



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

FACULTAD DE CIENCIAS

**IDENTIFICACIÓN DE LNCRNAS COMO
BIOMARCADORES MOLECULARES DE
PRONÓSTICO EN CÁNCER DE PRÓSTATA
LOCALMENTE AVANZADO MEDIANTE
TRANSCRIPTÓMICA**

T E S I S

**QUE PARA OBTENER EL TÍTULO DE:
BIÓLOGA**

P R E S E N T A

ANA PAULA ALARCÓN ZENDEJAS



**DIRECTOR DE TESIS:
DR. CRISTIAN GABRIEL OLIVERIO
ARRIAGA CANON**

Ciudad Universitaria, CD.MX., 2022



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

DATOS DE LA ALUMNA Y DEL JURADO

1. Datos de la alumna

Alarcón

Zendejas

Ana Paula

Universidad Nacional Autónoma de México

Facultad de Ciencias

Biología

313595710

2. Tutor y secretario

Dr.

Cristian Gabriel Oliverio

Arriaga

Canon

3. Presidente

Dr.

Luis Alonso

Herrera

Montalvo

4. Vocal

Dra.

Karla Fabiola

Meza

Sosa

5. Suplente 1

Dr.

Luis David

Alcaraz

Peraza

6. Suplente 2

Dra.

Claudia Andrea

Segal

Kischinevzky

7. Datos del trabajo escrito

Identificación de lncRNAs como biomarcadores moleculares de pronóstico en cáncer de próstata localmente avanzado mediante transcriptómica.

Instituto Nacional de Cancerología (INCan)

160p

2021

AGRADECIMIENTOS ACADÉMICOS

Al Dr. Cristian Arriaga, mi tutor, por darme la oportunidad de hacer este proyecto; por guiarme y adentrarme en el mundo académico; por su paciencia y su visión; pero principalmente por sus palabras de apoyo, por su confianza y por creer en mí.

Al Dr. Luis Herrera, jefe del laboratorio de Carcinogénesis en donde se llevó a cabo este proyecto, y a la Mtra. Clementina Castro, técnico del mismo, por todo el apoyo y atención que me brindaron para poder realizar los experimentos correspondientes.

Al Dr. Miguel Ángel Jiménez, la Dra. Anna Scavuzzo y a todo el departamento de Urología del INCan, por su apoyo en la parte clínica y selección de muestras.

Al Dr. Luis Herrera, la Dra. Karla Meza, el Dr. Luis David Alcaraz y a la Dra. Claudia Segal, por haber formado parte de mi jurado y haber aportado un poco de su vasto conocimiento a mi proyecto.

Al MSc. Rogelio Montiel, QFB. Laura Contreras, QFB. Xiadani Guajardo, Biól. Sara del Río y al Mat. Augusto Pérez, por su apoyo en los análisis y experimentos necesarios.

Al Instituto Nacional de Cancerología, a la Facultad de Ciencias y a la Universidad Nacional Autónoma de México, por mi educación, mi formación profesional y el espacio para desarrollarme académica y profesionalmente.

AGRADECIMIENTOS PERSONALES

A Xía, Mich y Sara, por no sólo ser mis compañeras de lab, sino mis amigas; por todas las risas, errores, aprendizajes y también por esas tardes de pláticas, chismecito y comida rápida.

A mis mejores amigas y amigos de la Fac y de la Grasa: Eddie, por su apoyo y cariño incondicional siempre, a Samuel, Daphis, Catita, Colu, Ash, Pau grande, Hiram, Kevin, Felipe y Woog, por mantenerme viva durante la carrera, por las reuniones de tupperes en el Darwin entre clases, por darme los momentos más felices y divertidos desde 2015 y por ser una familia completita.

Al resto de mis mejores amigas y amigos: Andy, Raúl y Sofie, por estar a mi lado tantos años, por haber crecido conmigo y seguir aquí; por la motivación infinita, las inyecciones de energía, los mensajes de apoyo y el cariño fraternal. A Ana Mari por siempre resolver mis dudas y crisis existenciales.

Y, por último, a mi familia, mi pilar más grande y el más importante. A mi mami, mi papá, Ceci, mi abuela, mi abuelo, mi tía Marce, Macu, Lolita y Frida, sin ustedes no sería nada, y, sin embargo, gracias a ustedes, siento que lo soy todo. Teniéndolos a ustedes puedo decir que estoy completa. Nunca terminaré de agradecerles por mi educación y por permitirme llegar hasta acá, por nunca dejarme sola, por las lecciones de vida y las herramientas para vivirla, pero sobre todo por amarme incondicionalmente, yo también los y las amo hasta el fin del tiempo.

TABLA DE CONTENIDO

Abreviaturas y siglas	1
Resumen gráfico	4
Resumen / <i>abstract</i>	5
1. Introducción	7
1.1. Cáncer de próstata	
1.1.1. Anatomía y fisiología de la glándula prostática	
1.2. Epidemiología del cáncer de próstata	
1.2.1. Factores de riesgo	
1.3. Diagnóstico y pronóstico del cáncer de próstata	
1.3.1. Diagnóstico	
1.3.2. Estratificación y sistemas de gradación del cáncer de próstata	
1.3.3. Pronóstico	
1.4. Recurrencia bioquímica	
1.4.1. Biomarcadores moleculares de recurrencia bioquímica en cáncer de próstata	
1.5. RNA largos no codificantes	
1.5.1. Biogénesis de RNA largos no codificantes	
1.5.2. Clasificación y funciones de RNA largos no codificantes	
1.5.3. RNA largos no codificantes en cáncer	
1.5.4. El uso de lncRNA como biomarcadores moleculares en cáncer de próstata	
1.6. Transcriptómica enfocada a cáncer de próstata	
2. Justificación	32
3. Planteamiento del problema	33
4. Pregunta de investigación	34
5. Hipótesis	35
6. Objetivos	36
6.1. Objetivo general	
6.2. Objetivos particulares	

7. Metodología	37
7.1. Abordaje	
7.2. Fase de descubrimiento	
7.2.1. Criterios de inclusión de pacientes	
7.2.2. Purificación de RNA total	
7.2.3. Secuenciación masiva en paralelo de RNA (RNA-Sec)	
7.2.4. Análisis bioinformático de expresión diferencial	
7.2.4.1. Control de calidad de los datos de secuenciación	
7.2.4.2. Cambio de plataforma de secuenciación	
7.2.4.3. Alineamiento de secuencias	
7.2.4.4. Cuantificación de transcritos	
7.2.4.5. Análisis de expresión diferencial	
7.2.4.6. Selección de datos del análisis de expresión diferencial	
7.3. Fase de validación experimental	
7.3.1. <i>Fase de validación experimental A</i> : análisis de expresión diferencial de tejido tumoral vs. tejido adyacente	
7.3.1.1. Obtención de muestras y criterios de inclusión (cohorte independiente)	
7.3.1.2. Purificación, cuantificación y determinación de la integridad del RNA total	
7.3.1.3. Análisis de expresión diferencial	
7.3.1.3.1. Tratamiento con DNasa	
7.3.1.3.2. Síntesis de cDNA	
7.3.1.3.3. PCR en tiempo real para arreglo de expresión de lncRNA	
7.3.1.3.4. Análisis cuantitativo de expresión diferencial	
7.3.2. <i>Fase de validación experimental B</i> : análisis de expresión diferencial en líneas celulares de tejido prostático	
7.3.2.1. Purificación de RNA total de líneas celulares	
7.3.2.1.1. Cultivo de líneas celulares	
I. LNCaP	

II.	PrEC	
7.3.2.1.2.	Purificación de RNA total	
7.3.2.1.3.	Cuantificación de RNA	
7.3.2.2.	Cuantificación del lncRNA H19	
7.3.2.2.1.	Diseño de oligonucleótidos	
7.3.2.2.2.	Síntesis de cDNA	
I.	Tratamiento con DNasa	
II.	Tratamiento de transcriptasa reversa	
III.	Evaluación de la calidad de cDNA por PCR	
7.3.2.2.3.	Validación del método de cuantificación de transcritos por PCR en tiempo real	
I.	Determinación de la eficiencia de amplificación de los oligonucleótidos	
II.	Cuantificación de la expresión relativa del lncRNA mediante el método $\Delta\Delta Ct$	
7.3.3.	<i>Fase de validación experimental C</i> : análisis de expresión diferencial de muestras de pacientes con recurrencia bioquímica vs. pacientes sin recurrencia bioquímica	
7.3.3.1.	Selección de muestras y criterios de inclusión	
7.3.3.2.	Análisis de expresión diferencial	
7.3.4.	Análisis estadístico	
7.4.	Caracterización <i>in silico</i> de los lncRNA candidatos	
8.	Resultados	55
8.1.	Análisis bioinformático	
8.1.1.	Selección de muestras de pacientes con cáncer de próstata localmente avanzado	
8.1.2.	Integridad y calidad de RNA de las muestras para secuenciación	
8.1.3.	Selección de datos de RNA-Sec de acuerdo a los criterios de control de calidad para análisis bioinformáticos	
8.1.4.	Identificación del perfil de expresión diferencial de lncRNA	
8.2.	Validación experimental	

8.2.1. Análisis de integridad y calidad del RNA total para la fase de validación	
8.2.2. Determinación de un panel de lncRNA con expresión diferencial en muestras de pacientes con cáncer de próstata localmente avanzado	
8.2.3. Validación de la expresión del lncRNA H19 en líneas celulares de próstata por RT-qPCR	
8.2.4. Determinación de un panel de lncRNA asociado al pronóstico de recurrencia bioquímica en muestras de pacientes con cáncer de próstata localmente avanzado	
8.2.5. Análisis de conjuntos	
8.3. Caracterización <i>in silico</i> de los lncRNA candidatos como biomarcadores moleculares en cáncer de próstata localmente avanzado	
8.3.1. Características genéticas generales y determinación del potencial codificante de lncRNA	
8.3.2. Determinación de la expresión diferencial de los lncRNA en tejido prostático sano y tumoral basada en datos de acceso público	
8.3.3. Determinación de la localización y acumulación subcelular de los lncRNA analizados	
8.3.4. Determinación de la interacción entre los lncRNA caracterizados y otras moléculas	
9. Discusión	84
10. Conclusiones	91
11. Perspectivas	92
12. Referencias	93
Artículo de revisión	107

ABREVIATURAS Y SIGLAS

°C: grados Celsius.

APE: antígeno prostático específico (del inglés *PSA: Prostate Specific Antigen*).

AUA: Asociación Urológica Americana (del inglés *American Urological Association*).

AUC: área debajo de la curva (del inglés *Area Under the Curve*).

CaP: cáncer de próstata.

Cat.: catálogo.

cDNA: DNA complementario.

cm²: centímetros cuadrados.

Ct: del inglés *threshold cycle*.

CTC: células tumorales circulantes.

DNA: ácido desoxirribonucleico.

DRE: examen digital rectal (del inglés *Digital Rectal Exam*).

FDA: Administración de Medicamentos y Alimentos (del inglés *Food and Drug Administration*).

Fw: sentido (del inglés *forward*).

INCan: Instituto Nacional de Cancerología.

ISUP: Sociedad Internacional de Patologías Urológicas (del inglés *International Society of Urologic Pathology*).

Kb: kilobase.

lncRNA: RNA largo no codificante (del inglés *long non-coding RNA*).

mCRPC: cáncer de próstata metastásico resistente a castración.

min: minuto.

miRNA: microRNA.

mL: mililitro.

mM: milimolar.

mRNA: RNA mensajero.

ng: nanogramo.

NGS: secuenciación de siguiente generación (del inglés *Next-Generation Sequencing*).

nt: nucleótidos.

ORF: marco de lectura abierto (del inglés: *Open Reading Frame*).

pb: pares de bases.

PBS: amortiguador de fosfatos (del inglés *Phosphate-Buffered Saline*).

PCA: análisis de componentes principales (del inglés *Principal Component Analysis*).

PCR: Reacción en cadena de la polimerasa (del inglés *Polymerase Chain Reaction*).

Pol II: polimerasa II.

qPCR: Reacción en cadena de la polimerasa cuantitativa (del inglés *quantitative*).

RA: Receptor de andrógenos (del inglés AR: *Androgen Receptor*).

RIN: Número de integridad de RNA (del inglés *RNA integrity number*).

RNA: ácido ribonucleico.

RNA-Sec: secuenciación masiva en paralelo de RNA.

rpm: revoluciones por minuto.

rRNA: RNA ribosomal.

RT-PCR: Reacción en cadena de la polimerasa con reverso-transcriptasa.

Rv: antisentido (del inglés *reverse*).

s: segundo.

SFB: suero fetal bovino.

Tm: del inglés *melting temperature*.

tRNA: RNA de transferencia.

