



**Identificación de marcadores para cáncer de próstata por medio de un análisis de expresión
diferencial**

Paredes Escobar, Michelle Marcela

Facultad de Ingeniería

Maestría en Biología Computacional

Dr. Avelar Rivas, Jesús Abraham

Quito, 2022

Dedicatoria

Para mi madre, Victoria Escobar Barberan, por ser esa fuente de apoyo incondicional durante todas las etapas de formación personal y profesional. La vida nos separó, pero ambas sabemos que fue por mi bienestar y nuestro futuro.

Para aquella Michelle de 3 años que soñaba con tener su propia PC. La vida nos cumplió el sueño y amamos trabajar en computadoras. Para aquella Michelle de 23 años que jamás se imaginó que una pandemia le otorgaría una de sus pasiones más grandes: la bioinformática y biología computacional.

A todas las mujeres madres, amigas, mentoras y científicas que estuvieron presente en mi vida, todas ellas fueron y son inspiración de la profesional que quiero llegar a ser. Espero provocar la misma inspiración y dar el mismo apoyo a futuras investigadoras en biología computacional y bioinformática

Michelle Paredes Escobar

“El universo tiene un comienzo, pero no un final. Infinito. Las estrellas también tienen sus comienzos, pero se encuentran con un final de su propia creación. Finito.”

Agradecimientos

A mi madre María Victoria que me brindó impulso para seguir este programa de maestría. Sus mensajes, llamadas y palabras de apoyo me motivaron día con día para cumplir la meta. Gracias por escucharme y darme aliento en esta dura etapa de mi vida. La vida nos volverá a unir.

Al Dr. Abraham Avelar que siempre estuvo dispuesto a compartir sus conocimientos y consejos durante el transcurso de este trabajo. Sus clases fueron espectaculares, es un referente como docente e investigador.

A la Pontificia Universidad Católica del Ecuador – PUCE por ofertar un programa de maestría necesario e innovador.

A mis hermanos. Antes los quería lejos, ahora solo los quiero cerca. Porque cada mensaje suyo alegra mis días. Ustedes son mi motivación para viajar y volver a compartir risas con ustedes. Gracias por tantos memes.

A mis mascotas Timaius †, Byakko y Suza. El amor que da un perrito es el más puro que las personas pueden recibir, y más aún cuando son adoptados. No tiene precio ver sus caritas de agradecimiento. Insisto, yo nos los salvé de un futuro incierto, ustedes me salvaron a mí con tanto amor.

A mi tía Ramona por ser esa segunda madre que me cuida como si fuera su hija. A mi Madrina Linda Cevallos por tantas anécdotas contadas y ser ese oído que escucha y luego aconseja, siempre habrá momento para un cafecito con pan.

A mi mejor amigo, David Vaca. La vida nos unió de una forma random y fue la mejor casualidad que me ha pasado. Sin tu apoyo este último año, este objetivo no se hubiera materializado. La vida nos depara más aventuras. Sé que el destino nos pone pruebas difíciles y que ahora no entendemos, pero con el tiempo cobran sentido.

A mis amigos Doménica, Brandon y Camila. En algún momento fui yo quien los apoyó con sus trabajos, hoy fueron ustedes que se presentaron como luz al final del túnel. Gracias por adoptarme hace muchos años atrás, gracias por siempre estar.

A mis compañeros de Maestría. Pese a ser un grupo pequeño, nos mostramos unidos y siempre dispuestos a ayudarnos cuando algún tema se volvía complicado. Pese a encontrarnos separados por una pantalla, la virtualidad no fue un limitante para formar un hermoso vínculo.

Michelle Paredes Escobar

Índice de contenido

Dedicatoria.....	2
Agradecimientos	3
Índice de contenido	5
Índice de tablas	7
Índice de figuras.....	8
Abreviaturas.....	10
Resumen	12
Abstract.....	13
Capítulo I: Introducción	14
Antecedentes	14
Justificación.....	16
Objetivos del Proyecto.....	18
Objetivo General.....	18
Objetivos específicos	18
Capítulo 2: Marco Teórico.....	19
Anatomía y fisiología de la próstata	19
Cáncer de próstata (CaP)	21
Genética del cáncer de próstata.....	22
Métodos de diagnóstico	24
Prueba de antígeno prostático específico (PSA) en sangre	24
Tomografía por Emisión de Positrones/Tomografía Computarizada (PET/CT)	25
Tacto rectal	25
Tecnologías de expresión diferencial de genes (DGE).....	25
Análisis de expresión diferencial en cáncer de próstata	27
Capítulo 3: Metodología	29
Esquema de desarrollo	29
Set de datos	29
Preprocesamiento.....	31
Análisis de expresión diferencial.....	32
Análisis de enriquecimiento funcional.....	33
Capítulo 4: Resultados	34

Análisis de calidad.....	34
Mapeo y cuantificación de transcritos.....	37
Análisis exploratorio de los datos	38
Análisis de expresión diferencial.....	41
Análisis de enriquecimiento funcional.....	44
Capítulo 5: Discusión.....	49
Capítulo 6: Conclusiones y Recomendaciones.....	54
Referencias.....	56

Índice de tablas

Tabla 1

Lecturas mapeadas con Kallisto 37

Tabla 2

Análisis de enriquecimiento para genes sobreexpresados 45

Tabla 3

Análisis de enriquecimiento para genes subexpresados GO función molecular 46

Tabla 4

Análisis de enriquecimiento para genes subexpresados GO componente celular 46

Índice de figuras

Figura 1	
<i>Tasas de incidencia y mortalidad estandarizadas por edad estimadas en 2020 en cáncer de próstata</i>	17
Figura 2	
<i>Ubicación de la glándula prostática.....</i>	19
Figura 3	
<i>Anatomía de la glándula prostática.....</i>	20
Figura 4	
<i>Esquema de progresión del cáncer de próstata</i>	22
Figura 5	
<i>Factores de riesgo asociados al cáncer de próstata</i>	23
Figura 6	
<i>Flujo de trabajo para el análisis de expresión diferencial</i>	30
Figura 7	
<i>Matriz de recuentos no normalizados.....</i>	32
Figura 8	
<i>Estadística básica del análisis de secuencias en FastQC</i>	34
Figura 9	
<i>Análisis de calidad de secuencias en FastQC</i>	35
Figura 10	
<i>Contenido GC por base del análisis de FastQC.....</i>	36
Figura 11	
<i>Contenido GC por base del análisis de FastQC.....</i>	36
Figura 12	
<i>Análisis de exploratorio de genes expresados diferencialmente del dataset-1</i>	39
Figura 13	
<i>Análisis de componentes principales (PCA) del dataset-2.....</i>	40
Figura 14	
<i>Gráfico MA para el dataset2.....</i>	41
Figura 15	
<i>Gráfico de Volcán para el dataset2.....</i>	43

Figura 16

Mapa de calor de valores z para el dataset2 de los 20 primeros genes expresados diferencialmente 44

Figura 17

Principales GO de función moleculares para genes sobreexpresados diferencialmente en carcinoma GO de función molecular..... 45

Figura 18

Principales GO de función moleculares para genes subexpresados diferencialmente en carcinoma GO de función molecular 47

Figura 19

Principales GO de función moleculares para genes subexpresados diferencialmente en carcinoma GO de componentes celulares 48

Abreviaturas

CaP	Cáncer de próstata
PSA	Antígeno prostático específico
AR	Receptor de andrógenos
DRE	Examen rectal digital
IRM	Imágenes por resonancia magnética
PIN	Intraepitelial prostática
SNPs	Polimorfismos de nucleótido único
SVN	Variantes de un solo nucleótido
SCNA	Alteraciones del número de copias somáticas
PSMA	Antígeno prostático específico de membrana
DGE	Expresión diferencial de genes
RNA-seq	Secuenciación de ARN
GEO	Gene Expression Omnibus
GAI	Illumina genome analyzer II
ENA	European Nucleotide Archive
%GC	Contenido de guaninas y citocinas

CSV	Archivo separado por comas
PCA	Análisis de componentes principales
GO	Ontología génica

Resumen

El cáncer de próstata es el segundo cáncer con mayor incidencia en el mundo en hombres entre los 45 y 60 años. Debido a la alta tasa de falsos positivos que arroja la prueba de antígeno prostático, que conlleva a procedimientos invasivos y tratamientos innecesarios en displasias benignas, se vuelve imperante determinar biomarcadores de detección temprana y no invasivos. Gracias a los avances en la secuenciación de ARN el estudio del transcriptoma toma un papel relevante en la identificación de genes expresados diferencialmente con potencial para actuar como biomarcadores pronósticos o predictivos. En este trabajo se utilizaron datos de RNA-seq obtenidos de tejidos sanos y cancerosos publicados en bases de datos públicos con la finalidad de realizar un análisis de expresión diferencial y posteriormente realizar un análisis de enriquecimiento funcional para identificar las funciones metabólicas donde interfieren los genes diferencialmente expresados. Como resultado se obtuvo que las secuencias disponibles en el proyecto código de acceso GSE22260 de GEO mostraron tener poca o nula expresión diferencial, por lo que se descartó este set de datos del análisis y se optó por trabajar con las secuencias disponibles en el proyecto PRJEB2449 de la plataforma ENA. El análisis de expresión diferencial de genes de este conjunto de datos determinó que un total de 105 genes se expresaron diferencialmente. Además, el análisis de enriquecimiento funcional identificó que los genes sobre expresados se involucran en la síntesis de tubulina, mientras que los sub expresados en los procesos de adhesión celular. Los genes identificados en este trabajo pueden ser considerados en ensayos de validación de biomarcadores y generar pruebas de identificación de tejidos cancerosos.

Palabras clave:

- **CÁNCER DE PRÓSTATA**
- **EXPRESIÓN DIFERENCIAL**
- **BIOCONDUCTOR**
- **BIOMARCADORES**

Abstract

Prostate cancer is the second most common cancer in the world among men aged 45 to 60. Due to the high rate of false positives generated by the prostate-specific antigen test, which leads to invasive procedures and unnecessary treatments in benign dysplasias, it becomes imperative to determine early and non-invasive biomarkers for detection. Thanks to advances in RNA sequencing, the study of the transcriptome plays a significant role in identifying differentially expressed genes with the potential to act as prognostic or predictive biomarkers. In this study, RNA-seq data obtained from healthy and cancerous tissues published in public databases were used to perform a differential expression analysis, followed by a functional enrichment analysis to identify metabolic functions in which the differentially expressed genes interfere. As a result, it was found that the sequences available in the GSE22260 accession code project from GEO showed little or no differential expression, so this dataset was excluded from the analysis, and the sequences available in the PRJEB2449 project from the ENA platform were chosen for further analysis. The analysis of differential gene expression in this dataset determined that a total of 105 genes were differentially expressed. Furthermore, the functional enrichment analysis identified that the overexpressed genes are involved in tubulin synthesis, while the underexpressed genes are involved in cellular adhesion processes. The genes identified in this study can be considered for biomarker validation assays and the development of tests for identifying cancerous tissues.

Keywords:

- **PROSTATE CANCER**
- **DIFFERENTIAL EXPRESSION**
- **BIOCONDUCTOR**
- **BIOMARKERS**

Capítulo I: Introducción

Antecedentes

Reconocida como la enfermedad del hombre adulto mayor, el cáncer de próstata (CaP) es el segundo cáncer más común en hombres entre 45 y 60 años. A nivel mundial, en el año 2020 se estimó 1.414.000 nuevos casos y 375.304 muertes convirtiendo a esta patología en la quinta causa principal de muerte por cáncer en hombres, superado por cáncer de pulmón, hígado, estómago, esófago y páncreas (Sekhoacha et al., 2022; Sung et al., 2021; Waldron et al., 2023).

El CaP se caracteriza por proliferación descontrolada de células de la glándula prostática dando como resultado un crecimiento anormal de la glándula. La transformación maligna comienza como una neoplasia intraepitelial, continua a un cáncer de próstata localizado, seguido de un adenocarcinoma prostático avanzado con invasión local y finalmente un cáncer de próstata metastásico llegando a ganglios linfáticos pélvicos y retroperitoneales, médula espinal, vejiga, recto, hueso y cerebro. La mayoría de las muertes se produce por la metástasis originada y no por la invasión localizada (Schatten, 2018; Wang et al., 2018).

Uno de los principales factores de riesgos son los antecedentes hereditarios. Se estima que los hombres con antecedentes familiares de cáncer de próstata tienen un riesgo del 50% de desarrollo de esta patología. Además, estudios en gemelos afirman que es uno de cánceres más hereditarios. Las anomalías genéticas involucradas incluyen los reordenamientos cromosómicos, mutaciones puntuales en los genes de reparación y alteraciones en el número de copias de células somáticas (Sekhoacha et al., 2022; Wang et al., 2018).

Es una enfermedad heterogénea ya que dentro de un mismo paciente se puede encontrar múltiples focos tumorales con diferencias genéticas, patológicas y funcionales. Esta característica complica la

comprensión clínica de la enfermedad obstaculizando la estimación del grado de diseminación metastásica y la selección terapéutica. Considerar la heterogeneidad tumoral al momento de realizar el estudio del cáncer es un factor importante para determinar la respuesta farmacológica y recaída de la enfermedad (Wang et al., 2018; Wasim et al., 2022).

Durante las primeras etapas de la enfermedad, los pacientes no experimentan síntomas lo que conduce a un diagnóstico tardío disminuyendo la probabilidad de supervivencia al ser detectado en etapas metastásicas. El principal método de cribado es la prueba de antígeno prostático específico (PSA), una glicoproteína dependiente transcripcionalmente del receptor de andrógenos (AR), principal impulsor oncogénico del cáncer de próstata. El PSA es producida exclusivamente por las células prostáticas malignas o benignas por lo que aumentos de sus niveles es indicador de esta enfermedad. No obstante, este método de tamizaje continúa siendo discutido debido a la alta tasa de falsos positivo y por lo tanto tratamientos innecesarios. Otros métodos diagnóstico incluyen: examen rectal digital (DRE), biopsia guiada por ultrasonido, imágenes por resonancia magnética (IRM) e histopatología (Sekhoacha et al., 2022; Tian et al., 2018). Sin embargo, hasta el momento, no existe una prueba que permita el diagnóstico diferenciado entre cáncer de próstata y displasias benignas (Wasim et al., 2022).

Tras este objetivo, la búsqueda de nuevos biomarcadores que mejoren la toma de decisiones diagnósticas y terapéuticas avanza constantemente. Un biomarcador se define como cualquier sustancia, estructura o proceso que puede medirse y evaluarse de forma objetiva y que actúa como indicador de procesos biológicos normales y patológicos, o respuesta farmacológica. De acuerdo con su función, los biomarcadores pueden clasificarse en predictivo o pronóstico (Strimbu & Tavel, 2010).

