patients with colorectal cancer. *Anticancer Res*. 2011;31(6):2249-54.

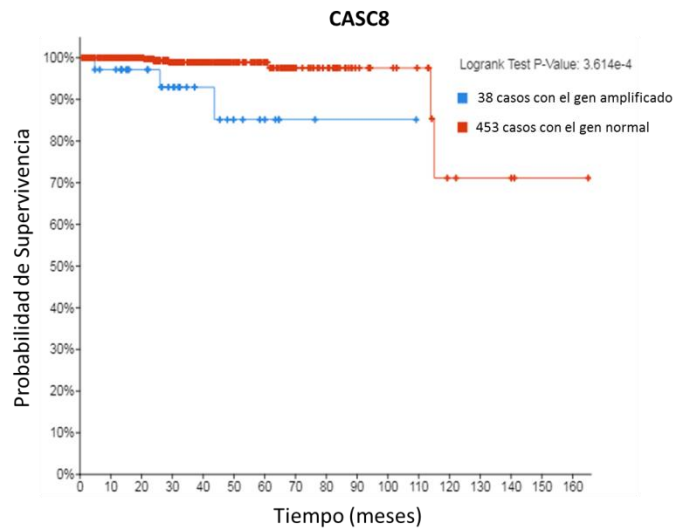
125. Stephan C, Yousef GM, Scorilas A, Jung K, Jung M, Kristiansen G, et al. Hepsin is highly over expressed in and a new candidate for a prognostic indicator in prostate cancer. *J Urol*. 2004;171(1):187-91.

126. Ritch E, Fu SYF, Herberts C, Wang G, Warner EW, Schönlau E, et al. Identification of hypermutation and defective mismatch repair in ctDNA from metastatic prostate cancer. *Clin Cancer Res*. 2019;

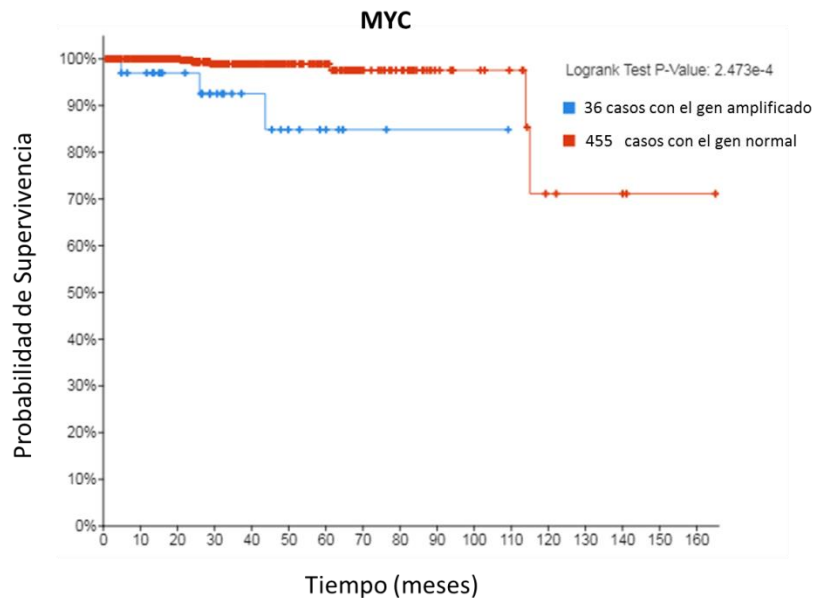
127. Liu X, Gomez-Pinillos A, Loder C, Carrillo-De Santa Pau E, Qiao R, Unger PD,

- et al. KLF6 loss of function in human prostate cancer progression is implicated in resistance to androgen deprivation. *Am J Pathol.* 2012;181(3):1007-16.
128. Zhuang M, Calabrese MF, Liu J, Waddell MB, Nourse A, Hammel M, et al. Structures of SPOP-Substrate Complexes: Insights into Molecular Architectures of BTB-Cul3 Ubiquitin Ligases. *Mol Cell.* 2009;36(1):39-50.
129. Gao J, Chang MT, Johnsen HC, Gao SP, Sylvester BE, Sumer SO, et al. 3D clusters of somatic mutations in cancer reveal numerous rare mutations as functional targets. *Genome Med.* diciembre de 2017;9(1):4.
130. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio Cancer Genomics Portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* 2012;2(5):401-4.
131. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal.* 2013;6(269).
132. Lim SO, Kim H, Jung G. P53 inhibits tumor cell invasion via the degradation of snail protein in hepatocellular carcinoma. *FEBS Lett.* 2010;584(11):2231-6.
133. Ali M, Hom RA, Blakeslee W, Ikenouye L, Kutateladze TG. Diverse functions of PHD fingers of the MLL/KMT2 subfamily. Vol. 1843, *Biochimica et Biophysica Acta - Molecular Cell Research.* 2014. p. 366-71.
134. Rao RC, Dou Y. Hijacked in cancer: The KMT2 (MLL) family of methyltransferases. Vol. 15, *Nature Reviews Cancer.* Nature Publishing Group; 2015. p. 334-46.
135. Froimchuk E, Jang Y, Ge K. Histone H3 lysine 4 methyltransferase KMT2D. Vol. 627, *Gene.* Elsevier B.V.; 2017. p. 337-42.