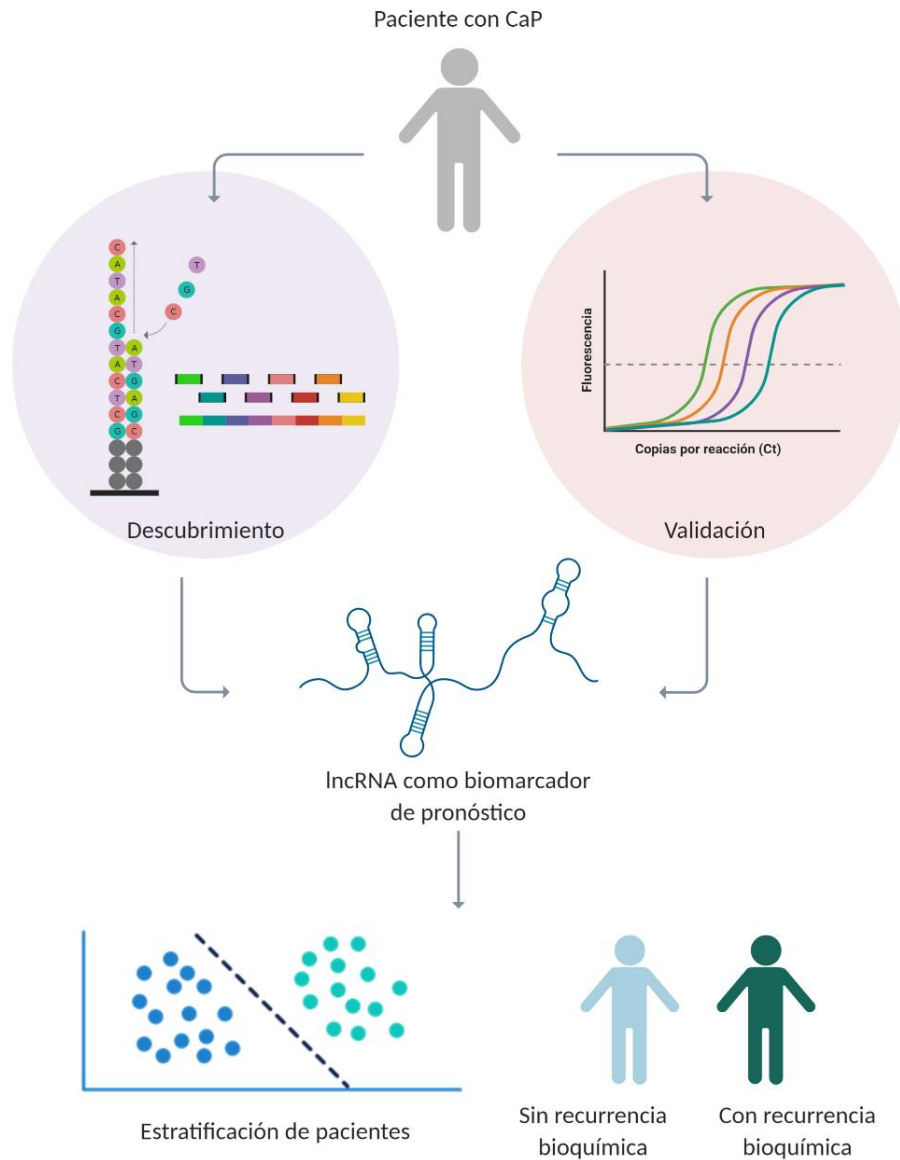
TRUS: ultrasonido transrectal (del inglés *transrectal ultrasound*).

U: unidad.

µg: microgramo.

µL: microlitro.

RESUMEN GRÁFICO



- A partir de muestras de pacientes con cáncer de próstata localmente avanzado de alto riesgo, se llevaron a cabo ensayos de descubrimiento y validación de lncRNA como candidatos para la búsqueda de biomarcadores moleculares de pronóstico.
- Mediante la técnica de RNA-Seq y ensayos de RT-qPCR, se encontró un panel de lncRNA con expresión diferencial que distingue al tejido tumoral del tejido adyacente, y un panel que distingue muestras de pacientes con recurrencia bioquímica de aquellas de pacientes que no recurrieron.

RESUMEN

El cáncer de próstata es un problema de salud pública en México y en el mundo, representando al segundo tumor maligno más común en hombres a nivel mundial. El 30% de los casos de cáncer de próstata que han sido tratados con prostatectomía radical presentan un relapso posterior a la cirugía y actualmente son muy pocos los biomarcadores moleculares de pronóstico que predigan la recidiva en el paciente oncológico, por lo que es necesario identificar nuevos marcadores moleculares involucrados en la recurrencia bioquímica del cáncer de próstata. Los lncRNA son transcritos mayores a 200 pares de bases que no tienen potencial codificante, y se ha determinado que presentan una expresión y actividad desregulada en varios tipos de cáncer, incluyendo al cáncer de próstata, por lo cual se han postulado como biomarcadores moleculares de utilidad clínica. Sin embargo, su estudio como biomarcadores de pronóstico en cáncer de próstata no se ha abordado a profundidad. En este proyecto se llevó a cabo un análisis de la expresión diferencial de lncRNA a partir de datos obtenidos por RNA-Seq de 3 muestras de tejido tumoral y 3 muestras de tejido adyacente de pacientes con cáncer de próstata localmente avanzado. Además, se realizó nuevamente un análisis de expresión diferencial en la fase de validación experimental mediante el arreglo de expresión de lncRNA: *RT² Cancer Pathway*, esta vez en muestras de 16 pacientes de una cohorte independiente comparando tejido tumoral y tejido adyacente. A partir de estos resultados se identificó un conjunto de lncRNA con expresión tumor-específica conformado por los lncRNA LNCPC1, H19, LNCPC2 y LNCPC3. Posteriormente se identificó un conjunto de lncRNA con expresión específica de recurrencia bioquímica mediante el arreglo de expresión de lncRNA *RT²* en donde se compararon 5 muestras de pacientes con recurrencia bioquímica y 11 sin recurrencia bioquímica. Este panel se conforma por los lncRNA LNCRBQ1, LNCRBQ2, LNCRBQ3 y LNCRBQ4. Por lo anterior, se concluye que este conjunto ayudará a determinar qué pacientes presentarán recurrencia bioquímica, por lo que podría ser un posible biomarcador de pronóstico en cáncer de próstata localmente avanzado.

ABSTRACT

Prostate cancer is a public health issue in Mexico and the rest of the world, representing the second most common malignant tumor in men worldwide. Studies have shown that 30% of prostate cancers treated with radical prostatectomy relapse into biochemical recurrence after surgery. Currently, few prognostic biomarkers predict biochemical recurrence in patients, so it is necessary to identify new molecular markers involved in prostate cancer recurrence. LncRNAs are RNA transcripts longer than 200 base pairs that are not translated into proteins and, recently, it has been determined that they have altered expression patterns and a deregulated activity in several cancers, including prostate cancer. Thus, lncRNAs are considered specific molecular biomarkers with high potential clinical utility in prostate cancer. Here we analyzed differential expression of lncRNAs in both tumor tissue and adjacent tissue samples from 3 patients with locally advanced prostate cancer by RNA-seq. Additionally, we validated the obtained results by differential expression analysis of 16 tumor tissue samples and 16 adjacent tissue samples obtained by radical prostatectomy using the lncRNA expression array kit: RT² Cancer Pathway. We identified 4-lncRNA panel with tumor-specific expression: LNCPC1, H19, LNCPC2 y LNCPC3. The use of the lncRNA expression array kit RT² also let us to determine a lncRNA panel specific for biochemical recurrence for which 5 recurrent patient tumor samples and 11 non-recurrent patient tumor samples were analyzed. In this molecular firm, the lncRNAs LNCRBQ1, LNCRBQ2, LNCRBQ3 and LNCRBQ4 are included, and thus, we propose them as potential prognostic biomarkers in the prediction of biochemical recurrence in locally advanced prostate cancer patients after radical prostatectomy.

1. INTRODUCCIÓN

1.1 Cáncer de próstata

El cáncer de próstata (CaP) se caracteriza por el crecimiento anormal de la glándula prostática causado por la división descontrolada de las células que la conforman¹ (Figura 1). Es uno de los tipos de cáncer con mayor prevalencia en hombres alrededor del mundo y se distingue por ser una enfermedad heterogénea que en muchos casos se comporta de manera indolora o indolente (como frecuentemente se le refiere en el área clínica), es decir, la enfermedad es asintomática y no provoca la muerte; sin embargo, es progresiva y letal^{2,3}. Cuando el CaP se detecta a tiempo, el tratamiento tiene un mejor efecto y la tasa de mortalidad disminuye. A pesar de ello y de las técnicas de detección utilizadas en la actualidad, como lo son la prueba del antígeno prostático específico (APE) y el examen digital rectal (DRE por sus siglas en inglés), el CaP sigue representando uno de los problemas de salud pública de mayor importancia en los pacientes afectados y actualmente no existe un método para detectar esta enfermedad de manera temprana^{1,3}. En particular, este tipo de neoplasia es detectada principalmente en varones mayores a los 60 años de edad y se estima que mientras la proporción de adultos mayores crezca, también lo harán la incidencia y los casos positivos de CaP^{2,4}.

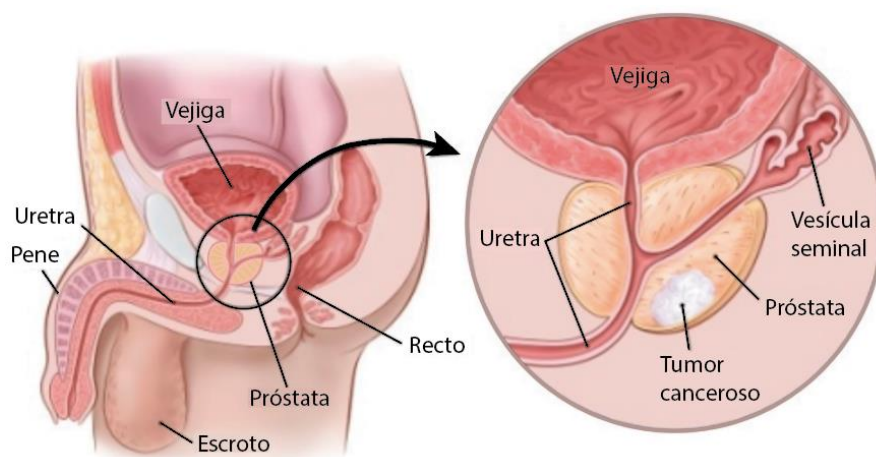


Figura 1. Anatomía de la próstata y tumor prostático. La próstata es una glándula sensible a la estimulación por andrógenos que crece conforme el adulto envejece. Se diferencia anatómica e histológicamente en tres regiones y los tumores se desarrollan mayoritariamente en la zona periférica. Modificada de la Sociedad Americana de Cáncer, 2019⁵.

1.1.1. Anatomía y fisiología de la glándula prostática

La próstata es una glándula exocrina sensible a andrógenos, de gran tamaño, que forma parte del aparato reproductor masculino⁶. Esta glándula se divide en tres zonas principales fusionadas: la zona transicional que abarca cerca del 5-10% del tejido glandular de la próstata y que rodea la parte de la uretra que cruza a través de ella; la zona central que ocupa entre el 20-25% de la glándula; y la zona periférica que ocupa alrededor del 70% del tejido (Figura 2A)⁶⁻⁸. Además, estas zonas no sólo se diferencian anatómicamente e histológicamente, sino que poseen una predisposición específica hacia una transformación benigna o maligna de células sanas a cancerosas. Por ejemplo, la hiperplasia prostática benigna se desarrolla, generalmente, en la zona transicional; mientras que alrededor del 60-75% del CaP surge en la zona periférica⁹⁻¹². Por otra parte, a nivel celular, la próstata presenta un epitelio pseudoestratificado con tres tipos de células diferenciadas: luminales, basales y neuroendócrinas (Figura 2B)¹³⁻¹⁵. A pesar de que no se sabe con exactitud cuál es el tipo celular que origina el cáncer, se ha observado que la patología se desarrolla a partir de células de tipo basal así como también de las luminales¹⁶⁻¹⁸.

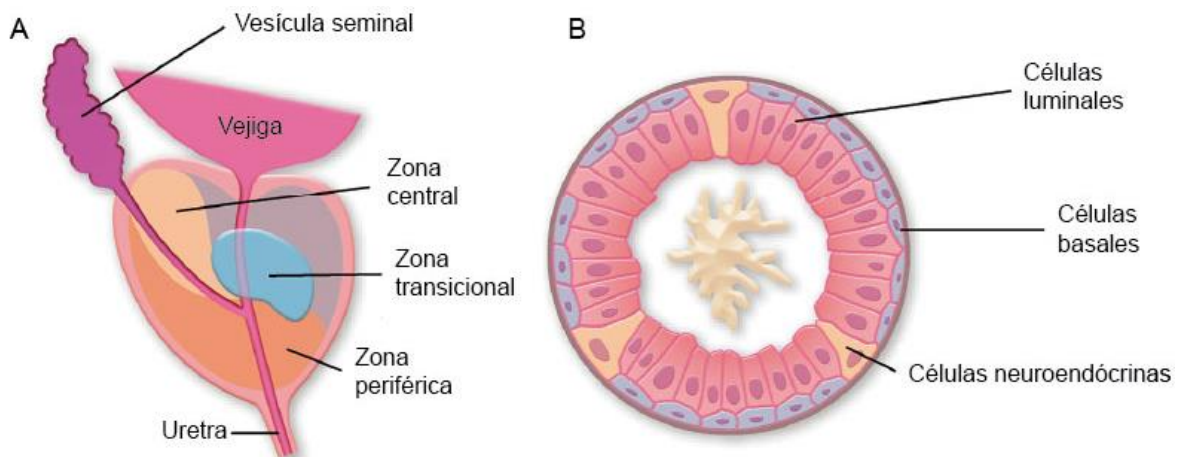


Figura 2. Anatomía y arquitectura celular de la glándula prostática humana. A) La anatomía zonal divide a la próstata en una zona fibromuscular anterior y en tres zonas glandulares: la zona central rodea los ductos eyaculatorios, la zona transicional rodea a la uretra, y la zona periférica rodea a ambos. **B)** Composición celular del epitelio prostático. A nivel molecular, el epitelio basal expresa citoqueratina 5 y el factor transcripcional p63; el epitelio luminal expresa citoqueratina 8 y KLK3; y las células neuroendócrinas expresan cromogranina A¹⁹. Modificada de Wang *et al.*, 2018³.