Los biomarcadores predictivos permiten predecir el efecto de determinado tratamiento, por lo que para demostrar su característica predictiva es necesario contar con un grupo expuesto al tratamiento y un grupo control de tal forma que se pueda determinar la significancia estadística entre la interacción del

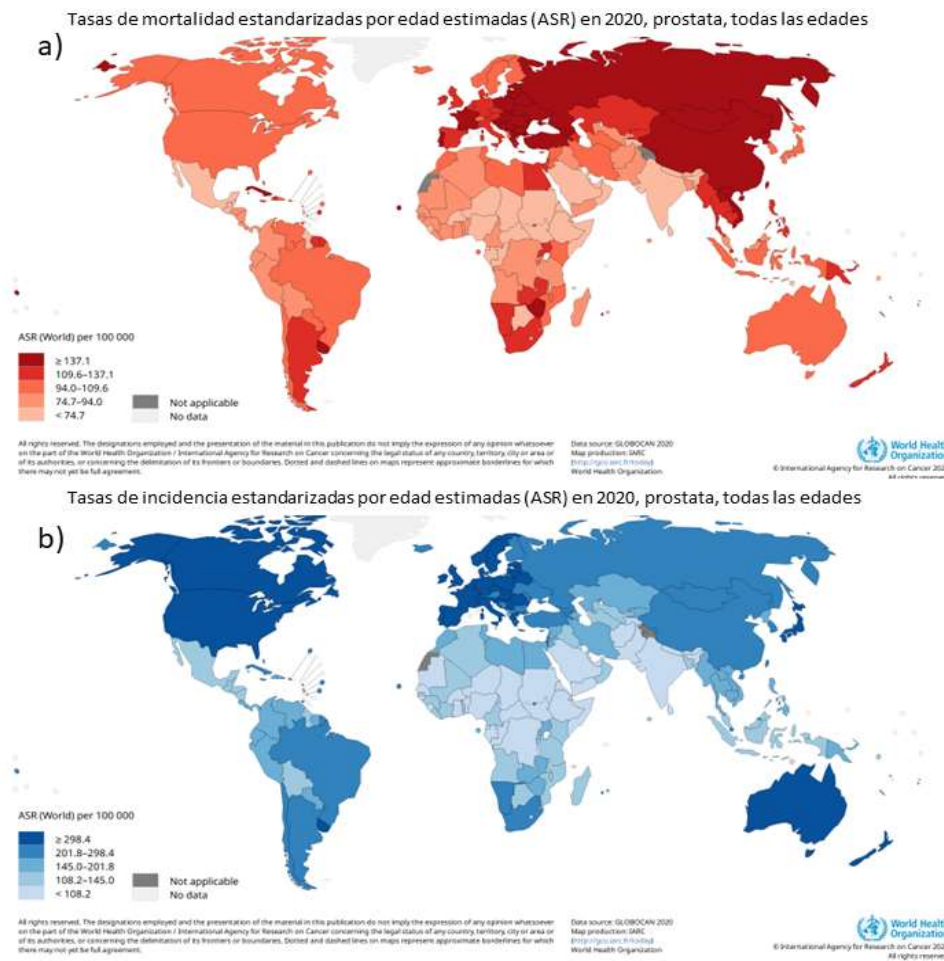
tratamiento y el biomarcador. Por el contrario, los biomarcadores pronósticos otorgan información sobre la recurrencia y progresión de la enfermedad sin intervención de un tratamiento. Por lo tanto, un buen marcador pronóstico debe predecir el resultado sin presentar efecto del tratamiento. No obstante, pese a las diferencias entre ambos tipos, también existen biomarcadores que cumplen un rol predictivo y diagnóstico (Ballman, 2015; Frommlet et al., 2022).

Justificación

A nivel mundial, se reportó 1.414.000 nuevos casos de cáncer de próstata y 375.304 muertes en el 2020. Se estima que para el 2040 los nuevos casos incrementarán a 28,4 millones y las muertes aumentarán en un 47%. Estos valores pueden variar de acuerdo con la globalización y el crecimiento de la economía de tal forma que la mortalidad varía del 64% al 95% en países en transición (Figura 1a). El cáncer de próstata es el más diagnosticado en hombres en más de la mitad de los países del mundo cuya tasa de incidencia varía de acuerdo con la región (Figura 1b) (Sung et al., 2021).

En América Latina y el Caribe la tendencia se mantiene de tal forma que el 14% de los cánceres detectados corresponden a cáncer de próstata. No obstante, la incidencia de la etapa metastásica es mayor en esta región versus otras partes del mundo, principalmente, por falta de pruebas de cribado, por lo que el cáncer de próstata continúa siendo la primera causa de muerte en hombres (Reis et al., 2020).

Figura 1

Tasas de incidencia y mortalidad estandarizadas por edad estimadas en 2020 en cáncer de próstata

Nota: (a) Incidencia de cáncer de próstata a nivel mundial, 2020. (b) Mortalidad por cáncer de próstata a nivel mundial, 2020. Fuente: Ferlay et al., (2020)

En el Ecuador, el Registro Nacional de Tumores, SOLCA - Núcleo de Quito reporta que el CaP es el tipo de tumor más común en hombres con una incidencia de 35,7 casos por 100.00 hombres y una tasa de mortalidad 10,6 – 14,3 casos por 100.000 hombres (Caballero, 2021). Estos datos se contrastan con la estadística de SOLCA- Núcleo Guayaquil quienes reportan en el boletín epidemiológico del año 2019 una incidencia de 38,8 casos por cada 100.000 hombres (Real et al., 2020).

Debido a la falta de acceso a pruebas de tamizaje, tanto en la región como en Ecuador, las tasas de incidencia y mortalidad aumentan cada año. Además, la alta tasa de falso positivo arrojados por las pruebas de PSA provocan pruebas confirmatorias invasivas innecesarias. Por lo que, la investigación de nuevos marcadores son necesarios para el correcto diagnóstico y apropiada decisión terapéutica (Kretschmer & Tilki, 2017; Shi et al., 2022). Con el avance de tecnologías de secuenciación de ARN (ARN-seq) y de las herramientas de análisis de expresión génica, el estudio del transcriptoma de cáncer de próstata permite el descubrimiento de biomarcadores para un seguimiento adecuado de la enfermedad de tal forma que se reduzca o evita múltiples procedimientos invasivos para la confirmación del padecimiento así como la adecuada selección terapéutica (Pinskaya et al., 2019).

Objetivos del Proyecto

Objetivo General

Realizar un análisis de expresión diferencial en cáncer de próstata para selección de identificación de marcadores predictivos y pronósticos.

Objetivos específicos

- Obtener set de datos de secuencias de ARN de células prostáticas normales y cancerígenas disponibles en bases de datos
- Analizar la calidad de las secuencias de ARN
- Realizar el análisis de expresión diferencial
- Identificar genes que se expresan diferencialmente entre células normales y cancerígenas

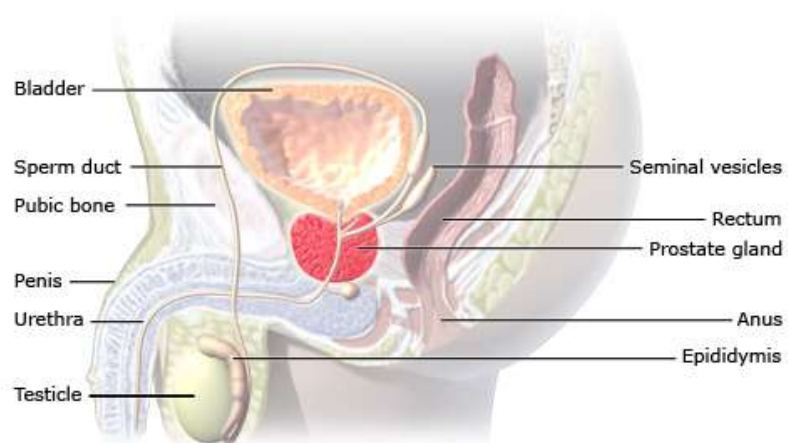
Capítulo 2: Marco Teórico

Anatomía y fisiología de la próstata

La próstata es un órgano que pertenece al sistema reproductor masculino, ubicada debajo de la vejiga y sobre los músculos del piso pélvico (figura 2). Es una glándula accesoria cuyo tamaño se asemeja al de una nuez y pesa 30 gramos aproximadamente. Su función es completar las secreciones del semen y mantener viables los espermatozoides. Además, los músculos prostáticos también se aseguran de que el semen sea expulsado con fuerza durante la eyaculación (Aaron et al., 2016; Institute for Quality and Efficiency in Health Care, 2016).

Figura 2

Ubicación de la glándula prostática



Location of the prostate gland

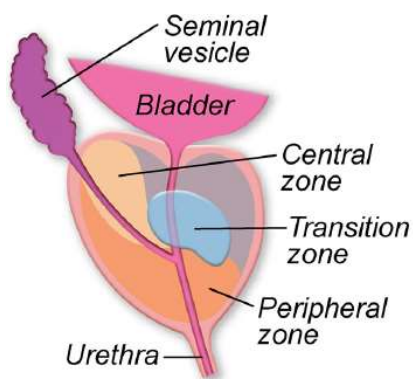
Nota: La próstata es una glándula se encuentra localizada debajo de la vejiga y sobre los músculos pélvicos. Fuente: Institute for Quality and Efficiency in Health Care (2016)

La glándula prostática se divide en tres principales áreas que son histológica y funcionalmente distintas: la zona de transición, la zona central y la zona periférica (figura 3). En conjunto rodean a la próstata de forma independiente por lo que se dice que están anatómicamente separadas (Wang et al., 2018).

La zona transición se encuentra en el interior de la glándula entre la zona central y periférica. Ocupa aproximadamente el 10% de la glándula siendo la más pequeña y es el responsable de la producción de líquido prostático. La zona central rodea a la zona de transición. Constituye la cuarta parte de la próstata. Es el punto de encuentro de los conductos seminales y prostático por lo que se conoce como conducto eyaculador, además se encarga de la producción de enzimas que contribuyen a la licuefacción del semen. Finalmente, el conducto periférico representa el 70% de la masa total de la glándula y está asociado con la producción del líquido seminal (Aaron et al., 2016; Institute for Quality and Efficiency in Health Care, 2016; Wasim et al., 2022).

Figura 3

Anatomía de la glándula prostática



Nota: La próstata se divide en tres zonas: la zona central, zona de transición y zona periférica. Del 60 al 75% de los cánceres se desarrollan en la zona periférica cuya función principal es la producción de líquido seminal. Fuente: Wang et al. (2018)

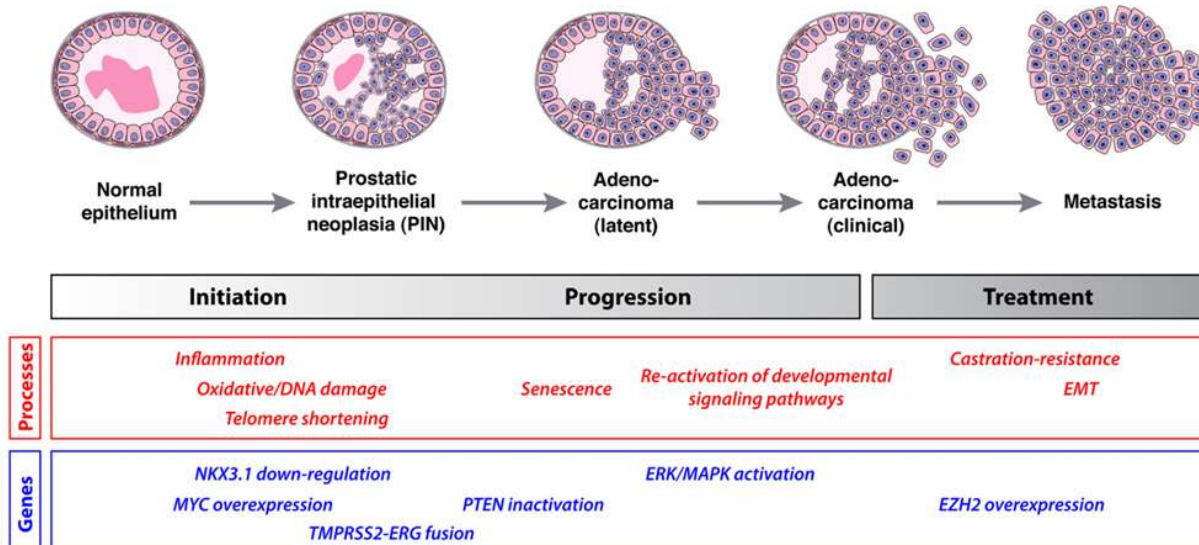
Cáncer de próstata (CaP)

El cáncer de próstata es una enfermedad definida por la proliferación descontrolada y anormal de células malignas de la glándula prostática lo que provoca el aumento de tamaño glandular. Del 60 al 75% de cánceres de próstata ocurren en la zona periférica (Wang et al., 2018).

El desarrollo del cáncer se esquematiza en la figura 4, el cual comienza con una neoplasia intraepitelial prostática (PIN). Esta es una condición precancerosa que se caracteriza por los cambios morfológicos en el tejido afectado. A nivel histológico, se observa la hiperplasia de células luminales, reducción de células basales, agrandamiento de núcleos y nucléolos, hiper cromasia citoplasmática y atipia nuclear (Shen & Abate-Shen, 2010).

La neoplasia avanza hasta la primera fase cancerosa denominado adenocarcinoma localizado en la cual las células malignas se encuentran confinadas dentro de la glándula, por lo que el tumor es pequeño y se encuentra limitado a la cápsula prostática. Posteriormente, se convierte en un carcinoma localmente invasivo de tal forma que el tumor llega a los tejidos cercanos. Finalmente, el cáncer alcanza tejidos lejanos como ganglios linfáticos pélvicos y retroperitoneales, médula espinal, vejiga, recto, hueso y cerebro provocando metástasis (Epstein et al., 2016).