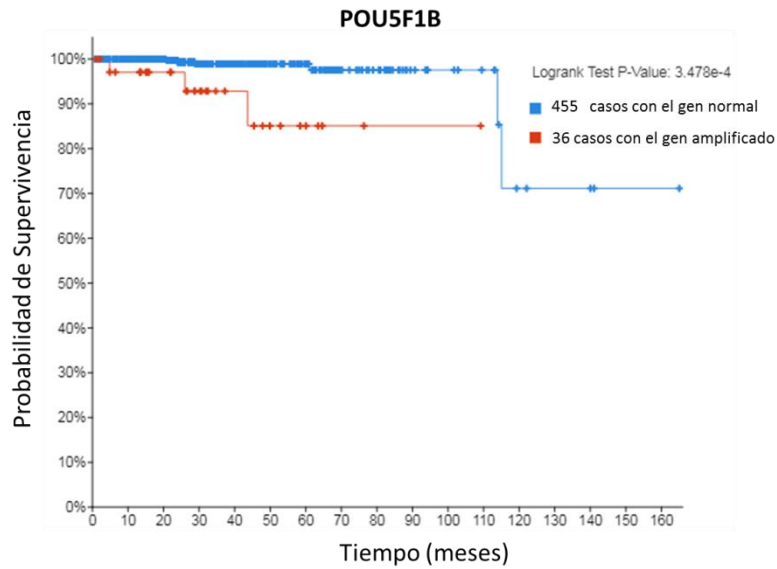
## 12. Anexos



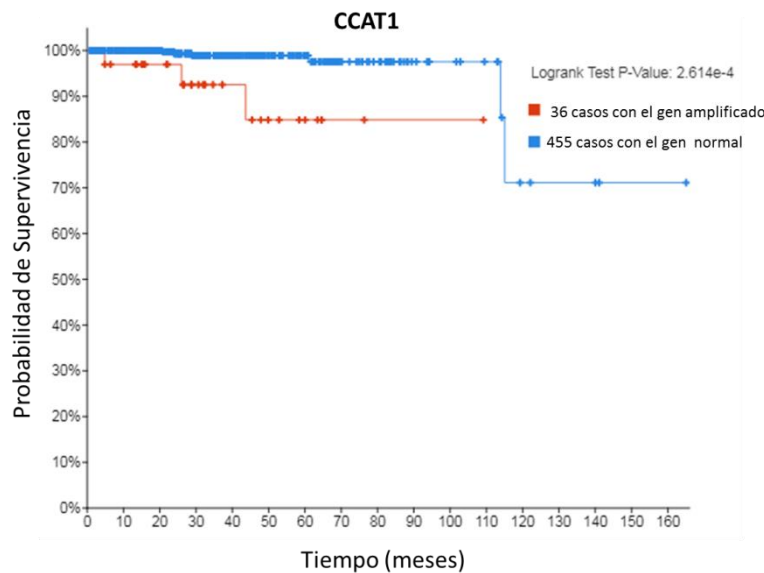
**Anexo 1. Grafica de KAPLAN-MEIER mostrando la sobrevida global de pacientes con y sin amplificaciones en el gen CASC8. Estudios de cBioPortal para CPp (N= 491).**