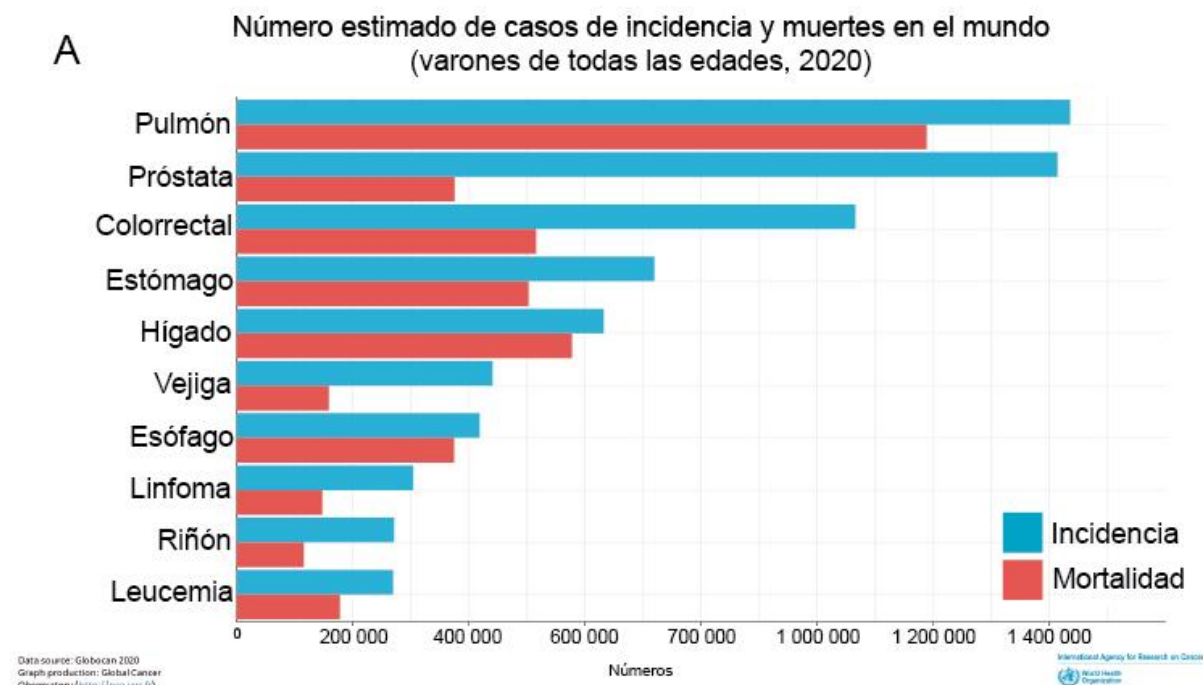
La principal función de la próstata es la formación del líquido seminal mediante la producción de una secreción ácida (pH ~6) que contiene zinc, ácido cítrico, poliamidas y enzimas proteolíticas, como el APE y la fosfatasa ácida (PAP por sus siglas en inglés). La fosfatasa ácida se encarga de la licuefacción del coágulo seminal para que éste se mezcle con el esperma y forme el semen⁶. Por otra parte, el APE (o PSA por sus siglas en inglés) es una glicoproteína enzimática cuya expresión es regulada por andrógenos. El APE se encarga de romper la semenogelina I y II presentes en el coágulo seminal, promoviendo la motilidad de los espermatozoides y facilitando su llegada al óvulo²⁰. Además, cerca del 75% del APE se encuentra unido a proteínas del plasma y es metabolizado en el hígado, mientras que el 25% restante es excretado en la orina⁶. El APE sérico circula de ambas formas, como APE libre y unido a proteínas, sin embargo, la prueba dirigida a este marcador mide únicamente la concentración del APE total, es decir, la suma total del APE libre y en complejo con otras proteínas²¹. Esta enzima ha resultado ser de gran importancia clínica en el área de detección y monitoreo de CaP debido a que la presencia de la enfermedad se asocia a una disminución en la producción del antígeno prostático. Sin embargo, durante estadios tempranos del CaP, se presenta la ruptura de la capa de células basales y la membrana basal, permitiendo el acceso directo de la glicoproteína a la circulación periférica y evitando que llegue al lumen de la glándula, provocando que se eleven significativamente sus niveles en sangre^{22,23}, dicho de otro modo, un valor alto de APE se asocia a un riesgo elevado de CaP.

1.2. Epidemiología del cáncer de próstata

A nivel mundial, el CaP es el tumor maligno más común en hombres después del cáncer de pulmón y ocupa el quinto lugar dentro de las principales causas de muerte por cáncer en varones. En 2020 se registraron 1,414,259 casos nuevos en el mundo, representando al 7.3% en incidencia de todos los tipos de cáncer y al 3.8% del total de muertes con 375,304 casos (Figura 3A)²⁴. En el mismo estudio de GLOBOCAN, se reportaron las tasas con mayor incidencia en Oceanía, Estados

Unidos, Alaska y Europa; a diferencia de Asia, África y gran parte de Sudamérica y Centroamérica en donde se presentaron tasas bajas en comparación con países desarrollados como los mencionados anteriormente²⁵. Asimismo, mientras que las mayores tasas de mortalidad fueron reportadas en América Central, Australia y Nueva Zelanda, las tasas más bajas fueron reportadas en Asia, el norte de África y Europa²⁵. Específicamente en México, el CaP representa el 29.9% de casos nuevos con 26,742 casos registrados, y el 16.9% de muertes por cáncer con 7,457 casos registrados (Figura 3B)²⁶. En particular, en el Instituto Nacional de Cancerología (INCan) se reciben alrededor de 500 pacientes diagnosticados con CaP anualmente, de los cuales el 37% se encuentra en etapas avanzadas y el 32% presenta metástasis al momento del diagnóstico (datos internos del INCan).

Consecuentemente, esta cifra ha aumentado a lo largo de los últimos años, por lo que el CaP representa un problema de salud pública en nuestro país y el mundo.



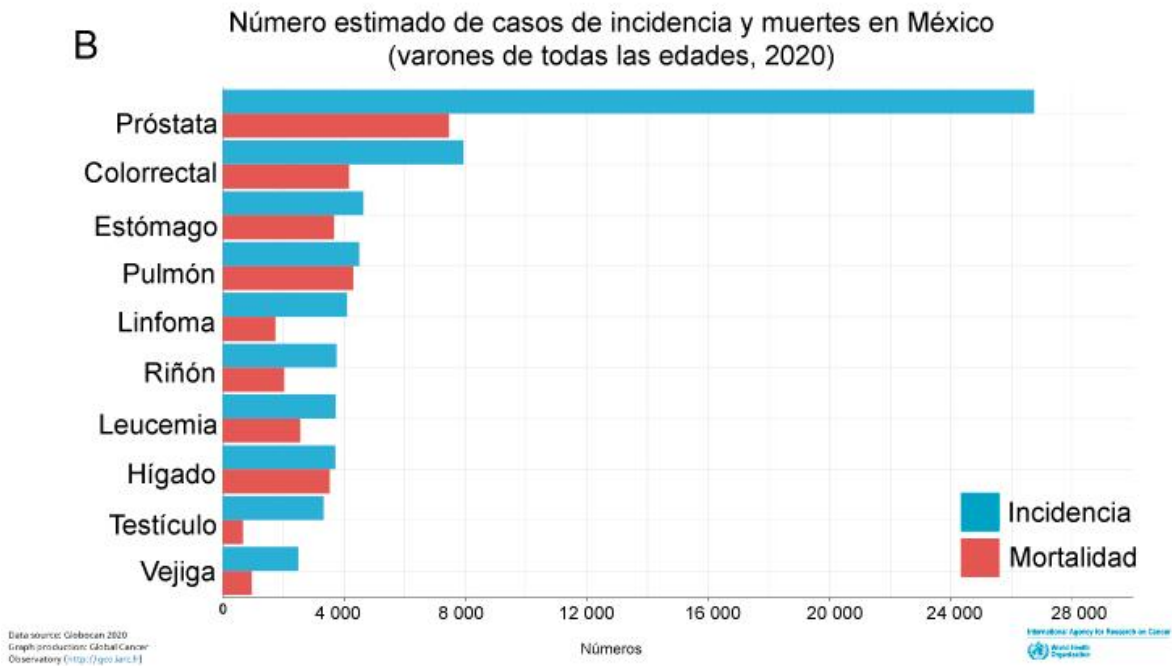


Figura 3. El cáncer de próstata es un problema de salud pública mundial y nacional. A) Gráfico de barras que representa los índices de incidencia y mortalidad de los diferentes tipos de cáncer en la población de hombres de todas las edades a nivel mundial. **B)** Gráfico de barras representativo de los índices de incidencia y mortalidad de los diferentes tipos de cáncer en la población de hombres de todas las edades en México. Modificada de *Cancer Today*²⁶.

1.2.1. Factores de riesgo

El desarrollo del CaP puede verse favorecido por algunos factores presentes en los pacientes que aumentan el riesgo de padecer esta enfermedad; los tres factores de riesgo confirmados para CaP son: la edad, la etnicidad y el historial familiar de CaP²⁷.

En primera instancia, la edad es el principal factor de riesgo en CaP, ya que la tasa de incidencia es cercana al 60% en hombres mayores a 65 años y ésta aumenta conforme el individuo envejece²⁸. Específicamente, se ha comprobado que el riesgo aumenta a partir de los 50 años en hombres caucásicos que no tienen antecedentes familiares de CaP, y después de los 40 años en hombres con ascendencia africana y antecedentes familiares de CaP²⁹; en otros casos, la probabilidad de que un hombre menor a 40 años presente CaP, es baja⁴. Como

segundo factor de riesgo a considerar se encuentra el grupo étnico al que el individuo pertenece. La población de hombres con ascendencia africana presenta la mayor incidencia y mortalidad por CaP en el mundo, además de tener una mayor probabilidad de desarrollar esta enfermedad en edades tempranas³⁰. Lo anterior se debe a que esta etnia posee una predisposición genética en donde genes, como *BRCA2*, son más susceptibles a mutar y sus variantes se asocian al alto riesgo de desarrollar CaP³¹. Por último, los antecedentes familiares de CaP son el tercer factor de riesgo a considerar. Se ha estimado que cerca del 20% de los pacientes con CaP tiene un historial familiar con presencia de la enfermedad. En estos casos, el cáncer pudo haberse generado por factores genéticos o por hábitos de vida similares a los de los familiares con CaP de los pacientes³²; incluso se ha comprobado que al tener algún familiar de primer grado con CaP, el riesgo a desarrollar cáncer aumenta al doble o triple³³. Gracias a estudios recientes, se han reconocido loci asociados a CaP; dentro de ellos se encuentra el locus 1q24-25 o el gen *HPC1* que codifica a la enzima ribonucleasa L (RNASEL) la cual participa en la regulación de la muerte celular por apoptosis³⁴. También el gen *HPC2* presente en el cromosoma 17p11 que codifica a la proteína ELAC2, la cual se une a SMAD2 promoviendo la proliferación celular y a su vez el desarrollo de CaP³⁵. De igual manera, el gen supresor de tumores *BRCA2* y las mutaciones en éste están asociadas a una mayor incidencia de CaP³⁶. Existen otros factores menos comunes que también están asociados al desarrollo o muerte por CaP, sin embargo, no están completamente comprobados. Dentro de ellos se encuentran el consumo de alimentos con altos niveles de grasa animal y carne roja, obesidad, diabetes y el hábito de fumar²⁷. En conjunto, estos tres factores han favorecido la prevención y la detección temprana del CaP.

1.3. Diagnóstico y pronóstico del cáncer de próstata

1.3.1. Diagnóstico

La detección temprana del CaP es un paso elemental para la disminución en las tasas de incidencia y mortalidad. Actualmente existen pruebas de rutina llevadas a cabo para la detección temprana de esta enfermedad. En las últimas décadas la

incidencia del CaP ha disminuido gracias al DRE y a la cuantificación de los niveles del biomarcador APE en sangre, seguida de una biopsia guiada por ultrasonido³⁷. A pesar de que estas pruebas han favorecido la detección del CaP, se han identificado algunas discrepancias en su eficacia, ya que pueden generar resultados inconsistentes y, por consecuencia, en muchos casos no se distingue el CaP indolente de las etapas agresivas del mismo³⁸.

La prueba del DRE se utiliza de manera sistemática para evaluar la disfunción miccional en varones, sin embargo, la detección de CaP a partir del DRE no es completamente efectiva ya que la mayoría de pacientes con CaP detectado por APE no presentan anomalías en el DRE³⁹. Esta prueba digital consiste en la exploración de la próstata y órganos adyacentes con el fin de encontrar anomalías morfológicas. En el caso de los hombres, se determinará si el tamaño de la próstata es normal y si hay presencia de nódulos o algún otro signo que salga fuera de los parámetros comunes.

Como se mencionó anteriormente, la detección del APE ha permitido identificar el CaP en etapas tempranas mediante radioinmunoanálisis en suero e inmunohistoquímica en secciones histológicas, considerándose con riesgo a cáncer a partir de los 4.0 ng/mL⁶. Sin embargo, esta prueba limita el diagnóstico de CaP, ya que su baja especificidad provoca la detección de falsos positivos y esto a su vez conlleva a un sobretatamiento de pacientes³⁸. Incluso se ha demostrado que cerca del 20-40% de los casos de CaP en Estados Unidos y Europa podrían deberse a un sobrediagnóstico por la prueba del APE⁴⁰. Un factor relevante sobre este biomarcador es que su detección es órgano-específica y no cáncer-específica, por lo que es factible que se dé un diagnóstico erróneo, es decir, que no se distinga entre CaP y una patología benigna como hiperplasia prostática benigna o prostatitis⁴¹. De acuerdo con lo reportado por la Sociedad Americana del Cáncer, la prueba de detección del APE ha presentado un valor de sensibilidad (parámetro que mide la proporción de casos positivos verdaderos, es decir, se detectan a los individuos con la enfermedad) y especificidad (parámetro que mide la proporción de casos negativos verdaderos, es decir, detecta a los individuos sanos) de 21% y

91%, respectivamente, al mantenerse el umbral de 4 ng/mL^{42,43}. El ultrasonido transrectal (TRUS por sus siglas en inglés) es una técnica complementaria llevada a cabo durante el diagnóstico de CaP en pacientes con niveles altos de APE mayores a 4.0 ng/mL¹. A partir del TRUS se obtienen imágenes de las áreas que serán seleccionadas para la toma de biopsias, sin embargo, si el paciente no ha desarrollado CaP a pesar de haber presentado concentraciones altas de APE, será sometido a biopsias innecesarias³⁷. Es por ello y por lo mencionado con anterioridad, que se requiere crear mejores estrategias para la detección y pronóstico del CaP.

1.3.2. Estratificación y sistemas de gradación del cáncer de próstata

Identificar la etapa en la que se encuentra el CaP es crucial para la elección del tratamiento que se le dará al paciente. Mediante la estratificación del CaP se puede describir en dónde se encuentra localizado el tumor, si se ha extendido a otros tejidos y si está afectando a otros órganos del cuerpo. La estratificación de adenocarcinomas de CaP se basa en el sistema de clasificación TNM, así como el sistema de gradación de Gleason y las pruebas de APE y DRE⁴⁴.