Figura 4

Esquema de progresión del cáncer de próstata

Nota: La ruta de progresión del cáncer de próstata se describe en cuatro etapas. El epitelio norma se transforma en una neoplasia intraepitelial la cual avanza hasta los tejidos circundantes y termina invadiendo órganos lejanos. Dentro de cada una de las etapas, se manifiestan diversos procesos biológicos y varios genes se activan e inactivan. Fuente: Shen & Abate-Shen (2010)

Genética del cáncer de próstata

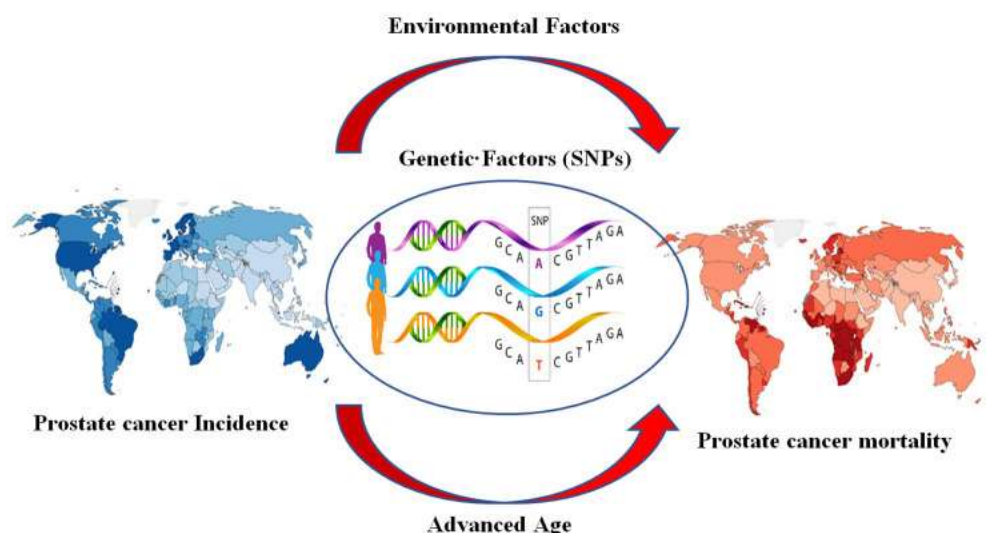
De manera general, las mutaciones genéticas son una de las causas frecuentes de cáncer. En el cáncer prostático los genes involucrados se relacionan a la ruta de los andrógenos y en el metabolismo de la testosterona (Sekhoacha et al., 2022). El linaje familiar es el principal factor de riesgo para el cáncer de próstata, quienes poseen parientes directos diagnosticados con esta patología, tienen un riesgo de 2 a 8 veces mayor de desarrollarlo en comparación con los hombres sin antecedentes familiares similares (Brandão et al., 2020; Ni Raghallaigh & Eeles, 2022). No obstante, un estudio realizado en gemelos monocigóticos y dicigóticos concluyó que el 57% del riesgo de cáncer prostático podría deberse a

factores hereditarios (Mucci et al., 2016). Este fenómeno se genera por una combinación de genes raros, pero de riesgo moderado o alto, así como, variantes genéticas comunes, pero de menor riesgo, exposición ambiental, etnicidad, migración, dietas y edad avanzada (Adhyam & Gupta, 2012; Allemailem et al., 2021; Wang et al., 2018).

Gracias a los avances en secuenciación de ADN, se ha logrado el reconocimiento de las alteraciones que promueven o conducen a una división celular incontrolada, lo que con lleva al cáncer de próstata, entre estas se tienen polimorfismos de nucleótido único (SNPs), variantes de un solo nucleótido (SVN), mutaciones puntuales y alteraciones del número de copias somáticas (SCNA) (figura 5) (Benafif et al., 2021; Sekhoacha et al., 2022). La identificación oportuna de estas mutaciones es un derrotero importante para la comprensión de la patología y para la obtención de estrategias de tratamiento adecuadas en un entorno metastásico (Brandão et al., 2020).

Figura 5

Factores de riesgo asociados al cáncer de próstata



Fuente: Allemailem et al. (2021)

Las mutaciones en la maquinaria de reparación de los errores de emparejamiento del ADN se asocian con un riesgo mayor de desarrollar la enfermedad, al menos el 12% de hombres que padecen cáncer metastásico de próstata poseen una mutación germinal patógena de los genes BRCA1/2, ATM, CHEK2, y NBN y aquellos con alteraciones en el gen BRCA2 poseen mayor probabilidad de adquirir la enfermedad y peores pronósticos (Sokolova et al., 2021; Wang et al., 2018). En el estudio conducido por Richard C. en 2016, se demostró que el 67% de los casos existía inactivación del segundo alelo en los genes de reparación del ADN, lo que da sustento a la importancia de las alteraciones germinales en los mismos (Sokolova et al., 2021). Los genes mencionados previamente junto con los genes HOX, Rase L, MSR1, ANXA7, CDKN1B, PON1, PTEN, NK3, moda, MYC y ELAC2/HPC2 se emplean ampliamente como biomarcadores para el cáncer de próstata (Sekhoacha et al., 2022).

Métodos de diagnóstico

La detección temprana y precisa de la enfermedad es un aspecto importante para aumentar la tasa de supervivencia, sobre todo cuando el cáncer ha avanzado hasta etapa metastásica. Debido a que el cáncer de próstata es una enfermedad heterogénea y la falta de especificidad en las técnicas de imagen tradicionales, no existe un método universalmente aprobado para el diagnóstico (Tian et al., 2018). A continuación, se describen los métodos más utilizados.

Prueba de antígeno prostático específico (PSA) en sangre

La PSA es una serina proteasa codificada en el cromosoma 19 secretada por la glándula prostática, se encuentra en el líquido seminal y en bajas proporciones en la sangre (American Cancer Society, 2019; Fundora et al., 2020). Su producción está relacionada con la presencia de andrógenos y el tamaño que presente la glándula prostática. La principal función del PSA es la desintegración del coágulo seminal a través de la proteólisis de la fibronectina en péptidos que favorecen la motilidad de los espermatozoides. Durante el examen se toman muestras de sangre para evaluar los niveles, siendo

aquellos superiores a 4 ng/ml los que se deben considerar para más análisis debido a la posibilidad de presentar cáncer (Sekhoacha et al., 2022).

Tomografía por Emisión de Positrones/Tomografía Computarizada (PET/CT)

La PET/CT es una técnica que permite la combinación de la información metabólica y morfológica demostrando una alta precisión al momento de realizar el diagnóstico debido a que se obtienen imágenes semicuantitativas. El método se basa en el uso de radiofármacos emisores de positrones junto con la detección de fotones colineales producidos por el choque de un positrón con un electrón. Existen diversos radiotrazadores, dentro de estos se encuentra el 68Ga-PSMA que básicamente utiliza la actividad enzimática del antígeno prostático específico de membrana (PSMA) y usa inhibidores enzimáticos radiomarcados o agentes aglutinantes para el diagnóstico. En el país no se cuenta con el acceso a marcadores específicos como la 68Ga-PSMA limitando el diagnóstico de alta eficiencia (Colmener et al., 2018; Fundora et al., 2020).

Tacto rectal

Otro de los métodos que se emplean para el diagnóstico es el tacto rectal, sirve para detectar alteraciones en el tamaño, consistencia o uniformidad de la próstata. Se basa en la introducción de un dedo en el recto con la finalidad de identificar algún abultamiento o área firme en la próstata que puede derivar en cáncer, es una técnica poco sensible debido a que se detecta entre 55 y 90% de los casos (American Cancer Society, 2019; Robles et al., 2019).

Tecnologías de expresión diferencial de genes (DGE)

Los estudios de expresión diferencial (DGE) tienen como objetivo comprender los mecanismos biológicos detrás del funcionamiento de genes. Dado que, la expresión génica varía entre tejidos de un mismo individuo, son de gran utilidad para estudios de variación entre organismos y dentro de un

mismo individuo. Por lo que los análisis DGE son utilizados para determinar biomarcadores relacionados con determinada enfermedad o condición y para estudios de sensibilidad farmacológica (Chung et al., 2021).

Inicialmente, los análisis DGE se realizaban por medio de tecnologías de microarreglos las cuales están basadas en hibridación de sondas sobre una superficie sólida. El ARNm extraído se convierte a ADNc marcado fluorescentemente. A continuación se hibrida la muestra sobre la micromatriz que será escaneada para evaluar el nivel de fluorescencia emitida y de esta forma determinar el nivel de expresión (Tao et al., 2017). No obstante, esta tecnología es limitada para catalogar y cuantificar la diversidad del transcriptoma, así como la baja reproducibilidad (Costa-Silva et al., 2017; Oszolak & Milos, 2011).

Con el avance de las tecnologías de secuenciación, los estudios del transcriptoma se han profundizado a lo largo de los años ampliando el rango de detección y proporcionando información cualitativa y cuantitativa sobre los transcritos de uno o varios organismos. La secuenciación de ARN (RNA-seq) evalúa de forma masiva el perfil de ARN que incluye moléculas codificantes y no codificantes como: ARNm, ARNr, miARN, ARNt y otros ARN pequeños (Chung et al., 2021). Uno de los avances más recientes son los estudios de una sola célula (Single-cell RNA-seq) que permite obtener información biológica sobre una única célula, analizar heterogeneidad celular e identificar bases moleculares detrás de determinadas enfermedades (Hong et al., 2020).

Los estudios de transcriptoma han contribuido en la investigación del cáncer de tal forma que se ha logrado comprender la heterogeneidad y evolución de la enfermedad, niveles de expresión, resistencia a tratamientos, desarrollo e influencia del microambiente, inmunoterapia, y selección de biomarcadores específicos (Hong et al., 2020). En primera instancia, la investigación clínica buscaba identificar variantes de secuencias que se asocien con el desarrollo de la enfermedad por medio de secuenciación de

genoma completo o secuenciación de exoma completo. No obstante, debido a la heterogeneidad inter e intratumoral y las ganancias/perdidas cromosómicas, los estudios genómicos se vuelven más específicos al estar complementados con análisis de RNA-seq (Van Allen et al., 2016). De esta forma, Nikas et al. (2020) han considerado el exoma completo y RNA-seq para generar un modelo de clasificación de cáncer de próstata (tejido tumoral versus sano) con sensibilidad y especificidad de alrededor del 95%. Este estudio mostró que la aplicación de herramientas genómicas y transcriptómicas poseen potencial como herramienta diagnóstica.

Análisis de expresión diferencial en cáncer de próstata

Debido a su heterogeneidad, los estudios realizados para esta enfermedad exploran los mecanismos moleculares relacionados a la biosíntesis y el metabolismo de los andrógenos y testosterona, así como mutaciones genéticas y reordenamientos cromosómicos asociados a las etapas de desarrollo como la carcinogénesis, progresión tumoral y metástasis. Estos estudios tienen como finalidad determinar nuevos marcadores diagnósticos, predictivos y pronósticos que, a diferencia del PSA, puedan identificar correctamente el tejido normal del tejido canceroso y sus variantes más agresivas. Por consiguiente, se lograría evitar múltiples biopsias o procesos quirúrgicos innecesarios producto del diagnóstico impreciso, inclusive otorgar tratamientos adecuados o terapias dirigidas (Bergez-Hernández et al., 2022; Sekhoacha et al., 2022).

Bajo esta premisa, el receptor de prostaglandina E4 fue categorizado como un mediador con tendencia tumor génica a través del análisis de perfil transcriptómica unicelular de diferentes pacientes en el estudio realizado por Peng et al. (2022). Se realizó el análisis de perfiles transcriptómicos de una sola célula en muestras de tejido residual de tumores ortotópicos y metastásicos cuyo diagnóstico fue confirmado por medio de histopatología. Se encontró que prostaglandina E4 está envuelta en diferentes

rutas de señalización como el promover la respuesta migratoria del tumor o su metástasis, así como actuar sobre reacciones celulares relacionadas con el sistema inmune.

Por otro lado, las rutas del receptor de andrógenos están relacionadas con las etapas tempranas del cáncer de próstata, y de igual manera, se encuentra la influencia de estas en las vías para el desarrollo de cáncer de próstata resistente a la castración (Jamroze et al., 2021). Por lo que S. Ge et al. (2020) sugiere apuntar a estas rutas para el estudio de tratamientos no hormonales de casos específicos, generan una alternativa a la orquiectomía y se evita el avance de cáncer que no responde a terapias hormonales.

El estudio abordado por Samaržija (2021) realizó el metaanálisis de datos de expresión génica de conjuntos de datos sobre células metastásicas de cáncer de próstata en ganglios, hígado y hueso, y células de tumor primario. Los genes con expresión diferencial en todos los sitios metastásicos analizados pertenecen a la clase de filamentos, adhesión focal y señalización del receptor de andrógenos. Se determinó que en metástasis óseas se experimentan cambios transcripcionales en la señalización de quimiocinas, en la metástasis de ganglios linfáticos se generan cambios en la cascada de señalización y en las metástasis hepáticas hay cambios en la expresión génica del órgano diana (Aurilio et al., 2020).

En resumen, estos estudios presentan diferentes enfoques de análisis en las vías de señalización relacionadas con el desarrollo del cáncer de próstata, ya sea en su etapa inicial o metastásica, logrando un entendimiento amplio sobre de la dinámica del desarrollo de la enfermedad por medio del análisis de la expresión génica. De esta forma se ha encontrado relación entre la prevención, diagnóstico y tratamientos de casos generales y específicos (Liu et al., 2020). Los usos del análisis de expresión diferencial para el estudio del cáncer de próstata van en aumento, desde biomarcadores para su diagnóstico hasta el convertir casos no tratables a tratables mediante inmunoterapia (Li et al., 2022).

Capítulo 3: Metodología

Esquema de desarrollo

El siguiente trabajo consistió en realizar el análisis de expresión diferencial de cáncer de próstata de secuencias obtenidas en bases de datos para identificar biomarcadores en esta patología. En la figura 6 se muestra el esquema de trabajo para el preprocesamiento de secuencias de RNA-seq sugerido en EMBL-EBI Training (2023) donde se optó por utilizar softwares que se acoplen a los recursos computacionales. En este flujo, se utilizaron lecturas crudas en formato FASTQ para posteriormente realizar el análisis de calidad y filtrado de lecturas. A continuación, se realizó el mapeo y pseudoalineación con el transcriptoma de referencia y finalmente el análisis de expresión diferencial. Una vez identificados los genes de interés, se procedió con el análisis de enriquecimiento funcional. Todos los análisis se realizaron en una computadora personal con memoria RAM de 16 GB y procesador Intel® Core™ i5 CPU 2.50 GHz con 4 núcleos y sistema operativo Unix distribución Ubuntu 22.04 LTS.