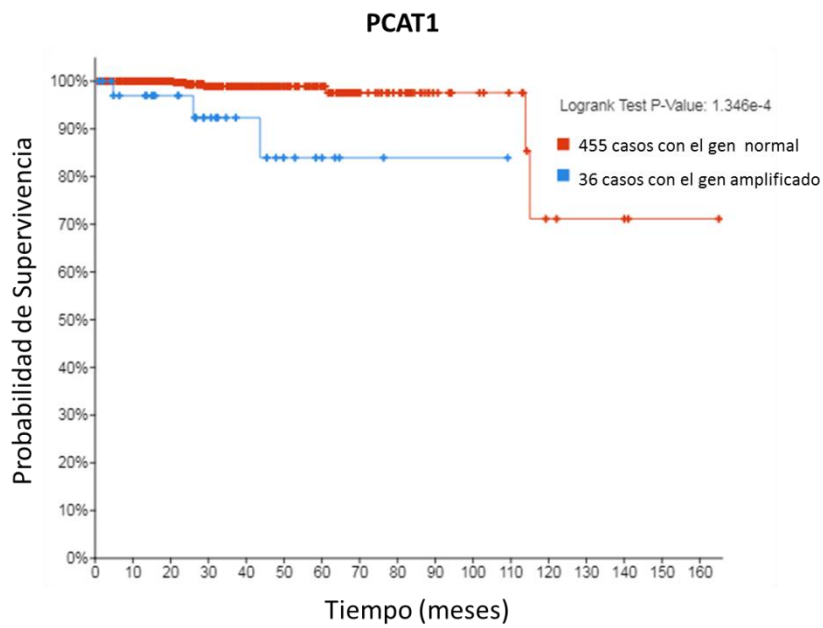
**Anexo 2. Grafica de KAPLAN-MEIER mostrando la sobrevida global de pacientes con y sin amplificaciones en el gen MYC8. Estudios de cBioPortal para CPp (N= 491).**



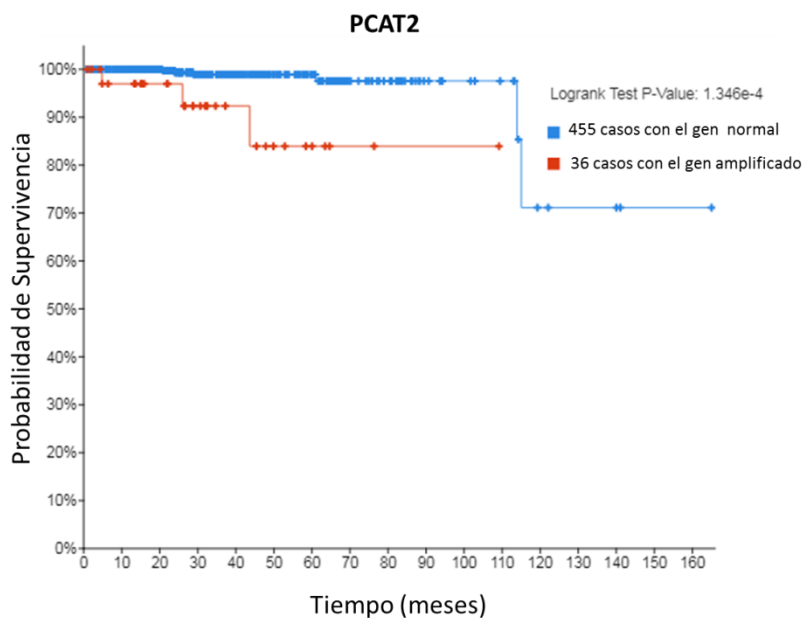
**Anexo 3. Grafica de KAPLAN-MEIER mostrando la sobrevida global de pacientes con y sin amplificaciones en el gen POU5F1B. Estudios de cBioPortal para CPp (N= 491).**



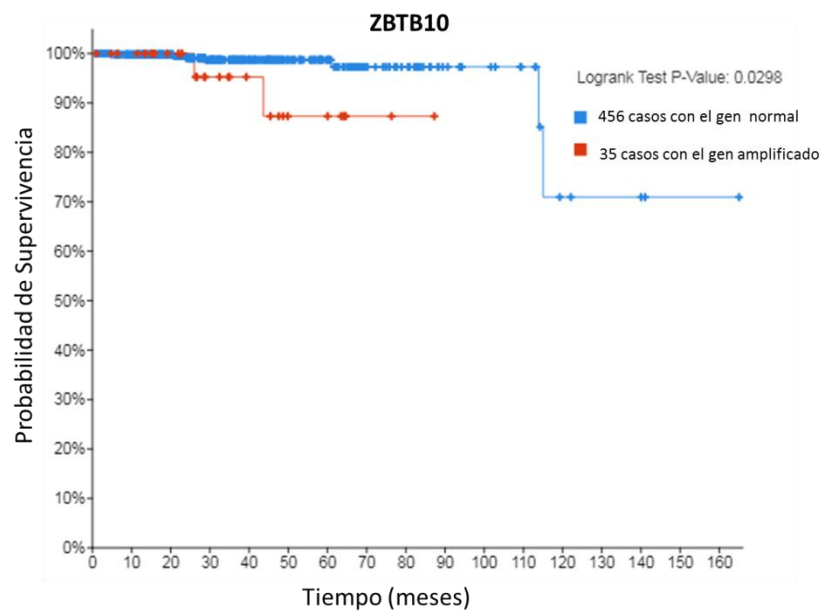
**Anexo 4. Grafica de KAPLAN-MEIER mostrando la sobrevida global de pacientes con y sin amplificaciones en el gen CCAT1. Estudios de cBioPortal para CPp (N= 491).**