El sistema de clasificación TNM (refiriéndose a T: Tumor, N: Nudo y M: Metástasis) describe el tamaño del tumor y su localización; si éste se ha extendido hasta nodos linfáticos; y si el cáncer ha invadido otras partes del cuerpo, así como la magnitud de la invasión (metástasis). A estas tres categorías se les asigna un número entre el 0 y el 4 según la extensión del cáncer: T1, T2, T3 y T4; N0, N1, N2 y N3; y M0 o M1^{45,46}. El puntaje de Gleason se emplea para medir el grado de agresividad del CaP. Este sistema se basa en identificar los dos patrones histológicos más comunes en las células de la biopsia tumoral. Según el nivel de irregularidad observado se les asigna un número del 1 al 5; entre menor sea el puntaje, la arquitectura celular se asemeja más a la de una glándula sana. La sociedad Internacional de Patologías Urológicas (ISUP por sus siglas en inglés) propone nueve posibles puntajes dentro de 5 grupos: **1** (3+3), **2** (3+4), **3** (4+3), **4** (3+5, 5+3, 4+4), y **5** (4+5, 5+4, 5+5)⁴⁷, en donde una puntuación total de 7

corresponde a un cáncer de grado medio y una puntuación total de 8-10 significa un cáncer de grado alto (Figura 4).

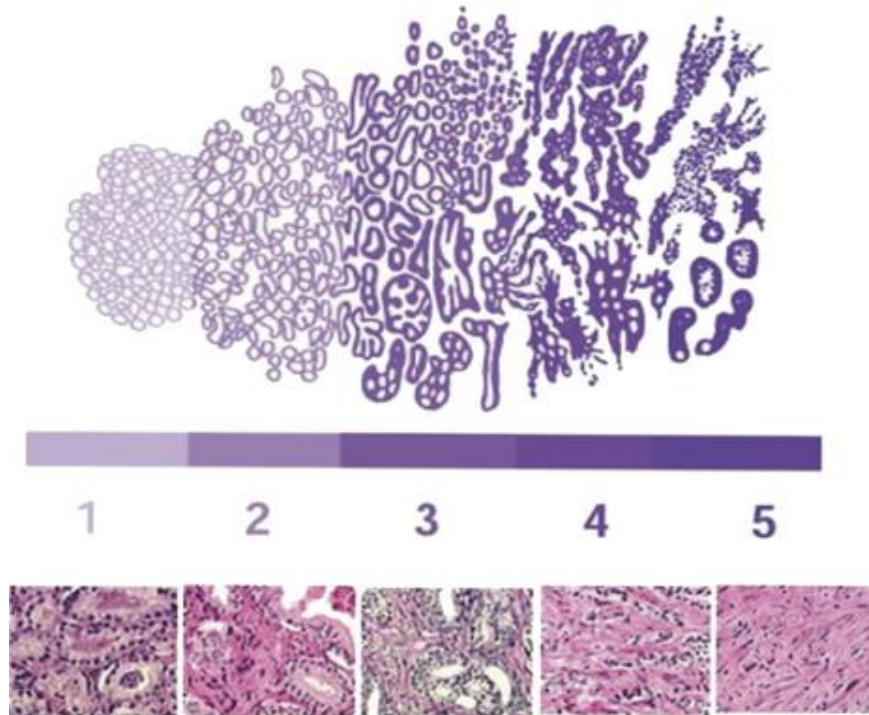


Figura 4. El sistema de gradación de Gleason se basa en la arquitectura celular del tejido prostático. Patrones histológicos correspondientes al sistema de Gleason adaptado por el consenso de 2015 por la ISUP. Tomada de Harnden *et al.*, 2007⁴⁸.

1.3.3. Pronóstico

Actualmente, las características clínico-patológicas son las principalmente utilizadas como factores de pronóstico de CaP. Dentro de ellas, la edad del paciente, las características morfológicas y patológicas del tumor, la concentración del APE en sangre, la escala de Gleason, la extensión extracapsular del tumor, el estadio clínico del CaP y la invasión de tejidos cercanos, entre otras, son tomadas en cuenta para la predicción de supervivencia global y la probabilidad de progresión o recurrencia de la enfermedad⁴⁹. A pesar de la gran utilidad que pueden tener los factores clínico-patológicos mencionados anteriormente, gran parte de la heterogeneidad del CaP se debe principalmente a factores genéticos y moleculares, es por ello que se han propuesto firmas génicas, genes individuales y moléculas circulantes obtenidas a

partir de biopsias líquidas como potenciales biomarcadores moleculares de pronóstico para CaP. Lo anterior, con el objetivo de ser detectadas principalmente en pacientes que hayan sido tratados con radioterapia o en pacientes que hayan sido sometidos a prostatectomía radical⁴⁹. Dentro de los biomarcadores moleculares de pronóstico se encuentran el receptor de andrógenos (RA o AR por sus siglas en inglés)⁵⁰, la fusión génica TMPRSS2:ERG como pronóstico de progresión^{51,52}, y la pérdida de PTEN para distinguir tumores potencialmente agresivos⁵³. También las firmas génicas de *Oncotype DX Genomic Prostate Score* para la predicción de agresividad en la enfermedad, muerte por CaP y metástasis en 10 años (AUC=0.67)⁵⁴; *Prolaris* asociado a mortalidad y recurrencia bioquímica^{55,56}; y *Decipher Prostate Cancer Test* para predecir metástasis en pacientes sometidos a prostatectomía radical (AUC=0.81)^{57,58}. Todas ellas se basan en paneles de genes asociados a cáncer y se utilizan actualmente en la clínica como biomarcadores moleculares de pronóstico^{59,60}.

Las herramientas clínicas mencionadas anteriormente guían al urólogo para establecer el tratamiento apropiado para pacientes con CaP, sin embargo, es muy importante continuar la búsqueda de marcadores moleculares que permitan predecir el riesgo de recurrencia bioquímica y progresión después del tratamiento localizado.

1.4. Recurrencia bioquímica

Muchos de los pacientes con CaP localizado que han sido sometidos a terapia por cirugía o radioterapia presentan una recidiva, es decir, muestran una concentración de APE sérico detectable y constante (o creciente, en caso del tratamiento con radioterapia)⁶¹. Los valores de APE se miden cada 3-6 meses después de haber sido sometidos a prostatectomía radical⁶²; y, de acuerdo con la Asociación Urológica Americana (AUA), se definen como “recurrentes” a los casos de cáncer que después de la cirugía presentan valores de APE a partir de los 0.2 ng/mL, también conocido como APE *nadir*⁶³. Sin embargo, no todos los pacientes con niveles altos de APE tendrán un relapso, y es difícil predecir quiénes lo desarrollarán y la velocidad con

la que progresará; por lo que es necesario implementar ensayos clínicos complementarios al valor absoluto de APE para definir la recurrencia bioquímica⁶². La probabilidad de recurrencia bioquímica posterior a la prostatectomía radical depende de factores como: la etapa del cáncer, el grado de extensión extracapsular, los márgenes quirúrgicos positivos y la invasión de vesículas seminales, entre otros. Se sabe que aproximadamente el 30% de los pacientes con CaP que fueron sometidos a prostatectomía radical tendrán un relapso, y que el 40% de ellos desarrollarán metástasis⁶⁴.

Debido a que ningún factor clínico-patológico puede predecir con exactitud el riesgo de que un paciente presente recurrencia bioquímica, se requiere implementar la búsqueda de marcadores moleculares de mayor precisión que mejoren el manejo clínico del paciente oncológico.

1.4.1. Biomarcadores moleculares de recurrencia bioquímica en cáncer de próstata

Los factores clínicos tradicionales que se toman en cuenta para el pronóstico del CaP se han asociado a la progresión del cáncer, pero ninguno de estos ha sido validado como un marcador pronóstico de recurrencia bioquímica⁶⁵. Hasta la fecha, la predicción de la recurrencia se basa en la prueba del APE y el puntaje de Gleason medido en biopsias, sin embargo, algunos estudios han determinado que estos parámetros no son adecuados para predecir recurrencia⁶⁶. También se ha implementado la detección de niveles altos del RA junto con co-factores y co-activadores asociados a él, como LSD1, FLH2 y TIF2^{67,68}. Asimismo, se ha propuesto a las células tumorales circulantes (CTC) como predictoras de recurrencia y progresión del CaP⁶⁹.

De igual manera, se han estudiado las alteraciones a nivel genómico, transcriptómico y proteómico que presentan las células tumorales en CaP. Dentro de ellas, se han propuesto las alteraciones en los niveles de expresión de transcritos no codificantes como posibles biomarcadores de recurrencia bioquímica en CaP⁷⁰.

Específicamente el biotipo conformado por los RNA largos no codificantes (lncRNA) que tienen un gran potencial como biomarcadores pronósticos debido a su expresión tejido-específica y a su papel en la regulación celular, pero son pocos los que se aplican en el área clínica. Gracias a diferentes estudios se ha encontrado un grupo de lncRNA asociados a CaP, dentro de ellos se encuentra *Prostate Cancer gene 3* (PCA3), el cual es un biomarcador de naturaleza no codificante y cuya sobreexpresión es específica de tumores prostáticos, por lo que es un biomarcador de utilidad clínica que se utiliza actualmente para la detección y diagnóstico de CaP, específicamente para decidir si un paciente con una biopsia negativa previa requiere de una segunda biopsia⁷¹. Muchos otros lncRNA han sido evaluados como predictores de recurrencia bioquímica en CaP, como es el caso de *Second Chromosome Locus Associated with Prostate-1* (SChLAP1), también conocido como LINC00913 o PCAT114, el cual es un RNA de biotipo no codificante con sobreexpresión específica de CaP y se ha reportado que promueve la invasión tumoral y la metástasis, además de que su detección puede predecir recurrencia bioquímica, así como la progresión a metástasis⁷².

Una de las particularidades de los lncRNA que ha llamado la atención es que presentan perfiles de expresión anormales en los distintos tipos de cáncer, por lo que se les ha vinculado directamente con el desarrollo y la patogénesis del mismo, asociándoseles con funciones oncogénicas, así como con funciones de supresor de tumores en cáncer. En la lista general de marcadores moleculares para cáncer se han incluido a varios lncRNA como HOTAIR⁷³, PCAT1⁷⁴, MALAT1⁷⁵ y PCA3⁷⁶, entre otros, y el número ha ido en aumento en los últimos años, lo que ha establecido la importancia de estudiar los lncRNA y su potencial como biomarcadores moleculares de pronóstico en patologías como el cáncer, así como de su aplicación clínica para la mejora en el manejo del paciente oncológico.

1.5. RNA largos no codificantes

De acuerdo con Hanguager y colaboradores⁷⁷, el 80% del genoma humano se transcribe, pero únicamente el 3% corresponde a genes codificantes de proteínas, es decir, el 77% corresponde a genes no codificantes; dentro de los cuales se encuentran los lncRNA⁷⁸. Los lncRNA se caracterizan por ser transcritos con una longitud mayor a 200 nucleótidos que no cuentan con potencial codificante; es decir, no presentan un marco de lectura abierto (ORF por sus siglas en inglés)⁷⁹ y por lo tanto, no dan origen a la síntesis de proteínas. Estas dos características principales los distinguen de los RNA mensajeros (mRNA) y del resto de los RNA no codificantes (microRNA, tRNA, rRNA, entre otros). Por otro lado, una de las principales características de los lncRNA es que se sintetizan a bajas tasas en comparación a un mRNA, su localización es mayoritariamente hacia el núcleo celular y presentan patrones diferenciales de expresión en distintos tejidos, estableciendo que este tipo de RNA se expresa de manera tejido-específica⁸⁰. Además, se ha determinado que los lncRNA pueden estar involucrados en distintos procesos biológicos como la metilación del DNA⁸¹, la remodelación de la estructura de la cromatina⁸², la formación de sub-compartimentos nucleares⁸³, la impronta génica^{84,85}, la regulación de la expresión génica^{82,85} y la regulación postranscripcional de mRNA^{65,66}, manifestando la gran importancia funcional de este tipo de transcritos en la regulación de la expresión genética.

En particular, los niveles de expresión desregulados de los lncRNA se han visto asociados a enfermedades como el cáncer. Algunos de los procesos patológicos en los que se han visto involucrados son: la proliferación celular, la tumorigénesis, la transición epitelio-mesénquima y la transformación de células sanas a cancerosas⁸⁷, es por ello que muchos estudios los han propuesto tanto como genes supresores de tumores así como oncogenes^{88,89}. Recientemente, los lncRNA han llamado la atención de la comunidad científica no sólo por su importancia en el desarrollo de cáncer sino por su potencial uso como biomarcadores moleculares debido a su expresión diferencial tejido- y tumor-específica^{90,87}. Además, los lncRNA son moléculas de RNA muy estables ya que pueden circular por fluidos corporales como la sangre, la saliva y la orina, mediante

microvesículas (exosomas y micropartículas), asociados a lipoproteínas de alta densidad o asociados a complejos proteicos⁹¹. Esto los convierte en una herramienta de diagnóstico molecular útil en el área clínica.

1.5.1. Biogénesis de RNA largos no codificantes

La biogénesis de los lncRNA es muy similar a la de un mRNA, ya que son sintetizados por la RNA polimerasa II (Pol II) y se encuentran en loci genómicos con marcas postraduccionales de cromatina similares a los de transcritos codificantes como la H3K4me3. En su mayoría, presentan la adición de la estructura denominada caperuza o “CAP” en el extremo 5', y pueden estar poliadenilados en su extremo 3', así como sufrir procesamiento (*splicing*) y presentar varias isoformas⁹². Visto desde el aspecto bioquímico, no existe alguna diferencia entre un transcrito de lncRNA y uno de mRNA, únicamente que los lncRNA no presentan un ORF dentro de su secuencia de nucleótidos⁹³ (Figura 5). Los lncRNA presentan sitios específicos de *splicing*, ya que tienen exones de longitud similar a la de los mRNA pero en una menor cantidad (en su mayoría presentan únicamente dos exones con una longitud promedio de 149 pb) y se encuentran principalmente enriquecidos en el núcleo celular en la fracción de nucleoplasma y cromatina⁸⁰.