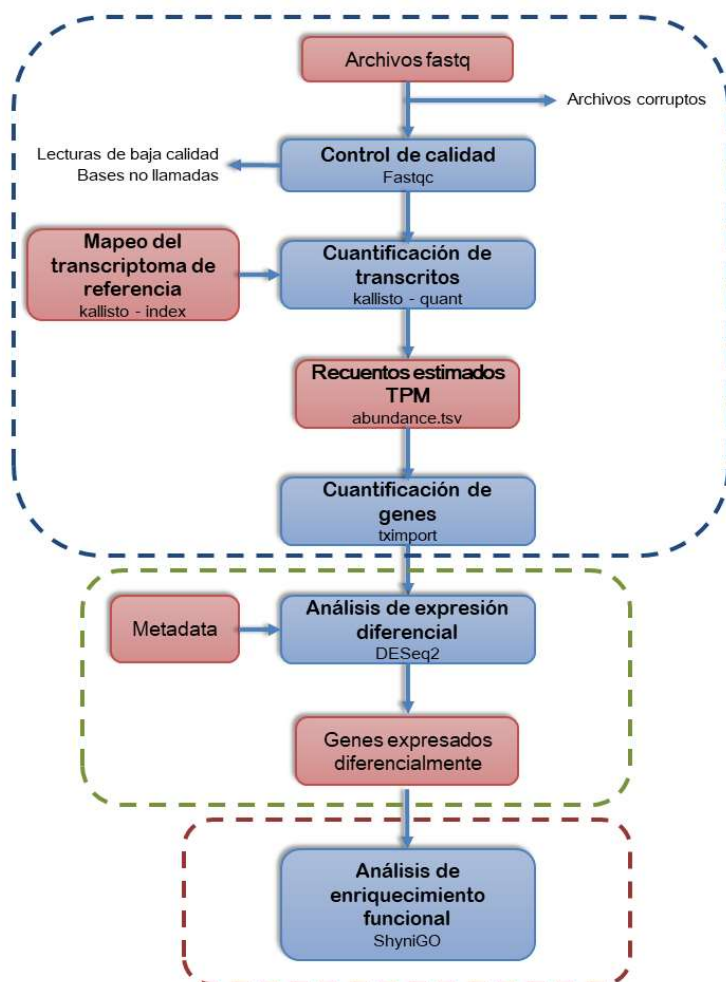
Set de datos

Se trabajó con dos conjuntos de datos. El primer *dataset* (*dataset-1*) fue tomado del Gene Expression Omnibus (GEO) bajo el número de acceso GSE22260. Este set de datos estuvo conformado por 20 muestras de tejido prostático canceroso y 10 muestras de tejido adyacente normal. Las muestras fueron obtenidas de tejido proveniente de prostatectomía radical del Núcleo de Tejidos *Baylor Prostate Specialized Programs of Research Excellence* (SPORE). Los pacientes se encontraban en un rango de edad entre los 39 a 73 años. La extensión tumoral se clasificó de acuerdo con el sistema de estadificación TNM (tumor, ganglios, metástasis). Las muestras se encontraban en estado T2 y T3 (cáncer localizado y cáncer localmente avanzado respectivamente) y puntuación Gleason de 6, 7, 8 y 9. Luego, se prepararon librerías *paired-end* en el sistema Illumina genome analyzer II (GAII) del Centro de Epigenética del

Cáncer Solexa Sequencing Core en la Universidad de Texas-M. Centro Oncológico D. Anderson (Kannan et al., 2011).

Figura 6

Flujo de trabajo para el análisis de expresión diferencial



Nota: El siguiente flujo comienza el preprocesamiento de los datos. Se analiza la calidad con los archivos fastqc y las lecturas de baja calidad son descartadas del análisis. Luego, el genoma de referencia es utilizado para generar un índice de genes que será utilizado para realizar el conteo de la expresión de genes secuenciados. Finalmente, los archivos de abundancia generados son utilizados para el análisis de expresión diferencial. Fuente: EMBL-EBI Training, (2023).

El segundo data set (*dataset2*) fue recuperado del European Nucleotide Archive (ENA) bajo el código de proyecto PRJEB2449. El conjunto de datos está conformado por 28 secuencias pareadas obtenidas de 14 pacientes divididas de la siguiente manera: 14 tumorales y 14 adyacentes normales. Las muestras fueron recolectadas en el Hospital Shanghai Changhai – China y su patología fue confirmada por medio de tinción hematoxilina-eosina. Los pacientes se encontraban en rango de edad entre 50 a 80 años. Los tumores se encontraban en estado T1-T4 y puntuación Gleason de 7,8 y 9. Debido a que existían archivos dañados, se descartaron 4 secuencias. A continuación, se realizó la extracción de ARN, el ARNm fue aislado por medio de perlas con oligos dT. Finalmente, las librerías fueron secuenciadas en la plataforma Illumina HiSeq™ 2000 (Ren et al., 2012).

Preprocesamiento

Las secuencias del *dataset-1* fueron descargadas a través de SRA Toolkit 3.0.3, mientras que las secuencias del *dataset-2* se descargaron del ENA por medio de la herramienta de línea de comando *wget*. Del *dataset-2* se descartaron 2 secuencias tumorales y sus respectivas secuencias normales debido a que la descarga se encontraba bloqueada, su acceso estaba restringido o el archivo estaba corrupto, por lo que se utilizaron únicamente muestras de 11 pacientes (11 normales y 11 tumorales) de dicho set de datos. Se realizó el análisis de calidad en Fastqc (Andrews, 2010) donde se evaluó la presencia de adaptadores, calidad de las secuencias y porcentaje de contenido de guaninas y citocinas (GC). Las secuencias con baja calidad fueron sometidas a recorte y filtrado en Trimmomatic para eliminar adaptadores, bases no llamadas y lecturas con disminución de calidad (Bolger et al., 2014).

El mapeo y cuantificación de transcritos se realizó por medio de la herramienta Kallisto. El algoritmo cuantifica abundancia de transcrito, además presenta un alto rendimiento debido a que acelera el proceso de alineación generando pseudoalineamientos en lugar de alineaciones reales (Bray et al., 2016). Para el mapeo se utilizó en transcriptoma de referencia *Homo sapiens* GRCh38.p53. El índice

generado fue utilizado para la cuantificación de transcrito para cada una de las secuencias, obteniendo como salida archivos `abundance.tsv` que contienen información sobre la abundancia de transcritos.

Análisis de expresión diferencial

El análisis de expresión diferencial se realizó en R Studio por medio del paquete DESeq2 de Bioconductor (Love et al., 2014). DESeq2 utiliza como entrada una matriz recuentos estimados no normalizados (Love et al., 2019) donde las filas corresponden a los genes y las columnas a las muestras (figura 7). La matriz de recuentos fue obtenida a partir de las estimaciones de abundancia obtenidas en la sección anterior e importadas a R por medio del paquete de Bioconductor `tximport` (Soneson et al., 2016). Además, para el análisis se requirió una matriz con información detallada sobre las muestras. Esta metada fue creada de forma independiente en un archivo separado por comas (CSV) de acuerdo con la información descrita por Ren et al. (2012). Los puntos de corte utilizados para el análisis de expresión diferencial fueron $|\log \text{fold change}| > 1.5$ and $\text{valor-p} < 0.05$.

Figura 7

Matriz de recuentos no normalizados



Nota: La matriz de recuentos normalizados $i \times j$ donde las filas i representan los genes, las columnas j las muestras y el contenido de la matriz contiene información sobre el número los recuentos no normalizados. Fuente: Love et al. (2019).

Análisis de enriquecimiento funcional

Los genes expresados diferencialmente tras el análisis de expresión diferencial fueron utilizados para el análisis de enriquecimiento funcional el cual fue realizado en ShinyGO (S. X. Ge et al., 2020). El análisis de ontología génica (GO) fue realizado tanto para genes sobrepresados y subexpresados para las bases de datos de GO componentes celulares y GO función molecular

Capítulo 4: Resultados

Análisis de calidad

El análisis de calidad con FastQC proporciona un conjunto de estadística sencillas que otorga una idea de la calidad de la secuencia sin procesar. La estadística básica se presenta en forma de tabla donde se puede observar información con respecto a la longitud de la secuencia, tamaño de la secuencia en pares de bases (bp) y contenido GC (figura 8). Uno de los parámetros vitales de este análisis es la calidad por base, la cual se esquematiza en forma de gráfico de barras y bigotes. El plano se grafica en función de las posiciones de las lecturas (eje x) y la puntuación de calidad (eje y), donde las lecturas cuya puntuación es menor a 20 se consideran de baja calidad, entre 20 y 28 de calidad media, y mayor a 28 se consideran lecturas de buena calidad (Andrews, 2010). En la figura 9 se observa el resultado de calidad obtenidas en FastQC para una secuencia de cada set de datos después del proceso de limpieza en Trimmomatic.

Figura 8

Estadística básica del análisis de secuencias en FastQC

Basic Statistics

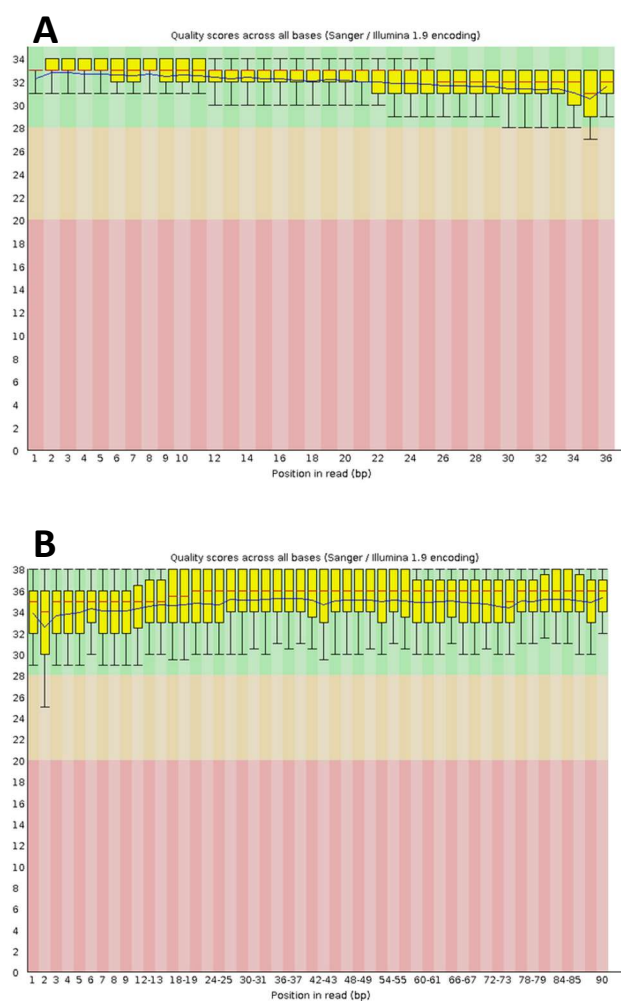
Measure	Value
Filename	paired_SRR057629_1.fastq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	277908
Total Bases	10 Mbp
Sequences flagged as poor quality	0
Sequence length	36
%GC	42

Nota: Cuadro de estadísticas básicas del análisis en FastQC. Se obtiene información general sobre la secuencia analizada como el nombre del archivo, tipo de archivo, tipo de secuenciación, total de

secuencias, total de bases, secuencias flanqueadas de baja calidad, longitud de las secuencias y contenido GC.

Figura 9

Análisis de calidad de secuencias en FastQC



Nota: El análisis de calidad se muestra en el diagrama de barras y bigotes, donde se considera buena calidad aquellas posiciones con puntuaciones mayores a 28. Se presenta el resultado obtenido para las secuencias SRR057629 (A) y ERR031017 (B) del *dataset-1* y 2 respectivamente.

El contenido GC también fue evaluado gráficamente comparando la distribución del contenido por lectura versus la distribución teórica (figura 10). Además, se descartó el contenido de adaptadores (figura 11).

Figura 10

Contenido GC por base del análisis de FastQC

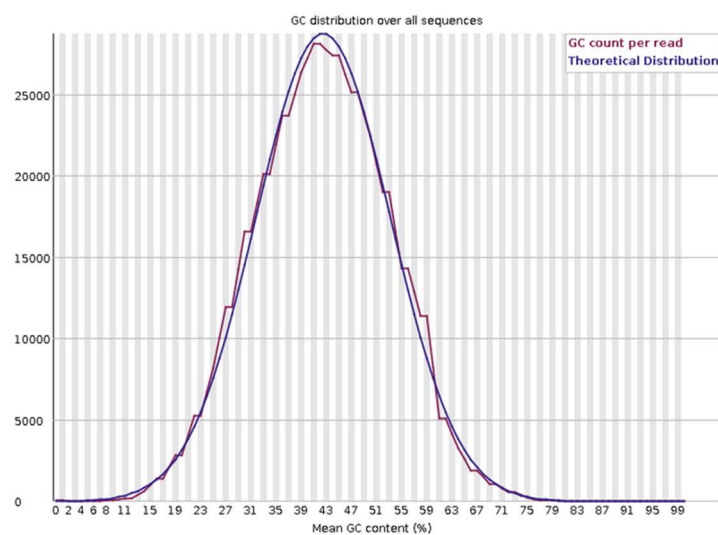
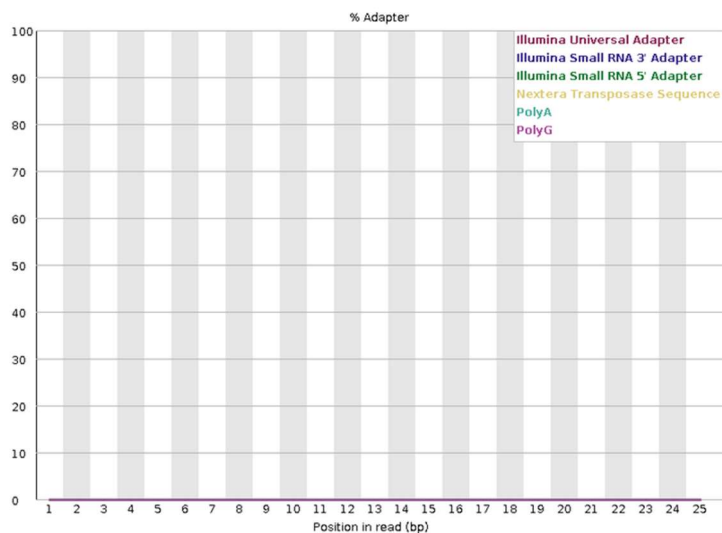


Figura 11

Contenido GC por base del análisis de FastQC



Mapeo y cuantificación de transcritos

El mapeo de transcritos se realizó con la herramienta Kallisto. El número de lecturas mapeadas por secuencia del *dataset-1* y *dataset-2* se muestran en la tabla 1.