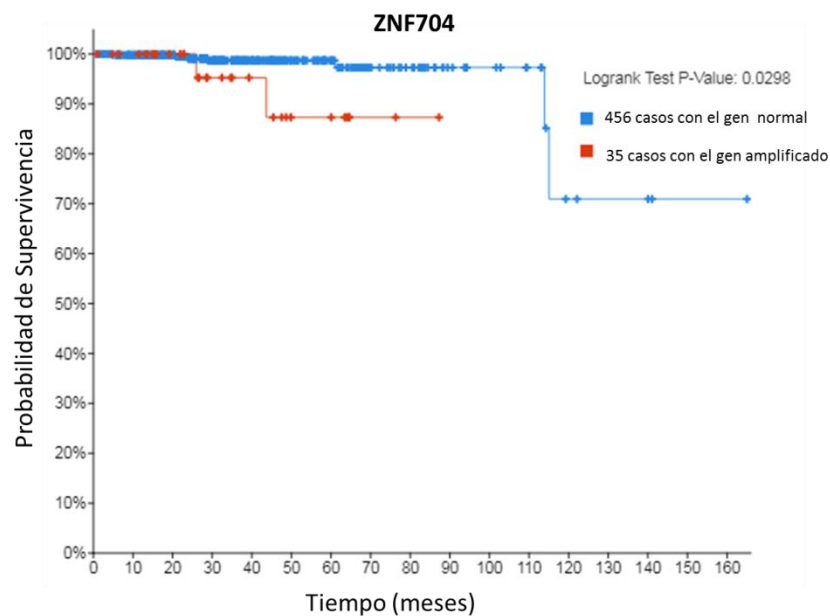
**Anexo 5. Grafica de KAPLAN-MEIER mostrando la sobrevida global de pacientes con y sin amplificaciones en el gen PCAT1. Estudios de cBioPortal para CPp (N= 491).**



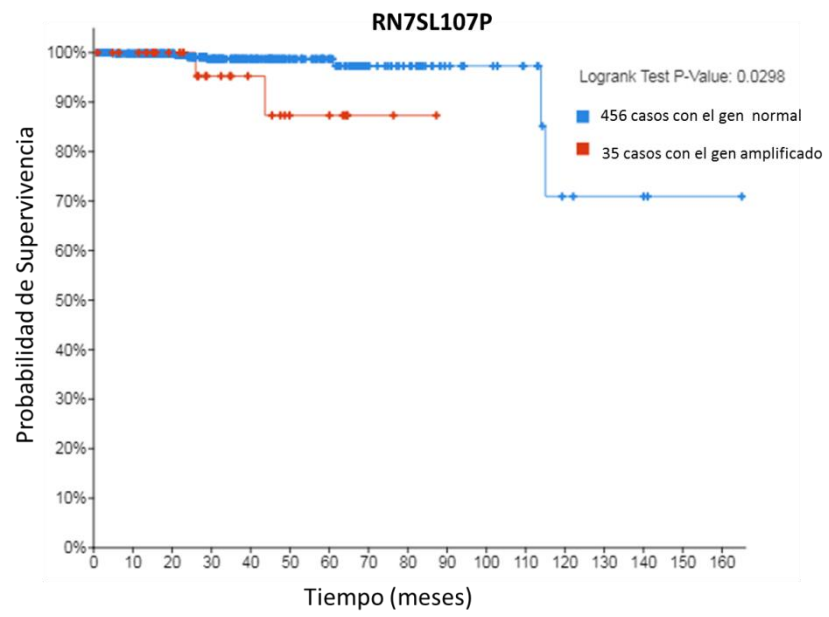
**Anexo 6. Grafica de KAPLAN-MEIER mostrando la sobrevida global de pacientes con y sin amplificaciones en el gen PCAT2. Estudios de cBioPortal para CPp (N= 491).**



**Anexo 7. Grafica de KAPLAN-MEIER mostrando la sobrevida global de pacientes con y sin amplificaciones en el gen ZBTB10. Estudios de cBioPortal para CPp (N= 491).**



**Anexo 8. Grafica de KAPLAN-MEIER mostrando la sobrevida global de pacientes con y sin amplificaciones en el gen ZNF704. Estudios de cBioPortal para CPp (N= 491).**



**Anexo 9. Grafica de KAPLAN-MEIER mostrando la sobrevida global de pacientes con y sin amplificaciones en el gen RN7SL107P. Estudios de cBioPortal para CPp (N= 491).**