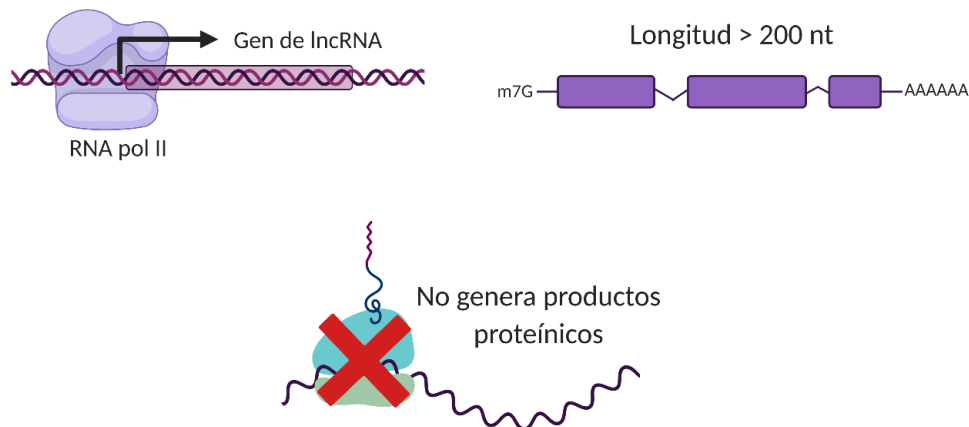


Figura 5. Biogénesis y características principales de los lncRNA. La definición de los lncRNA sigue modificándose, sin embargo, su esquema de clasificación universal se basa en que, a diferencia de los mRNA, los lncRNA no generan proteínas y su longitud sobrepasa los 200 nucleótidos. En general, los lncRNA son sintetizados por la RNA polimerasa II al igual que los mRNA y, en su mayoría, presentan las mismas modificaciones post-transcripcionales que un mRNA. Imagen generada en Biorender.com y modificada de Sun y Kraus, 2015⁸⁴.

1.5.2. Clasificación y funciones de RNA largos no codificantes

Debido a que existe una gran proporción de regiones no codificantes dentro del genoma humano (77%), el grupo de los lncRNA puede clasificarse de acuerdo a su localización genómica respecto a genes codificantes adyacentes (Figura 6).

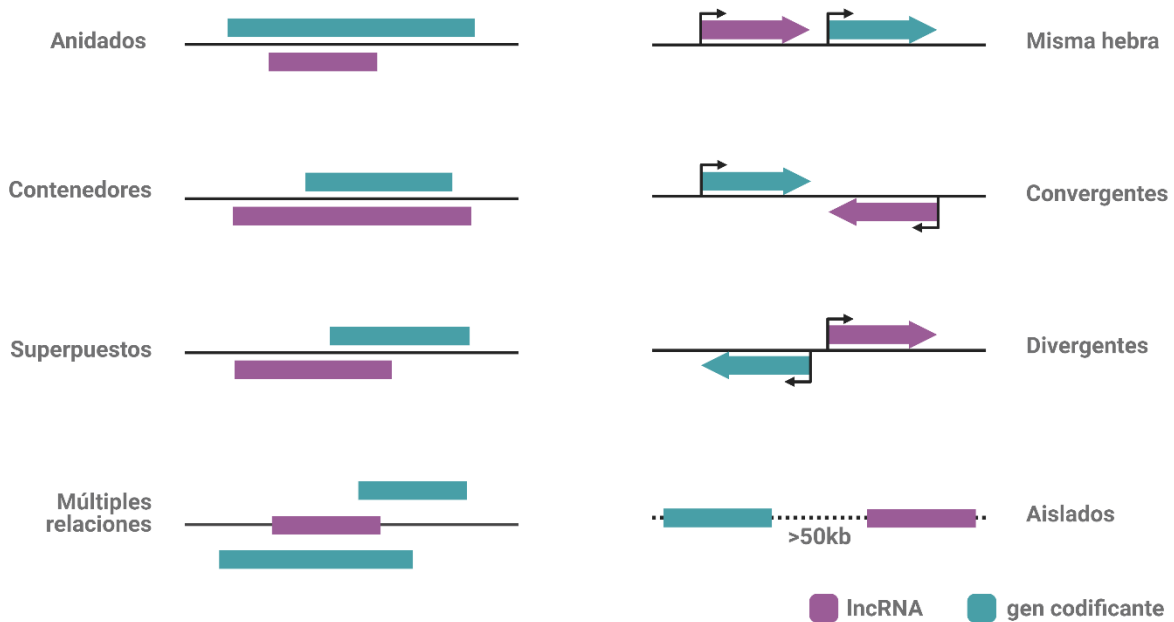


Figura 6. Clasificación de lncRNA de acuerdo a su localización. Clasificación de lncRNA basada en su posición genómica y el sentido en el que se transcriben con respecto a los genes codificantes adyacentes o cercanos a ellos. Modificada de Ransohoff *et al.*,2018⁹⁴. Imagen generada en Biorender.com.

A todos aquellos lncRNA localizados y sintetizados a partir de regiones encontradas entre dos genes codificantes, se les conoce como lncRNA intergénicos (lincRNA), por el contrario, aquellos lncRNA que son sintetizados a partir de la región intrónica de un gen codificante se conocen como lncRNA intrónicos⁹⁵. Otro de los factores tomados en cuenta para su clasificación, es la hebra desde la que se sintetizan, es decir, si el lncRNA se sintetiza desde la hebra sentido o si se transcribe a partir de la hebra antisentido. Los lncRNA sentido se sintetizan desde la misma hebra que los genes codificantes adyacentes a él, e, incluso, pueden llegar a empalmarse. Por el contrario, los lncRNA antisentido se transcriben desde la hebra opuesta a la de los genes codificantes cercanos⁹⁵. Sin embargo, la

clasificación que se usa puede variar de acuerdo al autor; Ransohoff y colaboradores⁹⁴ los clasifican dentro de los biotipos según su locus descritos en la tabla 1.

Tabla 1. Clasificación de lncRNA propuesta por Ransohoff y colaboradores⁹⁴.

lncRNA génicos	lncRNA que se empalman en uno o más nucleótidos sobre un gen codificante de proteínas.
Anidados	Genes de lncRNA contenidos en su totalidad dentro de un gen codificante. Sentido o antisentido.
Contenedores	Genes codificantes contenidos en su totalidad dentro de un lncRNA. Sentido o antisentido.
Superpuestos	Genes de lncRNA que no son clasificados como anidados ni contenedores. Sentido o antisentido.
De múltiples relaciones	lncRNA con más de una relación de las mencionadas anteriormente.
lncRNA intergénicos (lincRNA)	lncRNA que no se empalman con genes codificantes de proteínas.
Misma hebra	LincRNA transcritos desde la misma hebra y en la misma dirección que el gen codificante más cercano dentro de 50 Kb.
Convergentes	LincRNA transcritos de “cabeza a cabeza” con el gen codificante más cercano dentro de 50 Kb.
Divergentes	LincRNA transcritos de “cola a cola” con el gen codificante más cercano dentro de 50 Kb.
Aislados	LincRNA alejados por más de 50 Kb del gen codificante más cercano.

Los estudios enfocados en los mecanismos funcionales de los lncRNA han determinado que éstos juegan un papel importante en la regulación génica positiva y negativa en diversos procesos celulares. Esta regulación puede llevarse a cabo a nivel transcripcional y postranscripcional tanto de genes adyacentes (actuando en *cis*) como lejanos (actuando en *trans*) mediante el reclutamiento de proteínas, como factores transcripcionales y complejos remodeladores de la cromatina⁹⁶. También pueden actuar como andamios para facilitar la formación de complejos multipéptidos; como guía de moléculas reguladoras hacia sus genes blanco; como señuelos (del inglés *decoys*) para la captura de miRNA y proteínas; y también como

*enhancers*⁸⁵ (Figura 7). Asimismo, se han visto involucrados en procesos de *splicing* y de regulación traduccional del mRNA^{85,93}. Evidencias experimentales han establecido que la función de los lncRNA está relacionada a su localización dentro de la célula; los lncRNA que tienen funciones en el núcleo pueden encontrarse en el nucleoplasma como *Metastasis Associated Lung Adenocarcinoma Transcript 1* (MALAT1) o bien, en la cromatina como *X Inactive Specific Transcript* (XIST); por otra parte, están aquellos lncRNA que deben ser exportados al citoplasma para llevar a cabo sus funciones como los lncRNA *Differentiation Antagonizing Non-Protein Coding RNA* (DANCR) y *H19 Imprinted Maternally Expressed Transcript* (H19)⁹⁶. Debido a que la expresión de los lncRNA es específica de tejido y de tumor, así como a los procesos celulares en los que los lncRNA se ven involucrados, a los cambios en su expresión, a la regulación y a sus mecanismos de acción, es que pueden afectar la homeostasis celular, llegando incluso a favorecer el desarrollo de patologías como el cáncer.

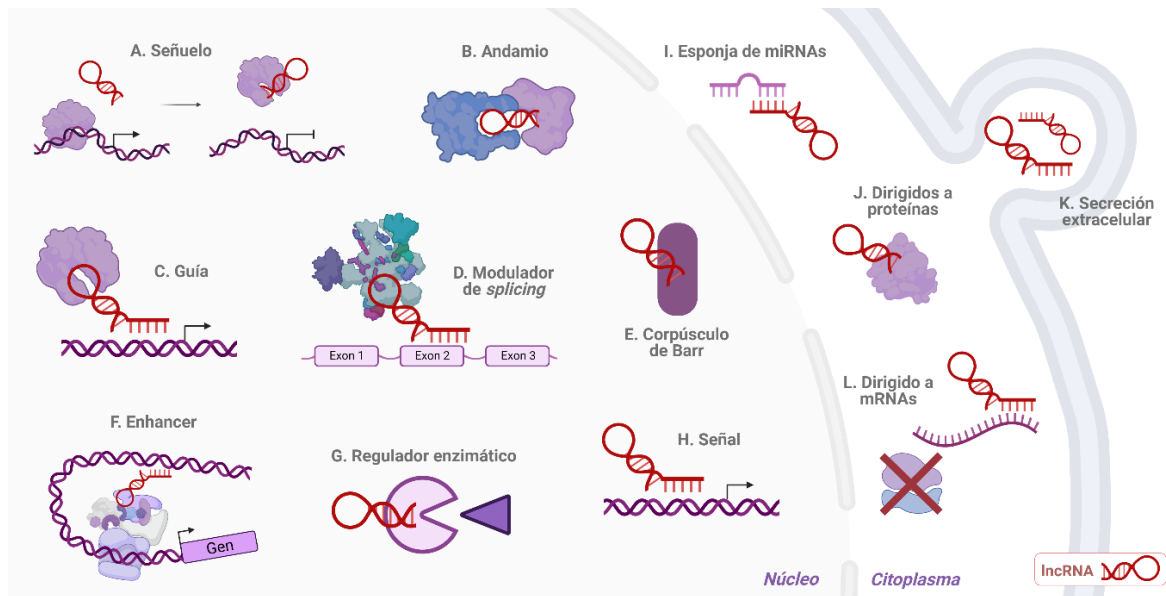


Figura 7. Modelos de los mecanismos de acción propuestos para lncRNA. Principales mecanismos de regulación llevados a cabo por los lncRNA dentro de la célula, del lado izquierdo (A-H) se muestran los mecanismos de acción nucleares y del lado derecho (I-L) los llevados a cabo en el citoplasma. Modificada de Rinn y Chang, 2012⁸⁵, y de Salviano-Silva *et al.*, 2018⁹⁷. Imagen generada en Biorender.com.

1.5.3. RNA largos no codificantes en cáncer

En enfermedades como el cáncer, existen alteraciones a nivel molecular que modifican los niveles de expresión génica, como la sobreexpresión del RA y del lncRNA PCA3, así como en las redes de regulación celular tales como la vía de señalización PI3K-AKT-mTOR; y que, además, promueven la generación del microambiente tumoral adecuado para el desarrollo y la progresión del cáncer⁷⁶. Esto impacta en la homeostasis celular y promueve el crecimiento descontrolado de células tumorales^{98,99}. Los lncRNA han sido asociados a cada una de las características distintivas del cáncer (*hallmarks*), dentro de las cuales, las vías de señalización se modifican para poder mantener la proliferación celular, y promover la viabilidad y la motilidad celular⁹⁹ (Figura 8).

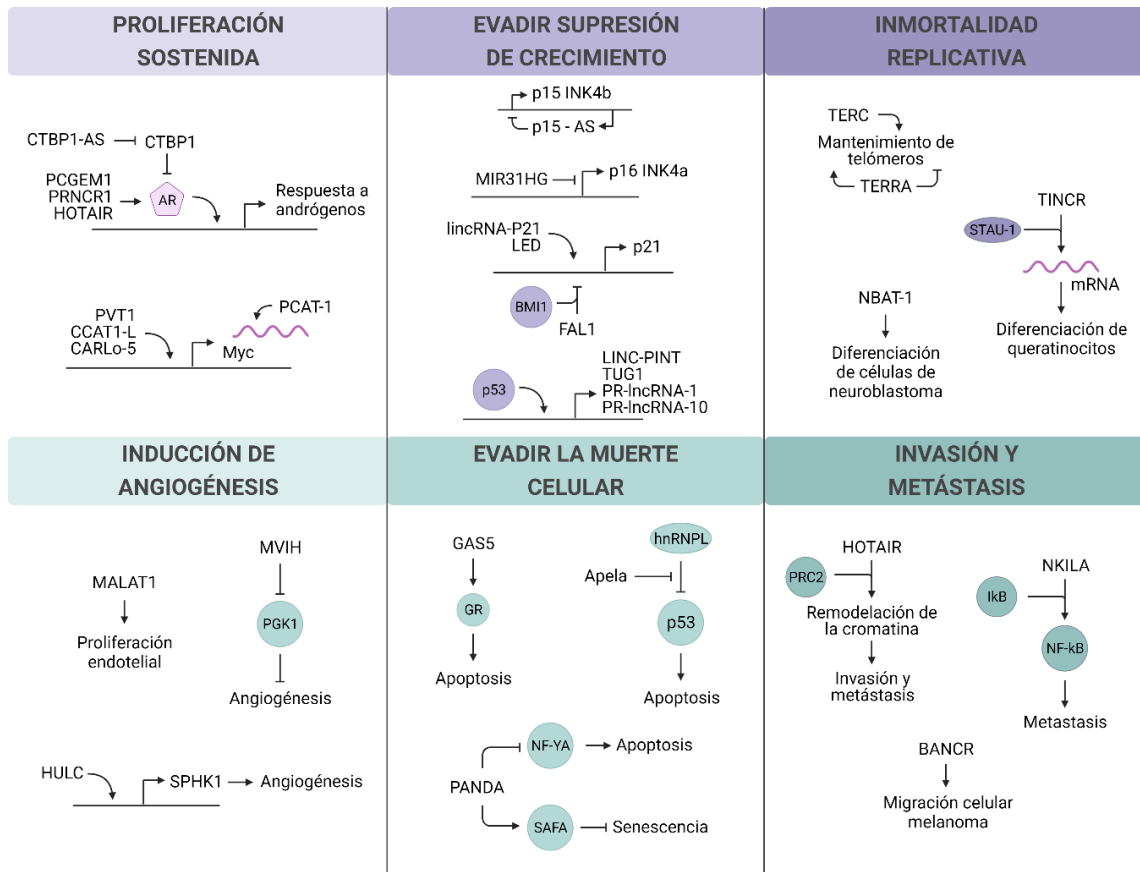


Figura 8. Función de los lncRNA en las principales características distintivas del cáncer. Los lncRNA llevan a cabo un papel importante en la regulación celular, por lo que la alteración en sus niveles de expresión puede favorecer el desarrollo y progresión del cáncer. Gracias a estudios enfocados en el genoma no codificante, se han identificado varios de los procesos carcinogénicos en los que están involucrados. Por ejemplo, PCGEM1 en la proliferación sostenida, MIR31-HG en la evasión de la supresión de crecimiento, TERC en la inmortalidad replicativa, MALAT1 en la angiogénesis, GAS5 en la apoptosis y HOTAIR en la invasión y metástasis. Modificada de Schmitt y Chang, 2016⁹⁹. Imagen generada en Biorender.com.