Tabla 1

Lecturas mapeadas con Kallisto

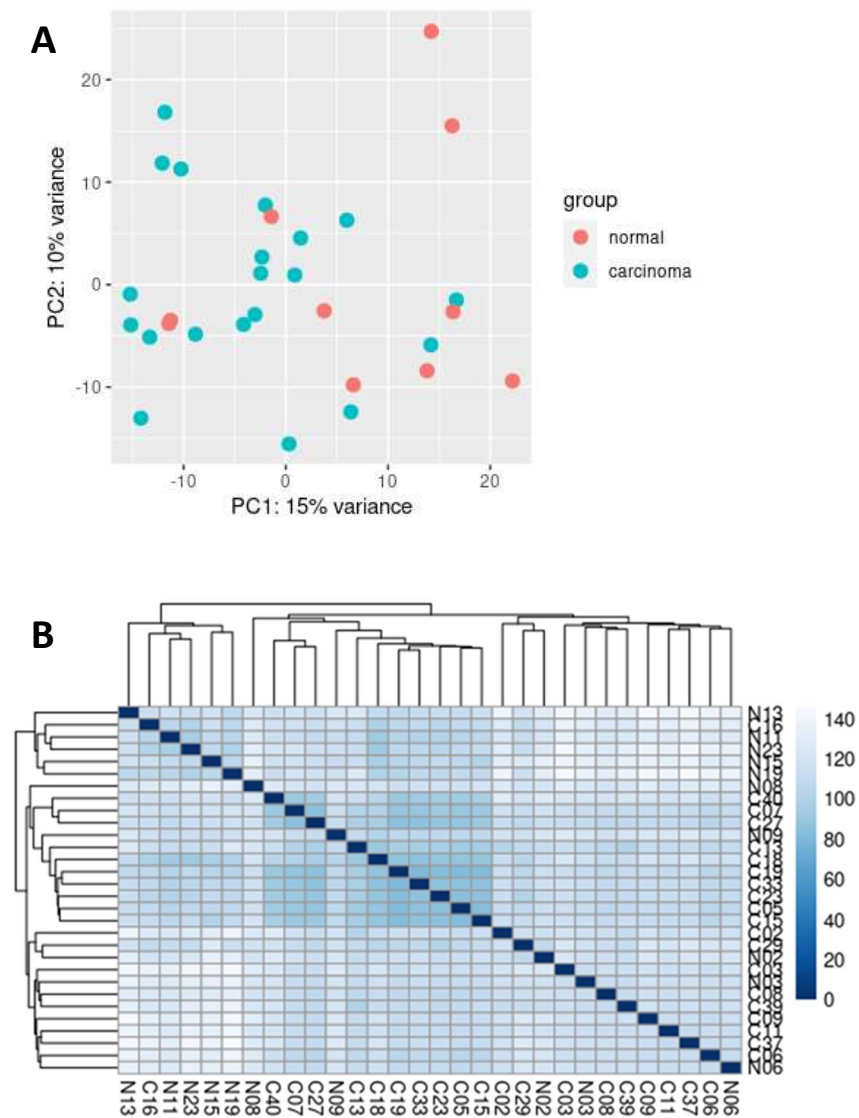
DATASET-1		DATASET-2	
ID Secuencia	Número de lecturas mapeadas	ID Secuencia	Número de lecturas mapeadas
SRR057629	91794	ERR031017	127588
SRR057630	70268	ERR031018	187705
SRR057631	595512	ERR031019	49085
SRR057632	153736	ERR031022	46097
SRR057633	573055	ERR031023	21841
SRR057634	86390	ERR031024	84159
SRR057635	90506	ERR031025	39277
SRR057636	29337	ERR031026	28985
SRR057637	692141	ERR031027	24193
SRR057638	720313	ERR031028	176020
SRR057639	268072	ERR031029	25862
SRR057640	755797	ERR031030	23739
SRR057641	779480	ERR031033	51056
SRR057642	706364	ERR031035	91906
SRR057643	688799	ERR031038	38836
SRR057644	474551	ERR031039	37593
SRR057645	786678	ERR031040	30973
SRR057646	44067	ERR031041	31220
SRR057647	182252	ERR031042	78141
SRR057648	799383	ERR299295	23679

SRR057649	58349	ERR299296	112080
SRR057650	59053	ERR299297	58033
SRR057651	55127	ERR299298	75111
SRR057652	37246	ERR299299	53054
SRR057653	221970		
SRR057654	269727		
SRR057655	425409		
SRR057656	320875		
SRR057657	721522		
SRR057658	487958		

Análisis exploratorio de los datos

Para el análisis exploratorio de datos se obtuvieron dos gráficas: el análisis de componentes principales (PCA) y mapa de calor de la matriz de distancia entre muestras con el objetivo de visualizar las diferencias y semejanzas entre las muestras de carcinoma y las de tejido sano adyacente. El gráfico PCA permite realizar una previsualización de la variabilidad de la expresión génica de los datos. Por otro lado, el gráfico de calor otorga una descripción general entre las similitudes y diferencias entre muestras (Batut et al., 2023; McDermaid et al., 2019). En la figura 12 se observan los resultados del análisis de componentes principales y del mapa de calor respectivamente para el *dataset-1* donde se comparan muestras con carcinoma (C) versus muestras de tejido sano (N). Se puede visualizar que existe poca diferenciación entre las condiciones de análisis. La figura 13 presentan los resultados de PCA y mapa de calor respectivamente para el *dataset-2*. Al contrario del *dataset-1*, el segundo conjunto de datos presenta una clara agrupación entre las condiciones de estudio.

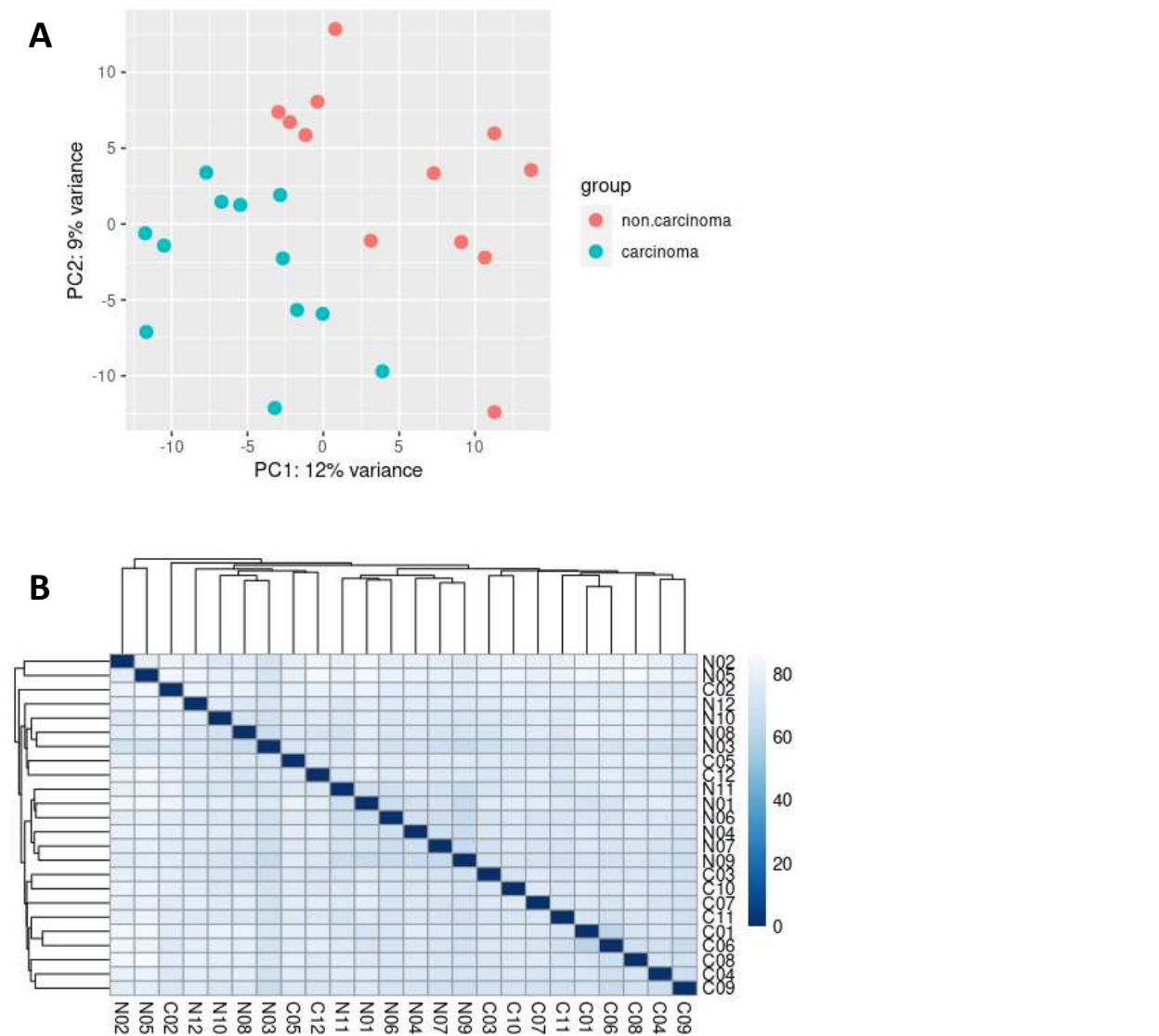
Figura 12

Análisis de exploratorio de genes expresados diferencialmente del dataset-1

Nota: Análisis de componentes principales (PCA) aplicado al primer conjunto de datos comparando tejido normal y carcinoma (A). Mapa de calor de distancias entre muestras cuya escala representa similitudes y diferencias entre muestras, azul (0) indica menor distancia dado los conteos normalizados (B).

Figura 13

Análisis de componentes principales (PCA) del dataset-2



Nota: Análisis de componentes principales (PCA) aplicado al segundo conjunto de datos comparando tejido normal y carcinoma (A). Mapa de calor de distancias entre muestras cuya escala representa similitudes y diferencias entre muestras, azul indica menor distancia dado los conteos normalizados (B).

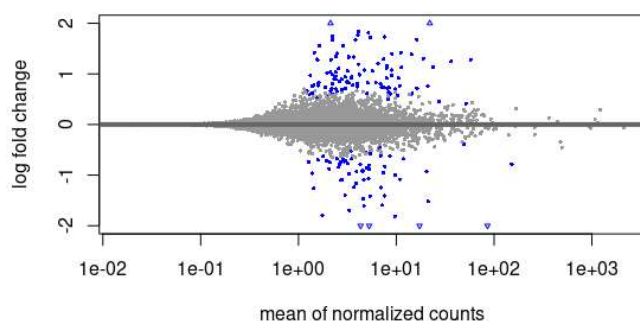
Análisis de expresión diferencial

Dado que el *dataset-1* no muestra diferencias claras entre las condiciones de, se continuó el análisis de expresión diferencial con el *dataset-2*. En este grupo de datos se pudo visualizar que existe un efecto que diferencia los tejidos con carcinoma de los tejidos sanos.

Para identificar genes con cambios de expresión e identificar patrones de agrupación de genes se utilizó un gráfico de comparación entre la diferencia de medias y el promedio (*MA plot, Mean-Average plot*). Este gráfico otorga una representación de la variabilidad global de la expresión génica entre las condiciones. El eje y muestra el cambio de expresión entre grupos o relaciones logarítmica (M), el eje x la fuerza de expresión promedio de los genes o media promedio (A). Los genes que se encuentran altamente diferenciados, es decir por debajo del umbral de significación, toman una coloración azul (Batut et al., 2023; McDermaid et al., 2019). En la figura 14 se muestra el *MA plot* para el *dataset-2*, donde se visualiza una serie de genes o puntos de color azul que se alejan del eje x representando los genes de interés para el estudio.

Figura 14

Gráfico MA para el *dataset2*



Nota: Perfiles de expresión de tejido prostático canceroso (log tumor/sano). Los genes que superaron el umbral de significación (valor p ajustado < 0,05) se representan como puntos azules.

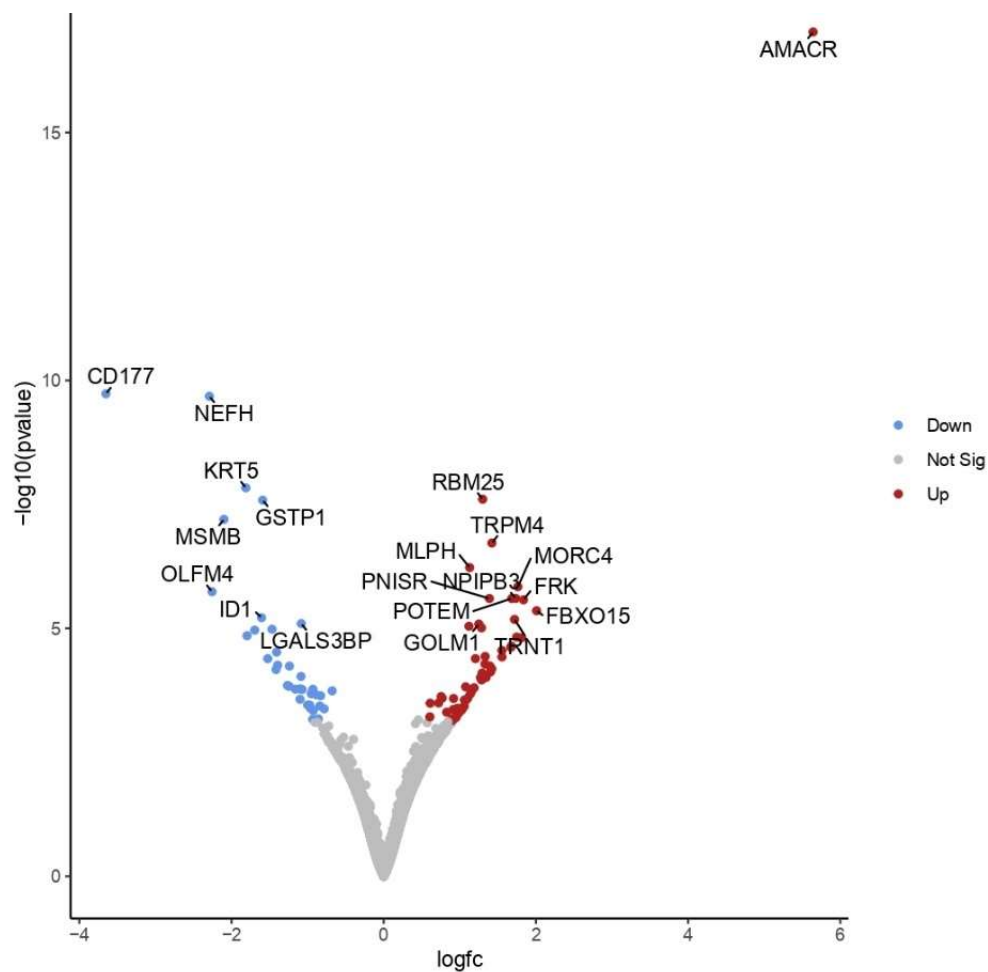
Dado que el gráfico anterior proporciona una visión global de la variabilidad permitiendo identificar patrones de expresión, se utilizó un gráfico de volcán para visualizar la magnitud de los cambios de expresión con significancia estadística. Este gráfico relaciona el valor-p ajustado frente al cambio logarítmico. Por lo tanto, el cambio logarítmico en el eje x indica diferencias en la expresión cuando estos se distribuyen hacia los extremos y muestra niveles similares de expresión cuando se acercan a cero (Batut et al., 2023; McDermaid et al., 2019) En la figura 15 se muestran los resultados del gráfico de volcán correspondiente al *dataset-2*, donde se observa los genes expresados al alza o sobre expresados hacia la derecha (rojo) y los genes regulados a la bajo o sub expresados a la izquierda (azul), en el centro se encuentran los genes que no presentaron expresión diferencial (gris).

Para complementar los análisis anteriores, la figura 16 muestra un mapa de calor de valores z en el cual se puede visualizar la expresión relativa de los genes en función al estadístico z permitiendo normalizar y comparar los niveles de expresión de los genes entre las muestras.

De esta forma, el análisis de expresión diferencial realizado con el paquete de Bioconductor DESeq2 se encontró un total de 105 genes que se expresaban de forma diferencial entre el tejido prostático con carcinoma y el tejido normal adyacente, de los cuales 65 genes se expresaron al alta y 40 genes se expresaron a la baja.