En el caso específico de CaP, la expresión desregulada de los lncRNA puede a su vez alterar las vías de señalización celulares que promueven el desarrollo y progresión del cáncer (Figura 9). Por ejemplo, la subexpresión del lncRNA supresor de tumores *Maternally Expressed Gene 3* (MEG3) ha sido identificada en tejido de CaP al compararse con tejido sano, y está asociada a la regulación del ciclo celular¹⁰⁰. MEG3 promueve la expresión de la caspasa 3 y de la proteína proapoptótica BCL4, y, por el contrario, inhibe la expresión de la ciclina D1 favoreciendo el arresto del ciclo celular y evitando la proliferación celular exacerbada¹⁰¹ (Figura 9). Aunado a ello, la disminución en los niveles del lncRNA PCA3 provoca la

sobreexpresión de los marcadores epiteliales E-cadherina, claudina-3 y citoqueratina-18, así como la subexpresión del marcador mesenquimal vimentina. En consecuencia, la sobreexpresión de PCA3, característica de CaP, favorece la transición epitelio-mesénquima (EMT por sus siglas en inglés)¹⁰² y la tumorigénesis en CaP consecuencia de la subexpresión del supresor de tumores PRUNE2¹⁰³ (Figura 9).

Los lncRNA que participan en vías de señalización distintivas del cáncer son biomarcadores potenciales de esta enfermedad, según la etapa y el proceso en los que se vean involucrados. Es por ello que se requiere de más estudios enfocados en la identificación, caracterización y validación de lncRNA involucrados en cáncer, para ser propuestos como biomarcadores moleculares de pronóstico con utilidad clínica para el manejo del CaP.

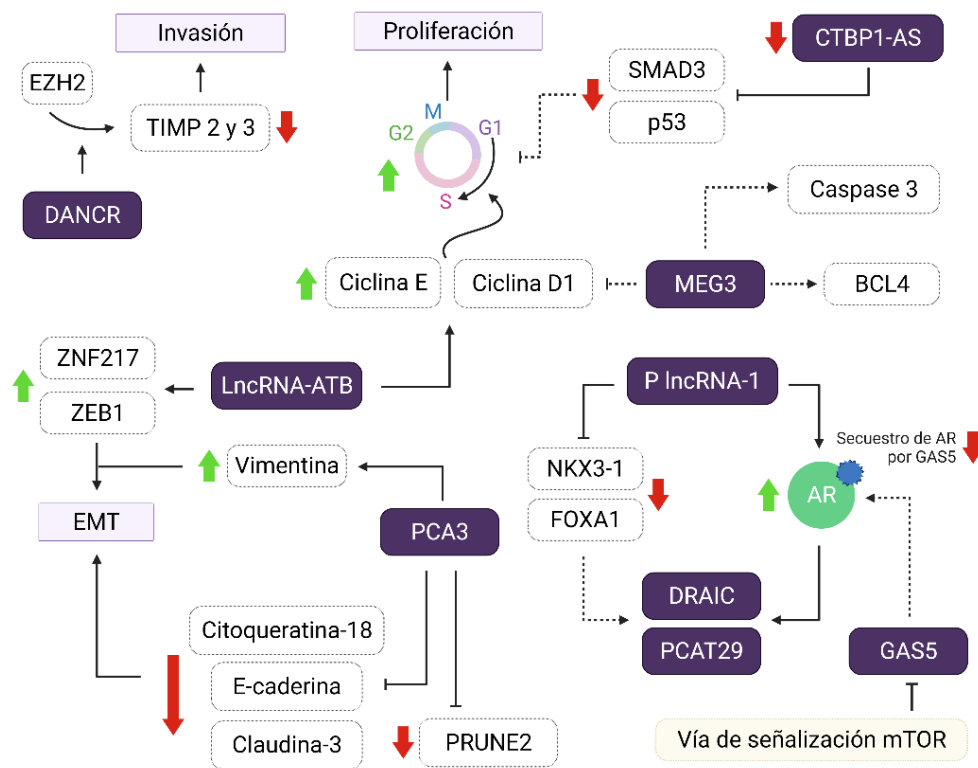


Figura 9. Funciones moleculares de lncRNA en CaP. Imagen representativa de algunos de los lncRNA involucrados en CaP. Los rectángulos morados representan lncRNA. Las flechas rojas indican subexpresión mientras que las verdes, sobreexpresión. Las líneas punteadas indican vías inactivas consecuencia de una subexpresión de lncRNA, genes o proteínas. Modificada de Smolle *et al.*, 2017¹⁰¹. Imagen generada en Biorender.com

1.5.4. El uso de lncRNA como biomarcadores moleculares en cáncer de próstata

Algunos lncRNA como PCA3, SChLAP1, PCAT1, PCGEM1 y UCA1 han sido propuestos como biomarcadores moleculares de pronóstico y diagnóstico para CaP principalmente por su expresión diferencial asociada a la progresión, recurrencia bioquímica y metástasis en pacientes con CaP.

El oncogén PCA3 o DD3, fue descubierto por Bussemakers en 1999, quien reportó que sus altos niveles de expresión eran específicos de CaP⁷⁶. Este lncRNA está sobreexpresado en el 95% de los tumores prostáticos y es sensible a andrógenos¹⁰⁴. Además, PCA3 está involucrado en la regulación de muchos genes asociados a cáncer incluyendo los cofactores del RA¹⁰² así como en el silenciamiento del gen supresor de tumores PRUNE2 en CaP, uniéndose a él y formando un RNA de doble cadena¹⁰³. También juega un papel importante en la progresión del CaP mediante la captura (mecanismo de esponja) del miRNA miR-218-5p, el cual participa en el proceso de invasión celular y, eventualmente, en el proceso de metástasis¹⁰⁵. Actualmente, PCA3 es considerado un biomarcador diagnóstico de CaP altamente específico¹⁰⁶ ya que su detección en muestras de orina reduce el número de biopsias innecesarias, siendo de gran utilidad en el diagnóstico urológico¹⁰⁷. PROGNSA (*Gene-Probe Inc.*) es una prueba comercial aprobada por la FDA en 2012 con una sensibilidad de 62% y una especificidad de 75%¹⁰⁸. Esta prueba es utilizada para medir las concentraciones de PCA3 y del mRNA de APE (KLK3) en muestras de orina de pacientes con DRE positivo y una biopsia negativa previa, determinando si es necesario repetir la toma de la biopsia o no^{109,110}. Dicha prueba ha reducido más del 40% de biopsias innecesarias en pacientes con CaP, demostrando la relevancia que tiene su uso en el área oncológica y su utilidad clínica basada en la expresión de lncRNA.

Otro lncRNA cuya utilidad clínica ha sido demostrada es SChLAP1 el cual se caracteriza por presentar niveles altos de expresión en CaP metastásico (mCRPC). Este lncRNA coordina la invasión metastásica al unirse a la subunidad SNF5 del complejo remodelador de la cromatina SWI/SNF y antagonizando su actividad en la regulación epigenética^{72,104}. Su efectividad como biomarcador pronóstico ha sido

probada en ensayos de hibridación de RNA *in situ*, y también se han desarrollado pruebas basadas en la detección de sus niveles de expresión para relacionar los resultados al estadio clínico del paciente y poder discernir entre CaP indolente y agresivo¹¹¹. Por lo tanto, SChLAP1 es considerado como un biomarcador pronóstico prometedor de utilidad clínica (sensibilidad de 24% y especificidad de 94%¹¹²) y uno de los mejores genes no codificantes para la predicción de metástasis y recurrencia bioquímica en pacientes con CaP^{72,113}.

Asimismo, *Prostate Cancer Associated lncRNA Transcript 1* (PCAT1) es otro RNA con niveles altos de expresión que se identificó originalmente en CaP mediante análisis de RNA-Seq⁷⁴. Este lncRNA es considerado como un oncogén debido a que está involucrado en procesos biológicos como la proliferación celular, apoptosis, invasión y migración celular, así como en la transición epitelio-mesénquima y en la progresión tumoral mediante la vía de señalización intracelular *Wnt/β-catenina*¹¹⁴. Además, PCAT1 reprime la expresión del gen supresor de tumores *BRCA2* provocando el rompimiento de doble cadena de DNA y la deficiencia funcional en la recombinación homóloga¹¹⁵. Este lncRNA también favorece la proliferación celular, ya que regula positivamente a la oncoproteína Myc mediante el mecanismo de señuelo en el que evita que el miRNA dirigido al transcrito cMyc lo degrade¹¹⁶. Todos estos hallazgos demuestran que el lncRNA PCAT1 representa un potencial biomarcador molecular de mal pronóstico asociado a la progresión del CaP.

Urothelial Carcinoma Associated 1 (UCA1) es un lncRNA que se identificó originalmente en carcinoma urotelial, sin embargo, su sobreexpresión se ha reportado en otros tumores como el de cáncer de mama, cáncer colorrectal, cáncer gástrico y CaP, entre otros¹¹⁷. Este lncRNA se asocia a proliferación celular, apoptosis y metástasis, así como en el desarrollo y progresión de cáncer¹¹⁷. Se ha reportado que la subexpresión de UCA1 en el CaP tiene como consecuencia la inhibición de la proliferación celular, ya que, cuando sus niveles de expresión aumentan, UCA1 actúa como esponja de miR-143, favoreciendo la expresión de su gen blanco, MYO6, el cual codifica a una proteína asociada a la migración celular y a la progresión del CaP¹¹⁸. En otro estudio realizado en células de CaP, se demostró que la depleción de UCA1 afecta la capacidad proliferativa y el ciclo celular de las

células tumorales, además de que les genera radiosensibilidad, por lo que UCA1 no sólo tiene valor pronóstico en CaP sino también muestra valor predictivo¹¹⁹.

Como ya se mencionó, existe un número relevante de lncRNA con implicaciones en la patogénesis del cáncer. Específicamente para el CaP, se ha validado ya la utilidad clínica de algunos de ellos, como es el caso de PCA3 y SChLAP1. El primero que es usado actualmente en la clínica mediante una prueba comercial; y el segundo, que junto con PCAT1, fue identificado en CaP mediante la técnica molecular de RNA-Seq. Todos estos descubrimientos demuestran la gran importancia y utilidad que tienen los lncRNA en el manejo del paciente oncológico y en la práctica clínica.

1.6. Transcriptómica enfocada a cáncer de próstata

En 2011, Prensner y colaboradores⁷⁴ identificaron 121 transcritos no codificantes asociados al CaP en una cohorte que incluía 81 muestras de tejidos prostáticos, de las cuales 20 muestras correspondían a tejido adyacente, 47 a tejido tumoral de CaP localizado y 14 a CaP metastásico (muestras de tejido obtenidas por prostatectomía radical), además de 21 líneas celulares de próstata, entre ellas VCaP, LNCaP y PrEC; las cuales son independientes de andrógenos, sensibles a andrógenos y de epitelio sano, respectivamente. Este grupo de lncRNA, nombrados como familia de *PCATs* (del inglés *Prostate Cancer Associated Transcripts*), no había sido anotado anteriormente y fue identificado mediante secuenciación masiva en paralelo de RNA. El RNA-Seq es una técnica de biología molecular basada en secuenciación de siguiente generación (NGS por sus siglas en inglés) de alto rendimiento cuya aplicación ha permitido el análisis de transcriptomas completos, es decir, el análisis de todos los transcritos que se están sintetizando en un tejido, tiempo y espacio determinado, el cual incluye genes codificantes y no codificantes como los lncRNA. Asimismo, permite la identificación de alteraciones en los niveles de expresión de los transcritos presentes en enfermedades. Una de sus principales aplicaciones es el análisis de expresión diferencial en modelos biológicos y sus implicaciones a nivel celular¹²⁰, es por ello que ha tenido un gran impacto en la generación de nuevos conocimientos acerca de los lncRNA, su relación con el

cáncer y la generación de nuevos biomarcadores. Esta herramienta ha mostrado ventajas sobre metodologías como microarreglos de expresión y PCR en tiempo real en el sentido de que el RNA-Sec explora todo el transcriptoma y no se limita a la detección de secuencias o transcritos ya conocidos, por lo que esta herramienta ha servido para descubrir lncRNA desconocidos o que no han sido caracterizados hasta la fecha.

Un análisis de expresión diferencial por RNA-Sec consiste en la extracción de RNA total y la síntesis de cDNA a partir de éste, así como la generación de una biblioteca de secuenciación con adaptadores seguida de la secuenciación por síntesis (Figura 10). La biblioteca preparada es secuenciada con una profundidad de 10 a 30 millones de lecturas por muestra, la profundidad depende del objetivo principal del estudio y se refiere al número total de lecturas secuenciadas obtenidas de una muestra. La secuenciación también puede llevarse a cabo por extremo sencillo (*single-end*) o por extremos pareados (*paired-end*), en esta última se analizan los productos provenientes tanto de la hebra sentido como de la hebra antisentido, lo que permite identificar transcritos provenientes de ambas hebras, así como el sentido en el que se sintetiza un transcrito. Como último paso se lleva a cabo el análisis bioinformático, éste conlleva el alineamiento y el ensamblaje de las lecturas secuenciadas a un genoma humano de referencia, la cuantificación de lecturas, y el modelado estadístico de los cambios significativos en los niveles de expresión¹²⁰. Gracias a la precisión con la que cuenta este tipo de herramientas, es posible identificar biomarcadores de utilidad clínica en CaP, ya que se puede evaluar la expresión génica tanto de células obtenidas de tejido tumoral, así como células obtenidas de tejido adyacente o de tejido sano para detectar expresión diferencial entre transcritos de lncRNA involucrados en la recurrencia bioquímica del CaP.

Los análisis de expresión diferencial derivados de los datos obtenidos por RNA-Sec, permiten llevar a cabo estudios de casos y controles que son ampliamente aplicados en la investigación de patologías como el CaP, ya que identificar los cambios a nivel transcriptómico en las células tumorales puede

mejorar las terapias en la clínica y detectar posibles blancos terapéuticos o biomarcadores moleculares específicos de utilidad clínica.

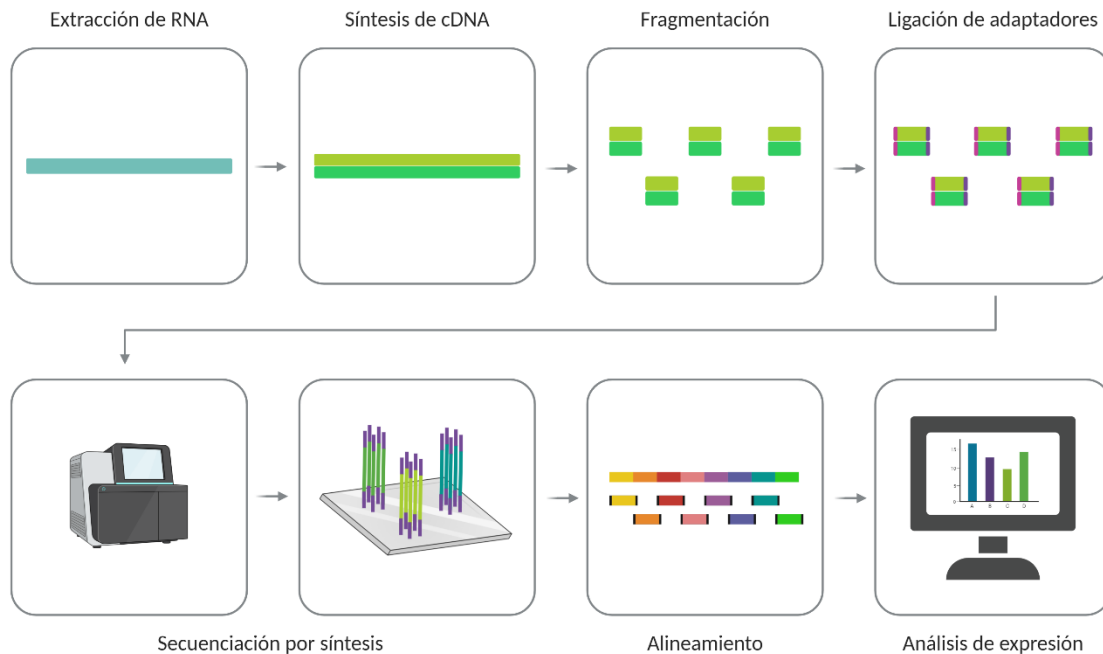


Figura 10. Secuenciación masiva en paralelo de RNA o RNA-Sec. Imagen representativa del flujo de trabajo llevado a cabo en un análisis de RNA-Sec. Para ello se debe extraer RNA total de la muestra biológica, sintetizar cDNA a partir de éste y, posteriormente, generar una biblioteca después de que se le hayan añadido adaptadores a cada fragmento para su identificación. La secuenciación se lleva a cabo por síntesis mediante PCR, en ella se utilizan cuatro nucleótidos marcados con etiquetas fluorescentes (dNTPs) para secuenciar los *clusters* formados en la superficie de la placa. La unión de los dNTPs en cada ciclo es detectada y registrada, y una vez realizado el alineamiento de las lecturas al genoma humano de referencia, se pueden determinar los niveles de expresión según el enfoque y método bioinformático utilizado. Por ejemplo, la herramienta *Salmon* no lleva a cabo un alineamiento base por base, sino un “cuasi-mapeo”¹²¹, mientras que la herramienta *STAR* utiliza una búsqueda de regiones semilla¹²². Imagen generada en Biorender.com.

2. JUSTIFICACIÓN

Actualmente, la prostatectomía radical es el principal tratamiento para el cáncer de próstata localmente avanzado; sin embargo, las técnicas para la predicción de la recidiva posterior a la cirugía son escasas y el biomarcador mayormente utilizado para su detección es el antígeno prostático específico. Por otra parte, los estudios genómicos y moleculares enfocados en cáncer continúan mejorando el manejo clínico de los pacientes y llevándolo a la era de la medicina de precisión basada en las características individuales de cada paciente. Sin embargo, a pesar de los grandes esfuerzos concentrados en la investigación de marcadores moleculares, actualmente no se ha introducido algún biomarcador molecular pronóstico para CaP basado en lncRNA que haya sido aprobado por la FDA. Por lo cual, es de suma importancia identificar marcadores moleculares característicos de la población mexicana que puedan apoyar el pronóstico como lo son los lncRNA. Estos transcritos pueden tener funciones tanto de oncogén como de gen supresor de tumores, es decir, pueden favorecer la carcinogénesis ya que su expresión anormal ha sido asociada al desarrollo de CaP a través de la regulación de procesos celulares implicados en proliferación, metástasis y apoptosis. El uso de herramientas de frontera como RNA-Seq ha permitido identificar este tipo de lncRNA con potencial biomarcador y, por ende, apoyar la medicina personalizada del paciente oncológico. En consecuencia, los estudios basados en RNA-Seq para la búsqueda de lncRNA como biomarcadores moleculares de pronóstico podrán generar información útil en el área clínica y oncológica para predecir el pronóstico de pacientes con CaP localmente avanzado de alto riesgo.

Es por ello que en este proyecto nos planteamos proponer un conjunto de RNA largos no codificantes como biomarcadores moleculares de pronóstico en cáncer de próstata. Estos se evaluarán en una cohorte de pacientes con recurrencia bioquímica sometidos a prostatectomía radical con el fin de poder dar un pronóstico temprano.

3. PLANTEAMIENTO DEL PROBLEMA

El cáncer de próstata es un problema de salud pública en México ya que ocupa el primer lugar en tasas de incidencia y mortalidad en hombres mexicanos de todas las edades. Debido a que el cáncer de próstata es una enfermedad que generalmente se comporta de manera indolente, ésta es detectada principalmente en estadios avanzados en los que el tratamiento correspondiente es la prostatectomía radical. Sin embargo, el 30% de los casos de cáncer de próstata tienen un relapso después de la cirugía y actualmente son pocos los biomarcadores moleculares de pronóstico que predigan recurrencia bioquímica en pacientes, por lo que es necesario identificar nuevos marcadores moleculares que estén involucrados en la progresión de cáncer y tengan expresión tumor- y etapa-específica como lo son los RNA largos no codificantes. Por lo anterior y con base en la tasa de mortalidad por CaP en México y la importancia de los lncRNA en el desarrollo y la progresión del cáncer, es necesaria la búsqueda de biomarcadores moleculares basados en lncRNA que puedan ser de utilidad clínica para identificar a los pacientes con CaP que presentarán un mayor riesgo de recurrencia temprana de la enfermedad.

4. PREGUNTA DE INVESTIGACIÓN

¿El perfil de expresión de lncRNA de muestras de pacientes mexicanos con cáncer de próstata localmente avanzado de alto riesgo, es capaz de predecir el desarrollo de recurrencia bioquímica?

5. HIPÓTESIS

Si el perfil de expresión de lncRNA es capaz de distinguir las muestras de pacientes con recurrencia bioquímica, entonces podrá considerarse como un biomarcador molecular con valor pronóstico en cáncer de próstata localmente avanzado de alto riesgo.

6. OBJETIVOS

6.1. *Objetivo general*

Identificar lncRNA que sirvan como biomarcadores moleculares de pronóstico en pacientes con cáncer de próstata localmente avanzado de alto riesgo.

6.2. *Objetivos particulares*

- I. Evaluar la calidad del RNA total extraído del tejido neoplásico obtenido de pacientes sometidos a prostatectomía radical, y la calidad de los datos de la secuenciación masiva en paralelo de RNA.
- II. Identificar los patrones de expresión de lncRNA en pacientes con cáncer de próstata localmente avanzado de alto riesgo comparando tejido tumoral contra tejido adyacente.
- III. Definir un lncRNA o un panel de biomarcadores moleculares de lncRNA con expresión específica de cáncer de próstata localmente avanzado de alto riesgo.
- IV. Proponer un lncRNA o un panel de biomarcadores moleculares de lncRNA que distinga a los pacientes que presenten recurrencia bioquímica de los que no.
- V. Realizar un análisis *in silico* para la caracterización de los lncRNA candidatos.

7. METODOLOGÍA

7.1. Abordaje

Para la identificación de lncRNA como marcadores moleculares de pronóstico en cáncer de próstata mediante análisis de transcriptoma, se llevarán a cabo dos fases, una de descubrimiento y una de validación a partir de una cohorte retrospectiva de pacientes con cáncer de próstata localmente avanzado de alto riesgo. En la fase de descubrimiento se utilizarán 3 muestras de tejido neoplásico obtenido mediante prostatectomía radical y 3 muestras de tejido normal (tejido adyacente al tumor). A partir del RNA purificado de los tejidos obtenidos por cirugía, se hará la preparación de librerías para su posterior secuenciación masiva en paralelo de RNA en la plataforma *Illumina* en un equipo *HiSeq 2500* con una profundidad de 30 millones de lecturas, esto nos dará una ventaja adicional para el análisis de toda la población de lncRNA, pudiendo detectar hasta lncRNA de muy baja abundancia y los no caracterizados. Posteriormente, mediante un abordaje bioinformático, se hará una búsqueda tanto de lncRNA anotados como no anotados y, mediante un análisis de componentes principales, se establecerá si hay un agrupamiento entre las muestras provenientes de tejido neoplásico en comparación con las que se obtuvieron del tejido adyacente. Además, mediante el análisis de expresión jerárquica sabremos si existe un panel de lncRNA que distinga a pacientes de mayor riesgo con el objetivo de determinar la existencia de un grupo de lncRNA que se expresen diferencialmente en los pacientes que muestren una recurrencia y/o progresión de la enfermedad; dichos lncRNA serán los candidatos a evaluar en la segunda fase. La fase de validación, que corresponde a la segunda fase del proyecto, estará formada por una cohorte independiente de 16 pacientes (con las mismas características clínico-patológicas de la fase de descubrimiento) en la cual se validará la presencia del conjunto de lncRNA diferencialmente expresados mediante PCR en tiempo real. Finalmente, se determinará una firma con expresión tejido-específica y una con expresión específica de recurrencia bioquímica, y éstas se caracterizarán de manera *in silico*. Los detalles del abordaje metodológico serán explicados en las siguientes secciones.

7.2. Fase de descubrimiento

A continuación, se describen los protocolos y procedimientos que se llevaron a cabo para la recolección de muestras y para el análisis bioinformático de los datos de secuenciación de RNA.

7.2.1. Criterios de inclusión de pacientes

La Unidad de Investigación Biomédica en Cáncer del INCan cuenta con un banco de tumores que incluye 320 muestras de pacientes mayores a 18 años, diagnosticados con cáncer de próstata y que forman parte de la población que ingresa al instituto. En el presente estudio se incluyeron las muestras correspondientes a pacientes con cáncer de próstata localmente avanzado de alto riesgo con confirmación histológica, sin tratamiento oncológico previo y sometidos a prostatectomía radical. Se analizó un total de 6 muestras correspondientes a 3 pacientes que firmaron con anterioridad el consentimiento informado aprobado por el comité de ética del INCan. El protocolo de investigación de la presente tesis se encuentra registrado ante los comités de ética del INCan con el número (019/004/DII) (CEI/1358/18).

7.2.2. Purificación del RNA total de muestras de pacientes

Se realizó la purificación del RNA total de las 6 muestras de tejido prostático de los 3 pacientes obtenida mediante prostatectomía radical, basándose en el protocolo de extracción de ácidos nucleicos de alta calidad para estudios genómicos descrito por Peña-Llopis y Brugarolas en 2013¹²³. Para ello se utilizó el kit *AllPrep* de QIAGEN (Cat. 80204) mediante el cual se extrajo simultáneamente RNA y DNA a partir de tejido prostático fresco. Posteriormente se cuantificó la concentración de cada muestra y se comprobó su calidad en el bioanalizador *Tape Station 2200* (*Agilent Technologies*) considerando un RIN mayor o igual a 7 como parámetro de buena calidad; por último, se almacenaron a -80°C hasta su posterior uso.

7.2.3. Secuenciación masiva en paralelo de RNA (RNA-Sec)

Se prepararon alícuotas de RNA correspondientes a cada una de las 6 muestras (tejido adyacente: 3 y tejido tumoral: 3) con 1 µg total de RNA. La secuenciación se llevó a cabo por extremos pareados, con una profundidad de 25 millones de lecturas y mediante la plataforma *HiSeq* 2500 de Illumina. Las muestras se entregaron y procesaron en el Laboratorio de Genómica de la Unidad de Red de Apoyo a la investigación (RAI), en el Instituto Nacional de Ciencias Médicas y Nutrición “Salvador Zubirán” para su secuenciación, la cual estuvo a cargo del Dr. Inti Alberto de la Rosa.