Figura 15

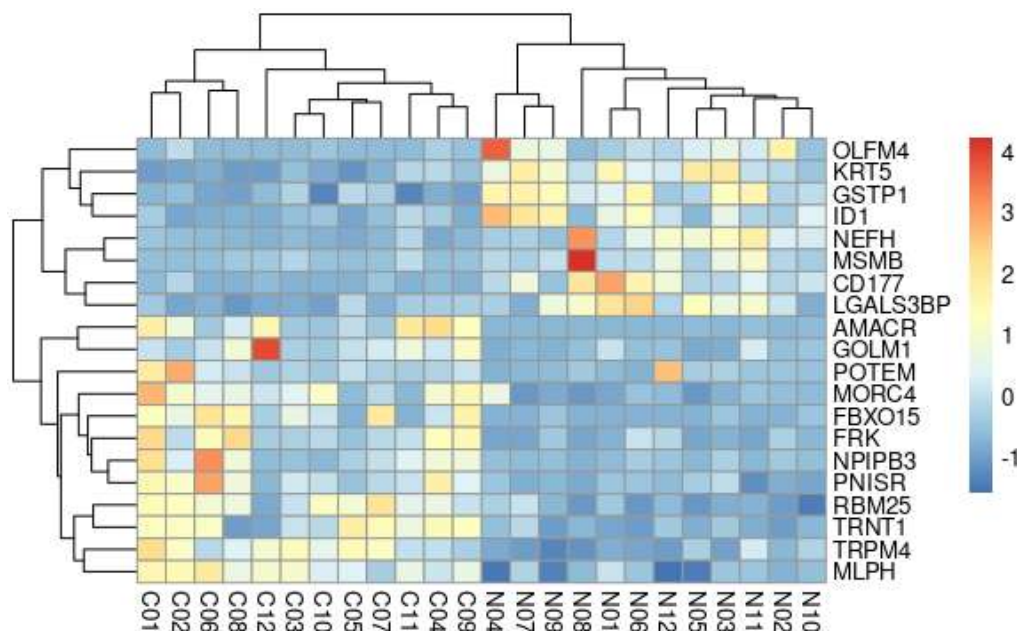
Gráfico de Volcán para el dataset2



Nota: Gráfico de volcán para genes expresados diferencialmente en tejido prostático canceroso. Se etiquetaron los 20 primeros genes que reportan expresión diferencial. Los genes sobre expresados se muestran en rojo y los sub expresados en azul.

Figura 16

Mapa de calor de valores z para el dataset2 de los 20 primeros genes expresados diferencialmente



Nota: Mapa de calor de los 20 primeros genes expresados diferencialmente. La escala indica el puntaje z de los conteos normalizados que indica el número de desviaciones estándar que se alejan de la media. En rojo se muestran las desviaciones estándar por arriba de la media y en azul el número de desviaciones por debajo de la media.

Análisis de enriquecimiento funcional

Una vez determinado los genes expresados diferencialmente, el objetivo es comprender en qué vías o redes estos genes están implicados. El análisis de enriquecimiento funcional permite comprender el mecanismo y funciones biológicas de los resultados obtenidos previamente (Batut et al., 2023).

Para los genes sobrexpresados se obtuvo un enriquecimiento funcional para GO de función molecular representado en la tabla 2. La figura 17 muestra los principales GO obtenidos en Shiny para los genes

sobreexpresados utilizando la base de datos GO de función molecular. Este conjunto de genes no mostró enriquecimiento para GO de componentes celular.

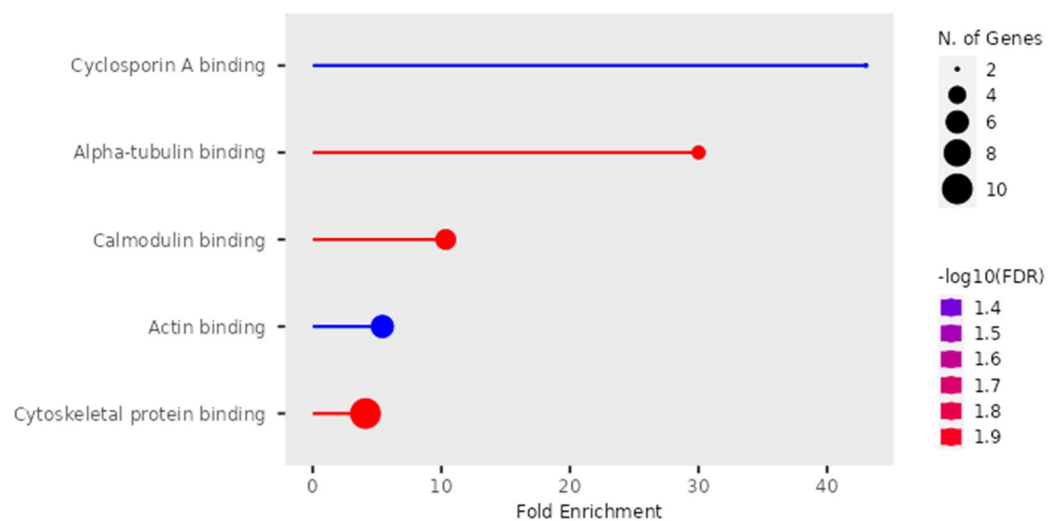
Tabla 2

Análisis de enriquecimiento para genes sobreexpresados

Enriquecimiento FDR	nGenes	Genes en la ruta	Enriquecimiento del pliegue	Ruta
4,89E+14	2	20	4,3011E+13	Unión a la ciclosporina A
1,14E+14	3	43	3,0008E+14	Unión alfa-tubulina
1,14E+14	5	208	1,0339E+13	Unión a calmodulina
4,89E+14	6	477	5,4102E+14	Unión a la actina
1,14E+14	10	1050	4,0963E+12	Unión a proteínas del citoesqueleto

Figura 17

Principales GO de función moleculares para genes sobreexpresados diferencialmente en carcinoma GO de función molecular.



Para los genes subexpresados se repitió el análisis. En esta ocasión se obtuvo enriquecimiento funcional para GO de función molecular y componente celular (tabla 3 y 4). La figura 18 y 19 muestra los

principales GO obtenidos en Shiny para los genes subexpresados utilizando la base de datos GO de función molecular y componente celular respectivamente.

Tabla 3

Análisis de enriquecimiento para genes subexpresados GO función molecular

Enriquecimiento FDR	nGenes	Genes en la ruta	Enriquecimiento del pliegue	Ruta
4,99E+13	1	3	2,2349E+14	Unión de óxido nítrico
4,99E+13	1	3	2,2349E+14	Constituyente estructural del citoesqueleto de filamento intermedio postsináptico
4,99E+13	1	3	2,2349E+14	Actividad ceramida-1-fosfato fosfatasa
1,79E+14	2	7	1,91563E+14	Actividad del transportador transmembrana de urea
2,45E+13	3	18	1,11745E+14	Unión de cadherina implicada en la adhesión célula-célula
1,16E+14	2	25	5,36376E+14	Actividad de glutatión peroxidasa
1,79E+14	3	49	4,10492E+14	Actividad mediadora de la adhesión célula-célula
1,19E+13	4	113	2,37335E+14	Constituyente estructural del citoesqueleto
4,99E+13	2	58	2,31197E+14	Actividad de peroxidasa
4,99E+13	2	62	2,16281E+14	Unión a proteínas de andamio
4,99E+13	2	63	2,12848E+14	Actividad oxidorreductasa, actuando sobre el peróxido como aceptor
2,17E+14	4	152	1,7644E+14	Unión de integrinas
3,77E+14	5	354	9,46992E+13	Unión de cadherina
2,45E+13	8	577	9,29595E+14	Unión de moléculas de adhesión celular
1,04E+13	8	784	6,84154E+14	Actividad de la molécula estructural

Tabla 4

Análisis de enriquecimiento para genes subexpresados GO componente celular

Enriquecimiento FDR	nGenes	Genes en la ruta	Enriquecimiento del pliegue	Ruta
9,50E+14	2	22	6,09519E+13	Apicolateral plasma membrane
4,36E+14	2	67	2,0014E+14	Tertiary granule lumen
1,49E+14	3	132	1,5238E+14	Bicellular tight junction
1,54E+14	6	265	1,51805E+14	Basal plasma membrane
5,28E+14	4	180	1,48993E+14	Adherens junction
1,55E+13	3	140	1,43672E+14	Tight junction
1,69E+14	6	283	1,42149E+13	Basal part of cell

1,88E+14	3	154	1,30611E+13	Apical junction complex
9,35E+14	4	240	1,11745E+14	Basolateral plasma membrane
7,15E+09	8	519	1,03348E+13	Cell-cell junction
1,14E+14	4	273	9,82374E+14	Intermediate filament
3,92E+14	3	209	9,62398E+14	Tertiary granule
3,92E+14	3	212	9,48779E+14	Specific granule
5,28E+14	5	368	9,10965E+13	Secretory granule lumen
5,28E+14	5	372	9,0117E+14	Cytoplasmic vesicle lumen
5,28E+14	5	374	8,96351E+12	Vesicle lumen
7,15E+09	10	907	7,39218E+14	Anchoring junction
3,31E+14	6	975	4,12597E+14	Secretory granule
1,04E+14	8	1323	4,05424E+14	Plasma membrane region

Figura 18

Principales GO de función molecular para genes subexpresados diferencialmente en carcinoma GO de función molecular

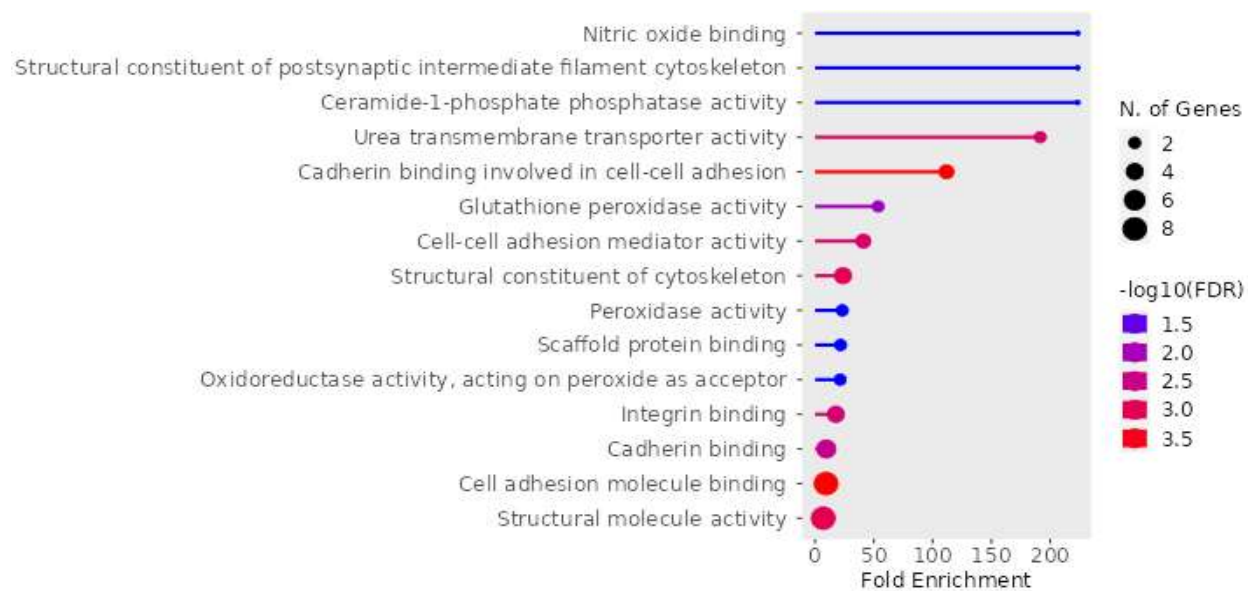
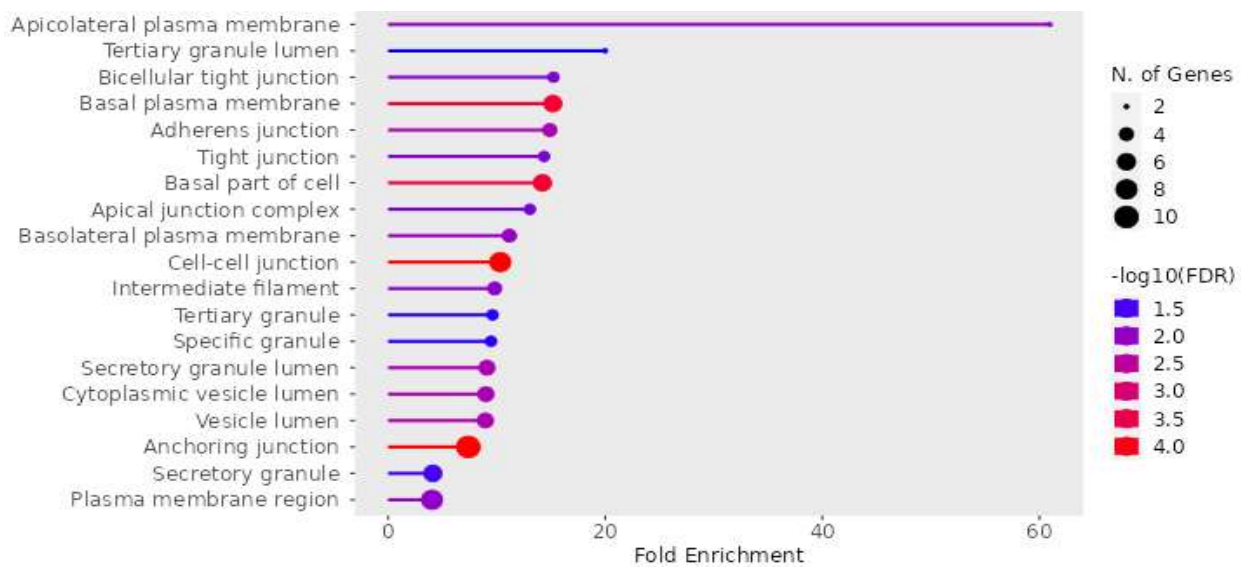


Figura 19

Principales GO de función moleculares para genes subexpresados diferencialmente en carcinoma GO de componentes celulares



Capítulo 5: Discusión

El cáncer de próstata es el segundo tipo de cáncer más común en hombres de edad madura en todo el mundo. La complejidad génica de esta enfermedad dificulta la comprensión clínica y por lo tanto la estimación oportuna de la etapa de desarrollo tumoral y selección terapéutica. Por ende, el entendimiento de los mecanismos genéticos detrás del cáncer proporciona un panorama amplio sobre la evolución de la enfermedad, resistencia al tratamiento y microambiente de desarrollo (Hong et al., 2020).

El estudio de la expresión génica se ha utilizado para la investigación del cáncer permitiendo el diagnóstico, predicción e identificación de biomarcadores. La secuenciación de ARN brinda un amplio catálogo de información sobre las células tumorales a diferencia de otras tecnologías de expresión como los microarreglos (Zhai et al., 2014).

En este trabajo se realizó un estudio de expresión diferencial de cáncer de próstata utilizando dos sets de datos. El primero publicado por Kannan et al. (2011) consistió en 30 secuencias divididas en 20 de carcinoma y 10 normales, mientras que el segundo fue reportado por Ren et al. (2012) donde se tomaron muestras pareadas de 14 pacientes (14 de carcinoma y 14 de tejido sano adyacente). Ambos estudios estuvieron conformados por pacientes masculinos de diversos rangos de edad y cuyo estado de enfermedad se clasificó como tumor localizado (T1-T2) y tumor localmente avanzado (T3-T4).

Para determinar las diferencias en el nivel de expresión entre ambas condiciones, se realizó un análisis de expresión diferencial el cual utiliza pruebas de hipótesis para determinar la diferencia entre la expresión de dos o más condiciones por medio de la evaluación de parámetros estadísticos. El estadístico valor-p permite estimar la probabilidad de que el gen analizado se exprese de manera diferencial entre las condiciones comparadas. Por otro lado, el cambio de pliegue (*foldchange*, FC)

establece el nivel de expresión de cada uno de los genes, de tal forma que un FC positivo representa un aumento en la expresión y un FC negativo una disminución (Deshpande et al., 2023). Makhijani et al. (2018) menciona que los métodos que utilizan FC para evaluar la expresión génica son más reproducibles y sus resultados poseen mayor relevancia biológica, además, el cambio de pliegue permite evaluar los efectos de la expresión génica en la regulación de las vías del cáncer y las funciones celulares a nivel de proteínas. Sin embargo, el nivel seleccionado puede tener diferentes efectos en el resultado por lo que su selección dependerá de la pregunta biológica.

En primer lugar, se realizó un análisis preliminar cuyos resultados proporcionaron una visión general sobre las condiciones de los datos y de esta forma permitieron evaluar el comportamiento de las muestras con carcinoma y tejido normal, y así identificar valores atípicos. El gráfico PCA se utilizó con el objetivo inferir la asociación entre las muestras, mientras que el mapa de calor de distancias permitió evaluar la variabilidad dentro y entre los conjuntos de datos, agrupando los transcriptomas de acuerdo con la distancia basada en la correlación entre dos muestras (Koch et al., 2018).

En la figura 12, correspondientes a los análisis preliminares del *dataset-1*, no se observó una agrupación clara entre ambas condiciones evaluadas. Son et al. (2018) explica que la baja calidad de los datos de RNA-seq, el incorrecto etiquetado de las bibliotecas, la falta de seguridad en la fuente de la muestra o la contaminación durante el proceso pueden producir agrupaciones e interpretaciones erróneas, además esto se puede corregir incluyendo replicas técnicas y biológicas en el experimento. Los datos provenientes del estudio de (Kannan et al., 2011) fueron obtenidos con el objetivo de caracterizar ARN quiméricos en cáncer de próstata por medio de secuencias de RNA-seq. En este estudio se utilizó la prueba Kolmogorov-Smirnov para identificar quiméricos que difieran entre muestras por lo que no se realizó un análisis de expresión diferencial propiamente dicho y, tal como menciona Makhijani et al. (2018) los métodos que no evalúan o consideran el cambio de pliegue son poco reproducibles. Por lo

tanto, la falta de diferenciación entre las condiciones de estudio en este set de datos dificultó realizar un análisis de expresión diferencial y por consiguiente un análisis de enriquecimiento funcional. Por otro lado, los resultados del PCA y heatmap para el *dataset-2* (figura 13) arrojaron resultados congruentes, donde se observa una clara diferenciación entre carcinoma y tejido sano lo que sugiere la existencia de genes expresados diferencialmente.

La identificación de patrones de expresión se realizó por medio de un MA plot (figura 14) donde se observa puntos en azul que sugiere que existen cambios de expresión de forma consistente.

(McDermaid et al., 2019) menciona que la comparación entre el cambio de pliegue logarítmico y la expresión media provocan que el gráfico tome forma de campana y que los puntos que se alejan hacia los extremos son relevantes para el estudio de expresión, mientras que los puntos que se acumulan en el centro poseen baja variabilidad.

Debido a que el gráfico MA no proporciona medidas de significancia estadística, no ofrece información sobre qué genes se expresan de forma diferencial con significancia estadística (McDermaid et al., 2019). Por lo que, se utilizó un gráfico de volcán para representar el valor P ajustado versus el cambio de pliegue logarítmico de tal forma que los genes expresados al alta se ubican hacia la derecha, los genes expresados a la baja hacia la izquierda y los genes con niveles de expresión similares se localizan en el centro (figura 15).

De forma complementaria, en la figura 16 se observa el mapa de calor del estadístico z. Por medio de este gráfico se puede observar los niveles de expresión por medio de la clave de color, además de visualizar los patrones de agrupación, donde se observa un grupo de genes sobre expresados para las muestras de tejido sano y otro grupo distinto de genes con niveles de expresión alta para las muestras de carcinoma. Utilizar los valores z para realizar este gráfico es una forma sencilla de interpretar el

resultado ya que se representa el número de desviaciones estándar se aleja de la media de los conteos normalizados para determinado gen (Batut et al., 2023).

En el estudio realizado por Ren et al. (2012) se encontraron genes de fusión y ARN largos no codificantes los cuales se encontraban fuertemente relacionados a la regulación transcripciones de múltiples genes. Dentro de este grupo se encontraba el gen PCA3 reportado en múltiples estudios como un biomarcador capaz de diferenciar patologías malignas de las benignas (Amaro et al., 2014). En contraste, el análisis de expresión diferencial realizado en este trabajo dio como resultado 105 genes expresados diferencialmente asociados con la síntesis de tubulinas y procesos de adhesión celular.

Las rutas enriquecidas para los genes expresados diferencialmente hacia la baja se relacionan con la síntesis de tubulina/microtubulos. La tubulina es una proteína cuya unión forma los microtúbulos cuyo rol estructural es fundamental en la separación y segregación cromosómica durante la mitosis, además facilitan el proceso de translocación de carga celular convirtiéndose en el andamio de movilización de diversos procesos como el arresto mitótico o la apoptosis (Zhu et al., 2008).

En el cáncer de próstata los microtúbulos desempeñan en el proceso de translocación del receptor de andrógenos, principal impulsor oncogénico prostático. Además, en cánceres de bajo y mediano grado se han detectado anomalías en los centrosomas conformados estructuralmente por tubulina provocando una desregulación génica que favorece la supervivencia de las células anormales y progresión del cáncer (Martin et al., 2014).

En el caso de las rutas enriquecidas para los genes regulados a la baja, estos se encuentran involucrados en el proceso de adhesión celular. Las cadherinas son proteínas de membrana que intervienen en la adhesión célula-célula, siendo las cadherinas clásicas (E, P, N y R) las que juegan un rol fundamental en la segregación de diferentes poblaciones celulares (Perez & Nelson, 2004).

En el estudio realizado por (Murali et al., 2013) se sostiene que la expresión anormal de E-cadherina se asocia con una disminución en la adhesión celular, lo cual se relaciona con la diseminación del cáncer de próstata. Además, se reporta que, durante el complejo proceso de metástasis, causado por la propagación de células anormales, se observa una pérdida de adhesión y desprendimiento celular desde el tumor primario. Adicionalmente menciona que, en estudios realizados en pacientes con metástasis ósea provocada por cáncer prostático, los niveles de E-cadherina son bajos.

Los genes involucrados en los procesos de síntesis de tubulina y adhesión celular son posibles biomarcadores. De esta manera, debido a que el trabajo realizado utilizó secuencias de RNA-seq obtenidas de muestras que no fueron expuestas a ningún tratamiento, los genes expresados diferencialmente pueden ser candidatos a biomarcadores pronósticos, mas no predictivos. Sin embargo, como menciona (Ballman, 2015) existen marcadores que cumplen la función predictiva y pronóstico. En la investigación realizada por (Zhu et al., 2008) se evidencia que el uso de fármacos que bloquean la señalización regulación de la actividad del receptor de andrógenos promueven la supresión del cáncer de prostático. Por lo tanto, los genes involucrados en la síntesis de tubulina podrían actuar como marcadores predictivos.

Capítulo 6: Conclusiones y Recomendaciones

El presente trabajo busca identificar posibles biomarcadores para cáncer de próstata por medio de un análisis de expresión diferencial. En primer lugar, se obtuvieron dos sets de datos de secuencias de ARN provenientes carcinoma y tejido sano a partir de dos proyectos independientes disponibles en los repositorios GEO y ENA.

De forma preliminar se realizó un análisis exhaustivo para evaluar la calidad de las secuencias de ARN para asegurar la integridad de los datos obtenidos. Los resultados del primer *dataset* se observó la ausencia de diferencias significativas entre las condiciones a evaluar producto de posibles inconvenientes aguas arriba como baja calidad de los datos de RNA-seq, incorrecto etiquetado de las bibliotecas, la falta de seguridad en la fuente de la muestra, contaminación o poca diferenciación transcriptómica entre las muestras. Por este motivo se optó por utilizar un nuevo set de datos cuyas anotaciones y resultados publicados presenten análisis de expresión. Los resultados de análisis preliminar para el *dataset-2* fueron favorables debido a que se observó claramente diferencias en la expresión de células con carcinoma versus las células sanas.

En análisis de expresión diferencial comparando ambas condiciones a un nivel de significación estadística de 0,05 utilizando el paquete de Bioconductor DESeq2, se obtuvo un total de 105 genes expresados diferencialmente de los cuales 65 se sobre expresaron y 40 se sub expresaron.

El análisis de enriquecimiento funcional en ShinyGO mostró que los genes expresados de forma diferencial se relacionaban con mecanismos asociados a la síntesis de tubulina y procesos de adhesión celular, procesos que ya se han asociado a cáncer. Dado que, los genes involucrados en dichos mecanismos son potenciales marcadores pronósticos, ya que la evaluación se realizó comparando dos condiciones sin la exposición a determinado tratamiento. No obstante, la bibliografía sugiere que los

fármacos enfocados a bloquear las rutas de señalización en la síntesis de tubulina suprimen el desarrollo tumoral prostático por lo que, en futuros trabajos, se propone evaluar la expresión diferencial en pacientes con cáncer expuestos al tratamiento con este tipo de medicación y así determinar si efectivamente estos genes son potenciales marcadores predictivos.

Referencias

- Aaron, L. T., Franco, O. E., & Hayward, S. W. (2016). Review of Prostate Anatomy and Embryology and the Etiology of Benign Prostatic Hyperplasia. *Urologic Clinics of North America*, 43(3), 279–288. <https://doi.org/10.1016/j.ucl.2016.04.012>
- Adhyam, M., & Gupta, A. K. (2012). A Review on the Clinical Utility of PSA in Cancer Prostate. *Indian Journal of Surgical Oncology*, 3(2), 120–129. <https://doi.org/10.1007/s13193-012-0142-6>
- Allemailem, K. S., Almatroudi, A., Alrumaihi, F., Almansour, N. M., Aldakheel, F. M., Rather, R. A., Afroze, D., & Rah, B. (2021). Single nucleotide polymorphisms (SNPs) in prostate cancer: Its implications in diagnostics and therapeutics. *American Journal of Translational Research*, 13(4), 3868–3889.
- Amaro, A., Esposito, A. I., Gallina, A., Nees, M., Angelini, G., Albin, A., & Pfeffer, U. (2014). Validation of proposed prostate cancer biomarkers with gene expression data: a long road to travel. *Cancer Metastasis Reviews*, 33(2), 657. <https://doi.org/10.1007/S10555-013-9470-4>
- American Cancer Society. (2019). Detección temprana , diagnóstico y clasificación por etapas ¿ Se puede descubrir el cáncer de endometrio en sus comienzos ? *American Cancer Society*, 1–22. <https://www.cancer.org/es/cancer/cancer-de-endometrio/deteccion-diagnostico-clasificacion-por-etapas/como-se-diagnostica.html>
- Andrews, S. (2010). *Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data*. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Aurilio, G., Cimadamore, A., Mazzucchelli, R., Lopez-Beltran, A., Verri, E., Scarpelli, M., Massari, F., Cheng, L., Santoni, M., & Montironi, R. (2020). Androgen Receptor Signaling Pathway in Prostate Cancer: From Genetics to Clinical Applications. *Cells*, 9(12), 1–14.

<https://doi.org/10.3390/cells9122653>

Ballman, K. V. (2015). Biomarker: Predictive or prognostic? *Journal of Clinical Oncology*, 33(33), 3968–3971. <https://doi.org/10.1200/JCO.2015.63.3651>

Batut, B., Freeberg, M., Heydarian, M., Erxleben, A., Videm, P., Blank, C., Doyle, M., Soranzo, N., Van Heusden, P., & Delisle, L. (2023). *Reference-based RNA-Seq data analysis*.

<https://training.galaxyproject.org/training-material/topics/transcriptomics/tutorials/reference-based/tutorial.html#functional-enrichment-analysis-of-the-de-genes>

Benafif, S., Ni Raghallaigh, H., McHugh, J., & Eeles, R. (2021). Genetics of prostate cancer and its utility in treatment and screening. *Advances in Genetics*, 108, 147–199.

<https://doi.org/10.1016/BS.ADGEN.2021.08.006>

Bergez-Hernández, F., Arámbula-Meraz, E., Alvarez-Arrazola, M., Irigoyen-Arredondo, M., Luque-Ortega, F., Martínez-Camberos, A., Cedano-Prieto, D., Contreras-Gutiérrez, J., Martínez-Valenzuela, C., & García-Magallanes, N. (2022). Expression Analysis of miRNAs and Their Potential Role as Biomarkers for Prostate Cancer Detection. *American Journal of Men's Health*, 16(5).

<https://doi.org/10.1177/15579883221120989>

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>

Brandão, A., Paulo, P., & Teixeira, M. R. (2020). Hereditary predisposition to prostate cancer: From genetics to clinical implications. *International Journal of Molecular Sciences*, 21(14), 1–23.

<https://doi.org/10.3390/ijms21145036>

Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016). Near-optimal probabilistic RNA-seq

quantification. *Nature Biotechnology*, 34(5), 525–527. <https://doi.org/10.1038/nbt.3519>

Caballero, H. (2021). *Boletín epidemiológico Cáncer de Protata Sociedad de Lucha Contra el Cáncer SOLCA NÚCLEO QUITO* (Vol. 02). <https://solcaquito.org.ec/boletin-epidemiologico-ano-1-vol-02-mayo-junio-2021/>

Chung, M., Bruno, V. M., Rasko, D. A., Cuomo, C. A., Muñoz, J. F., Livny, J., Shetty, A. C., Mahurkar, A., & Dunning Hotopp, J. C. (2021). Best practices on the differential expression analysis of multi-species RNA-seq. *Genome Biology*, 22(1), 1–23. <https://doi.org/10.1186/s13059-021-02337-8>

Colmener, L., Troconis, E., Cornejo, F., Noboa, A., Oswaldo, R., & Lupera, H. (2018). Aplicación de la Tomografía por emisión de positrones en el cáncer de próstata marcado con 68 Ga-PSMA: Revisión y actualidad. *Federación Ecuatorina de Radiología e Imagen*, 11(1), 13–20.

Costa-Silva, J., Domingues, D., & Lopes, F. M. (2017). RNA-Seq differential expression analysis: An extended review and a software tool. *PLoS ONE*, 12(12), 1–18. <https://doi.org/10.1371/journal.pone.0190152>

Deshpande, D., Chhugani, K., Chang, Y., Karlsberg, A., Loeffler, C., Zhang, J., Muszyńska, A., Munteanu, V., Yang, H., Rotman, J., Tao, L., Balliu, B., Tseng, E., Eskin, E., Zhao, F., Mohammadi, P., P. Łabaj, P., & Mangul, S. (2023). RNA-seq data science: From raw data to effective interpretation. *Frontiers in Genetics*, 14(March), 1–12. <https://doi.org/10.3389/fgene.2023.997383>

EMBL-EBI Training, E.-E. (2023). *Differential gene expression analysis | Functional genomics II*. <https://www.ebi.ac.uk/training/online/courses/functional-genomics-ii-common-technologies-and-data-analysis-methods/rna-sequencing/performing-a-rna-seq-experiment/data-analysis/differential-gene-expression-analysis/>

- Epstein, J. I., Egevad, L., Amin, M. B., Delahunt, B., Srigley, J. R., & Humphrey, P. A. (2016). The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma. *American Journal of Surgical Pathology*, *40*(2), 244–252.
<https://doi.org/10.1097/pas.0000000000000530>
- Ferlay, J., Ervik, M., Lam, F., Colombet, M., Mery, L., Piñeros, M., Znaor, A., Soerjomataram, I., & Bray, F. (2020). *Global Cancer Observatory: Cancer Today*. International Agency for Research on Cancer, Lyon, France. <https://gco.iarc.fr/today/home>
- Frommlet, F., Szulc, P., König, F., & Bogdan, M. (2022). Selecting predictive biomarkers from genomic data. *PLoS ONE*, *17*(6). <https://doi.org/10.1371/JOURNAL.PONE.0269369>
- Fundora, T., Hernández, L., & Chavéz, M. (2020). Tomografía por emisión de positrones/tomografía computarizada en la evaluación del resultado del tratamiento de linfomas. *Revista Cubana de Hematología, Inmunología y Hemoterapia*, *36*(4), 2–7. <http://orcid.org/0000-0001-9754-3113>
- Ge, S., Sun, C., Hu, Q., Guo, Y., Xia, G., Mi, Y., & Zhu, L. (2020). Differential expression profiles of circRNAs in human prostate cancer based on chip and bioinformatic analysis. *International Journal of Clinical and Experimental Pathology*, *13*(5), 1045.
[/pmc/articles/PMC7270690/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7270690/)
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7270690/?report=abstract>
- Ge, S. X., Jung, D., Jung, D., & Yao, R. (2020). ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics*, *36*(8), 2628–2629. <https://doi.org/10.1093/BIOINFORMATICS/BTZ931>
- Hong, M., Tao, S., Zhang, L., Diao, L. T., Huang, X., Huang, S., Xie, S. J., Xiao, Z. D., & Zhang, H. (2020). RNA sequencing: new technologies and applications in cancer research. *Journal of Hematology and*

Oncology, 13(1), 1–16. <https://doi.org/10.1186/s13045-020-01005-x>

Institute for Quality and Efficiency in Health Care. (2016, August 23). *How does the prostate work?*

Institute for Quality and Efficiency in Health Care (IQWiG).

<https://www.ncbi.nlm.nih.gov/books/NBK279291/>

Jamroze, A., Chatta, G., & Tang, D. G. (2021). Androgen receptor (AR) heterogeneity in prostate cancer and therapy resistance. *Cancer Letters*, 518, 1–9. <https://doi.org/10.1016/j.canlet.2021.06.006>

Kannan, K., Wang, L., Wang, J., Ittmann, M. M., Li, W., & Yen, L. (2011). Recurrent chimeric RNAs enriched in human prostate cancer identified by deep sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, 108(22), 9172–9177.

<https://doi.org/10.1073/pnas.1100489108>

Koch, C. M., Chiu, S. F., Akbarpour, M., Bharat, A., Ridge, K. M., Bartom, E. T., & Winter, D. R. (2018). A beginner's guide to analysis of RNA sequencing data. *American Journal of Respiratory Cell and Molecular Biology*, 59(2), 145–157. <https://doi.org/10.1165/rcmb.2017-0430TR>

Kretschmer, A., & Tilki, D. (2017). Biomarkers in prostate cancer – Current clinical utility and future perspectives. *Critical Reviews in Oncology/Hematology*, 120(September), 180–193.

<https://doi.org/10.1016/j.critrevonc.2017.11.007>

Li, Y., Shi, H., Zhao, Z., & Xu, M. (2022). Identification of castration-dependent and -independent driver genes and pathways in castration-resistant prostate cancer (CRPC). *BMC Urology*, 22(1), 1–13.

<https://doi.org/10.1186/s12894-022-01113-5>

Liu, S., Wang, W., Zhao, Y., Liang, K., & Huang, Y. (2020). Identification of Potential Key Genes for Pathogenesis and Prognosis in Prostate Cancer by Integrated Analysis of Gene Expression Profiles

and the Cancer Genome Atlas. *Frontiers in Oncology*, 10(June), 1–13.

<https://doi.org/10.3389/fonc.2020.00809>

Love, M. I., Anders, S., Kim, V., & Huber, W. (2019). *RNA-seq workflow: gene-level exploratory analysis and differential expression*.

<http://master.bioconductor.org/packages/release/workflows/vignettes/rnaseqGene/inst/doc/rnaseqGene.html>

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), 1–21. <https://doi.org/10.1186/s13059-014-0550-8>

Martin, S. K., Kamelgarn, M., & Kyprianou, N. (2014). Cytoskeleton targeting value in prostate cancer treatment. *American Journal of Clinical and Experimental Urology*, 2(1), 15–26.

<http://www.ncbi.nlm.nih.gov/pubmed/25374905><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4219288>

McDermaid, A., Monier, B., Zhao, J., Liu, B., & Ma, Q. (2019). Interpretation of differential gene expression results of RNA-seq data: Review and integration. *Briefings in Bioinformatics*, 20(6), 2044–2054. <https://doi.org/10.1093/bib/bby067>

Mucci, L. A., Hjelmborg, J. B., Harris, J. R., Czene, K., Havelick, D. J., Scheike, T., Graff, R. E., Holst, K., Möller, S., Unger, R. H., McIntosh, C., Nuttall, E., Brandt, I., Penney, K. L., Hartman, M., Kraft, P., Parmigiani, G., Christensen, K., Koskenvuo, M., ... Halekoh, U. (2016). Familial risk and heritability of cancer among twins in nordic countries. *JAMA - Journal of the American Medical Association*, 315(1), 68–76. <https://doi.org/10.1001/jama.2015.17703>

- Murali, A. K., Norris, J. S., Murali, A. K., & Norris, J. S. (2013). The Role of E-Cadherin-Catenin Complex in Prostate Cancer Progression. *Advances in Prostate Cancer*. <https://doi.org/10.5772/52751>
- Ni Raghallaigh, H., & Eeles, R. (2022). Genetic predisposition to prostate cancer: an update. *Familial Cancer*, 21(1), 101–114. <https://doi.org/10.1007/s10689-021-00227-3>
- Nikas, J. B., Mitanis, N. T., & Nikas, E. G. (2020). Whole Exome and Transcriptome RNA-Sequencing Model for the Diagnosis of Prostate Cancer. *ACS Omega*, 5(1), 481–486. https://doi.org/10.1021/ACSOMEGA.9B02995/SUPPL_FILE/AO9B02995_SI_001.XLSX
- Ozsolak, F., & Milos, P. (2011). RNA sequencing: advances, challenges and opportunities. *Plus One*, 12(2), 87–98. <https://doi.org/10.1038/nrg2934.RNA>
- Peng, S., Hu, P., Xiao, Y. T., Lu, W., Guo, D., Hu, S., Xie, J., Wang, M., Yu, W., Yang, J., Chen, H., Zhang, X., Zhu, Y., Wang, Y., Yang, Y., Zhu, G., Chen, S., Wang, J., Zhang, B., ... Ren, S. (2022). Single-Cell Analysis Reveals EP4 as a Target for Restoring T-Cell Infiltration and Sensitizing Prostate Cancer to Immunotherapy. *Clinical Cancer Research*, 28(3), 552–567. <https://doi.org/10.1158/1078-0432.CCR-21-0299>
- Perez, T. D., & Nelson, W. J. (2004). *Cadherin Adhesion: Mechanisms and Molecular Interactions*. 5(165), 3–21. https://doi.org/10.1007/978-3-540-68170-0_1
- Pinskaya, M., Saci, Z., Gallopin, M., Gabriel, M., Nguyen, H. T. N., Firlej, V., Describes, M., Rapinat, A., Gentien, D., De La Taille, A., Londoño-Vallejo, A., Allory, Y., Gautheret, D., & Morillon, A. (2019). Reference-free transcriptome exploration reveals novel RNAs for prostate cancer diagnosis. *Life Science Alliance*, 2(6), 1–12. <https://doi.org/10.26508/lsa.201900449>
- Real, J., Jaramillo, L., & Quinto, R. (2020). *Mortalidad por Cáncer de Próstata en Guayaquil. Periodo*

2010- 2019. [http://www.estadisticas.med.ec/Publicaciones/1 Boletin Epi Ca Próstata 2020.pdf](http://www.estadisticas.med.ec/Publicaciones/1%20Boletin%20Epi%20Ca%20Pr%C3%B3stata%202020.pdf)

Reis, R. B. dos, Alías-Melgar, A., Martínez-Cornelio, A., Neciosup, S. P., Sade, J. P., Santos, M., & Villoldo, G. M. (2020). Prostate Cancer in Latin America: Challenges and Recommendations. *Cancer Control*, 27(1), 1–8. <https://doi.org/10.1177/1073274820915720>

Ren, S., Peng, Z., Mao, J. H., Yu, Y., Yin, C., Gao, X., Cui, Z., Zhang, J., Yi, K., Xu, W., Chen, C., Wang, F., Guo, X., Lu, J., Yang, J., Wei, M., Tian, Z., Guan, Y., Tang, L., ... Sun, Y. (2012). RNA-seq analysis of prostate cancer in the Chinese population identifies recurrent gene fusions, cancer-associated long noncoding RNAs and aberrant alternative splicings. *Cell Research*, 22(5), 806–821. <https://doi.org/10.1038/cr.2012.30>

Robles, A., Roc, T., Huarte, G., Arregu, E. A., & Morales, S. (2019). *La próstata : generalidades y patologías más frecuentes*. 62, 41–54.

Samaržija, I. (2021). Site-specific and common prostate cancer metastasis genes as suggested by meta-analysis of gene expression data. *Life*, 11(7), 1–16. <https://doi.org/10.3390/life11070636>

Schatten, H. (2018). Brief overview of prostate cancer statistics, grading, diagnosis and treatment strategies. In *Advances in Experimental Medicine and Biology* (Vol. 1095, pp. 1–14). https://doi.org/10.1007/978-3-319-95693-0_1

Sekhoacha, M., Riet, K., Motloun, P., Gumenku, L., Adegoke, A., & Mashele, S. (2022). Prostate Cancer Review: Genetics, Diagnosis, Treatment Options, and Alternative Approaches. *Molecules*, 27(17), 1–33. <https://doi.org/10.3390/molecules27175730>

Shen, M. M., & Abate-Shen, C. (2010). Molecular genetics of prostate cancer: New prospects for old challenges. *Genes and Development*, 24(18), 1967–2000. <https://doi.org/10.1101/gad.1965810>

- Shi, H., Xu, H., Chai, C., Qin, Z., & Zhou, W. (2022). Integrated bioinformatics analysis of potential biomarkers for pancreatic cancer. *Journal of Clinical Laboratory Analysis*, 36(5), 455–460.
<https://doi.org/10.1002/jcla.24381>
- Sokolova, A. O., Obeid, E. I., & Cheng, H. H. (2021). Genetic Contribution to Metastatic Prostate Cancer. *Urologic Clinics of North America*, 48(3), 349–363. <https://doi.org/10.1016/J.UCL.2021.03.005>
- Son, K., Yu, S., Shin, W., Han, K., & Kang, K. (2018). A simple guideline to assess the characteristics of RNA-Seq Data. *BioMed Research International*, 2018. <https://doi.org/10.1155/2018/2906292>
- Soneson, C., Love, M. I., & Robinson, M. D. (2016). Differential analyses for RNA-seq: Transcript-level estimates improve gene-level inferences. *F1000Research*, 4, 1–19.
<https://doi.org/10.12688/F1000RESEARCH.7563.2>
- Strimbu, K., & Tavel, J. A. (2010). What are Biomarkers? *Current Opinion in HIV and AIDS*, 5(6), 463.
<https://doi.org/10.1097/COH.0B013E32833ED177>
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, 71(3), 209–249.
<https://doi.org/10.3322/caac.21660>
- Tao, Z., Shi, A., Li, R., Wang, Y., Wang, X., & Zhao, J. (2017). Microarray bioinformatics in cancer- A review. *Journal of B.U.ON.*, 22(4), 838–843.
- Tian, J. Y., Guo, F. J., Zheng, G. Y., & Ahmad, A. (2018). Prostate cancer: Updates on current strategies for screening, diagnosis and clinical implications of treatment modalities. *Carcinogenesis*, 39(3), 307–317. <https://doi.org/10.1093/carcin/bgx141>

- Van Allen, E. M., Robinson, D., Morrissey, C., Pritchard, C., Imamovic, A., Carter, S., Rosenberg, M., McKenna, A., Wu, Y. M., Cao, X., Chinnaiyan, A., Garraway, L., & Nelson, P. S. (2016). A comparative assessment of clinical whole exome and transcriptome profiling across sequencing centers: implications for precision cancer medicine. *Oncotarget*, *7*(33), 52888. <https://doi.org/10.18632/ONCOTARGET.9184>
- Waldron, N., Chowdhury, S., & McCahill, C. (2023). Prostate cancer. *Medicine (United Kingdom)*, *51*(1), 48–52. <https://doi.org/10.1016/j.mpmed.2022.10.013>
- Wang, G., Zhao, D., Spring, D. J., & Depinho, R. A. (2018). Genetics and biology of prostate cancer. *Genes & Development*, *32*, 1105–1140. <https://doi.org/10.1101/gad.315739.118.GENES>
- Wasim, S., Lee, S. Y., & Kim, J. (2022). Complexities of Prostate Cancer. *International Journal of Molecular Sciences*, *23*(22). <https://doi.org/10.3390/ijms232214257>
- Zhai, W., Yao, X. D., Xu, Y. F., Peng, B., Zhang, H. M., Liu, M., Huang, J. H., Wang, G. C., & Zheng, J. H. (2014). Transcriptome profiling of prostate tumor and matched normal samples by RNA-Seq. *European Review for Medical and Pharmacological Sciences*, *18*(9), 1354–1360.
- Zhu, M.-L., Horbinski, C., Garzotto, M., Qian, D. Z., Beer, T. M., & Kyprianou, N. (2008). Tubulin-Targeting Chemotherapy Impairs Androgen Receptor Activity in Prostate Cancer. *Bone*, *23*(1), 1–7. <https://doi.org/10.1158/0008-5472.CAN-10-0585.Tubulin-Targeting>