7.2.4. Análisis bioinformático de expresión diferencial

Con el fin de determinar los niveles de expresión de lncRNA en muestras de CaP, se comprobó que la calidad de los datos de secuenciación fuera adecuada para posteriormente realizar un alineamiento de lecturas al genoma de referencia (hg38) y una cuantificación de transcritos. Se consideró como grupo control al tejido adyacente y como grupo problema al tejido tumoral de cada uno de los pacientes sometidos a prostatectomía radical.

7.2.4.1. Control de calidad de los datos de secuenciación

Para comprobar que la calidad de las lecturas obtenidas de la secuenciación cumplieran con los criterios de calidad para llevar a cabo análisis bioinformáticos, los datos se analizaron con la herramienta bioinformática FastQC y MultiQC del programa en línea *Galaxy*¹²⁴ cuyo sistema analiza los archivos FastQ que contienen la secuenciación obtenida en formato FASTA. Los criterios utilizados para el análisis fueron los predeterminados por el programa. El análisis por FastQC se lleva a cabo mediante una serie de módulos de análisis a partir de los cuales se emite un reporte en donde se describen once parámetros de calidad que clasifican cada muestra como calidad aceptable, mala calidad o inaceptable, dentro de los cuales se encuentran: la calidad de la secuencia por base y el contenido N por base. Se asume

que la secuenciación cumple con los criterios de calidad deseables cuando cada muestra califica como “aceptable” en los parámetros mencionados anteriormente.

7.2.4.2. *Cambio de plataforma de secuenciación*

Con el fin de homogeneizar el formato de los archivos para adecuarlo al análisis en *Galaxy*, se realizó la conversión de formato de calidad de plataforma de los archivos *FastQ* modificando la codificación del puntaje de calidad que les fue asignado después de la secuenciación. Para ello se utilizó la herramienta *FastQ Groomer* de *Galaxy*.

7.2.4.3. *Alineamiento de secuencias*

Para comenzar con el análisis de datos de secuenciación, se inició mapeando las lecturas de la secuencia a un transcriptoma conocido o a un genoma anotado de referencia, convirtiendo cada lectura en una coordenada genómica. Para ello se utilizó la herramienta STAR también disponible en la plataforma de *Galaxy*. STAR permite hacer un mapeo y alineamiento de la secuencia en donde se toman en cuenta los espacios entre lecturas de genes que no hayan sido transcritos, además de descartar lecturas de baja calidad que siguieran presentes. El genoma de referencia contra el que se comparan las lecturas y sobre el que se realiza el mapeo fue hg38, la versión más actualizada del genoma humano (*Homo sapiens* (b38): hg38) disponible en *Galaxy*. Para el análisis de lecturas pareadas y para el resto de los parámetros se eligieron los predeterminados por el programa.

7.2.4.4. *Cuantificación de transcritos*

Para poder cuantificar la proporción de transcritos detectados durante la secuenciación, se utilizó la herramienta *Salmon* también incluida en *Galaxy*. Una vez que los transcritos fueron mapeados a una localización genómica específica fueron asociados a algún gen o transcrito para determinar su abundancia. Mediante *Salmon* se detectó la presencia de transcritos de cada muestra a partir de los archivos *FastQ* guiándose del genoma de referencia en formato FASTA para

transcritos (GRCh38.p13.genome.fa.gz)¹²⁵. En los parámetros se estableció que las lecturas eran pareadas y se especificó el archivo del genoma de referencia que sería utilizado para la cuantificación.

7.2.4.5. *Análisis de expresión diferencial*

Una vez cuantificados los transcritos secuenciados, se utilizó el programa *DESeq2* disponible en *Galaxy* para determinar qué transcritos son los que presentaron diferencias en sus niveles de expresión en tejido tumoral (grupo problema) a comparación del tejido adyacente (grupo control). La herramienta *DESeq2* generó un reporte en donde se presenta un gráfico de análisis de componentes principales (PCA), un mapa de calor y un gráfico de volcán; así como una lista de los lncRNA con expresión diferencial en las muestras de tejido tumoral con el identificador de *Ensembl* de cada gen, su tasa de cambio en expresión (log₂FC: logaritmo base 2 de la tasa de cambio de la expresión del transcrito), error estándar y el valor de *p*.

7.2.4.6. *Selección de datos del análisis de expresión diferencial*

La selección del listado de genes candidatos con mayor expresión diferencial se hizo con base en los resultados obtenidos a partir del análisis de expresión diferencial, los cuales fueron filtrados mediante la interfaz *R Studio*. La firma molecular de lncRNA se identificó utilizando la base de datos en ncRNA de *Ensembl*. Los criterios considerados para la selección de lncRNA fueron la asociación de los lncRNA a valores de FDR ≤ 0.05 y que, adicionalmente, tuvieran valores de log₂FC (logaritmo base 2 de la tasa de cambio de la expresión del transcrito) mayores o iguales al punto de corte, en este caso, 1.5 para los genes sobreexpresados y -1.5 para los subexpresados. Fue así como, basándose en la significancia estadística, se seleccionaron al subgrupo de lncRNA con mayor expresión diferencial, 5 sobreexpresados y 3 subexpresados.

7.3. Fase de validación experimental

7.3.1. Fase de validación experimental A: análisis de expresión diferencial de tejido tumoral vs. tejido adyacente

Con el fin de corroborar lo observado en los análisis bioinformáticos, se continuó con la fase de validación experimental en una cohorte independiente de pacientes en la que se evaluaron los niveles de expresión de lncRNA en muestras de tejido adyacente y tumoral mediante un arreglo de expresión de lncRNA.

7.3.1.1. Obtención de muestras y criterios de inclusión (cohorte independiente)

Los criterios de inclusión para la selección de muestras de la fase de validación A son los mismos que se describen en el apartado 7.2.1. Criterios de inclusión de pacientes de la fase de descubrimiento. Para esta etapa se analizaron 32 muestras de pacientes (16 de tejido tumoral y 16 de tejido adyacente).

7.3.1.2. Purificación, cuantificación y determinación de la integridad del RNA total

La extracción de RNA total de las muestras de pacientes, así como la cuantificación de éste y la evaluación de calidad e integridad, se llevó a cabo de la misma manera descrita en el apartado 7.2.2. Purificación de RNA total de muestras de pacientes. En esta fase de validación se consideraron valores de RIN menores a 7 como aceptables siempre y cuando los ratios 260/280 y 260/230 fueran óptimos.

7.3.1.3. Análisis de expresión diferencial

Con el fin de validar los resultados observados en el análisis de secuenciación masiva, se realizó un ensayo de RT-qPCR mediante arreglos de expresión lncRNA con los transcritos no codificantes más representativos en cáncer, para ello se dio uso al Kit *RT² lncRNA qPCR Assay (Human Cancer PathwayFinder)* de QIAGEN.

7.3.1.3.1. Tratamiento con DNasa

A pesar de que el kit RT² cuenta con un paso de eliminación de DNA genómico, se les dio un tratamiento previo con DNasa a cada muestra de RNA para asegurar que

estuvieran libres de contaminación por DNA. Para ello se usó a la DNasa Max de QIAGEN (Cat. 15200-50). El tratamiento consistió en preparar cada reacción con 1.25 μL de Solución Amortiguadora 10X para DNasa Max, 0.5 μL de enzima DNasa I (10 U/ μL), el volumen correspondiente a 500 ng de RNA y agua libre de nucleasas para llegar a un volumen de 10 μL totales. Posteriormente, cada muestra se incubó a 37°C durante 20 minutos y, por último, se les agregó 2.5 μL de la resina removedora de DNasa una vez homogeneizada por agitación, se incubaron a temperatura ambiente durante 10 minutos mientras se resuspendía suavemente cada 2 minutos. Pasado este tiempo, se centrifugaron a 13,000 rpm durante 1 minuto y, posteriormente, se transfirió el sobrenadante a un microtubo nuevo sin tocar la resina precipitada. Por último, se conservó a -80°C hasta su posterior uso.

7.3.1.3.2. Síntesis de cDNA

La síntesis de cDNA a partir del RNA extraído se llevó a cabo mediante oligonucleótidos *random primers* incluidos en el kit de la siguiente manera: se preparó una alícuota de 8 μL que contuvieran 1 μg total de RNA y se le agregaron 2 μL de buffer GE para la eliminación de las trazas restantes de DNA genómico. Se incubó a 42°C durante 5 minutos y se dejó reposando en hielo inmediatamente durante 1 minuto. Una vez eliminado el DNA genómico, en un microtubo de centrifuga de 1.5 mL nuevo, se preparó el Mix de Retrotranscripción (Mix RT) el cual se compone de 2 μL del Mix RE3 *Reverse Transcriptase*, 4 μL de la Solución Amortiguadora BC3, 1 μL del control P2 y 3 μL de agua libre de nucleasas para obtener un volumen final de 10 μL . Se agregaron 10 μL del Mix RT a cada microtubo con RNA, se incubó a 37°C durante 60 minutos e inmediatamente a 95°C durante 5 minutos, una vez cumplidos los tiempos, se dejó reposando en hielo y, finalmente, se les agregaron 91 μL de agua libre de nucleasas.

7.3.1.3.3. PCR en tiempo real para el arreglo de expresión de lncRNA

Una vez sintetizado el cDNA, se continuó con el ensayo de PCR tiempo real en las placas con el arreglo de expresión de lncRNA, para ello se preparó una mezcla con 1350 μL de *Mastermix 2X RT2 SYBR Green*, 102 μL de cDNA y 1248 μL de agua

libre de nucleasas. Se agregaron 25 μ L de reacción a cada pozo de la placa, se centrifugó durante 1 minuto a 1,000 rpm y se colocó en el termociclador programado con las condiciones descritas en la tabla 2.

Tabla 2. Programación de termociclador para PCR tiempo real del arreglo de lncRNA.

Ciclos	Duración	Temperatura
1	10 min	95°C
40	15 s	95°C
	1 min	60°C

Una vez terminada la reacción, se ajustó el valor del umbral para que las muestras pareadas coincidieran en él, es decir, la muestra tumoral debía tener el mismo valor de umbral que la muestra adyacente correspondiente.

7.3.1.3.4. Análisis cuantitativo de expresión diferencial

Para poder llevar a cabo el análisis cuantitativo del panel de lncRNA, el archivo con los resultados exportado del software del termociclador es importado a la plataforma en línea de QIAGEN para la evaluación de los resultados. Para ello se modificó el punto de corte de Ct a 40 ciclos y se descartaron como genes control a aquellos genes endógenos cuya desviación estándar y margen de error fuera alto, esto evaluando su variación en los valores de Ct resultantes en la comparación de tejido adyacente contra tejido tumoral. Para el resto de los parámetros y puntos de corte se mantuvieron los valores predeterminados por la herramienta (FoldReg: 2 y -2; *pvalue* < 0.05). Esta plataforma en línea genera los siguientes gráficos: mapa de calor, gráfico de volcán y gráfica de barras.

7.3.2. Fase de validación experimental B: análisis de expresión diferencial en líneas celulares de tejido prostático

Con el fin de corroborar lo observado en el análisis de expresión diferencial por el ensayo de RNA-Seq y por el arreglo de expresión de lncRNA, se llevó a cabo un ensayo de validación en un modelo celular. Para ello se utilizaron las líneas celulares de LNCaP y PrEC.

7.3.2.1. Purificación de RNA total de líneas celulares

En primera instancia, se estandarizaron los parámetros de los ensayos de expresión diferencial del lncRNA candidato H19. Se llevó a cabo la extracción de RNA total en líneas celulares de próstata para posteriormente ser cuantificado y determinar su integridad para continuar con los ensayos siguientes.

7.3.2.1.1. Cultivo de líneas celulares

Las líneas celulares seleccionadas para la detección de estos genes fueron dos líneas celulares humanas de próstata: LNCaP (tumoral) y PrEC (epitelial normal), las cuales fueron cultivadas bajo condiciones estériles y sin ningún tipo de tratamiento.

I. LNCaP (ATCC - CRL.1740)

Para llevar a cabo el cultivo celular, se eligió la línea celular de adenocarcinoma de próstata LNCaP de origen humano (*Homo sapiens*) proveniente del sitio metastásico desarrollado en un nodo linfático supraclavicular de un hombre caucásico de 50 años, su morfología es epitelial y con propiedad adherente. Su crecimiento se llevó a cabo en medio de cultivo RPMI suplementado con suero fetal bovino (SFB) al 10%, en botellas de cultivo de 25 cm² (*Corning*) y manteniendo las condiciones de incubación de 37°C y con un nivel de CO₂ al 5%. Se inició con una cepa en pase 10 hasta llegar al 15 haciendo levantamientos cada vez que la confluencia celular fuese cercana al 90%.

II. PrEC (Lonza - lote 0000654537)

Se eligió la línea celular de tejido prostático epitelial PrEC de origen humano (*Homo sapiens*) para realizar el cultivo del control sano. Éste se llevó a cabo en medio de cultivo PrEGM, en botellas de cultivo de 25 cm² (Corning) y manteniendo las condiciones de incubación a 37°C y con un nivel de CO₂ al 5%. Se inició con una cepa en pase 10 hasta llegar al 15 haciendo levantamientos cada vez que la confluencia celular fuese cercana al 90%.

7.3.2.1.2. Purificación de RNA total

Para llevar a cabo la extracción de RNA de las células cultivadas correspondientes a los cultivos de LNCaP y PrEC, a cada botella de cultivo de 25 cm² se les retiró el medio en el que crecieron y se lavaron con amortiguador de fosfatos (PBS por sus siglas en inglés) para eliminar cualquier residuo de medio restante. Las células fueron tratadas posteriormente con el reactivo de TRIzol (*Ambion, Life Technologies, de Thermo Fisher Scientific*) con el fin de lisarlas. Para ello se agregó 1 mL del reactivo a cada botella y se distribuyó homogéneamente sobre la superficie interna para recolectar el mayor número de células adheridas a la base. Las células en TRIzol fueron colocadas en un microtubo y se incubaron en hielo durante 3 minutos, posteriormente se les agregaron 200 µL de cloroformo y se centrifugaron a 14,000 rpm durante 15 minutos a 4°C para separar los componentes celulares.

El reactivo de TRIzol separa la muestra en 3 fases: 1. la fase orgánica en el fondo del microtubo, la cual contiene la membrana plasmática y proteínas, entre otras moléculas y componentes celulares por desechar; 2. la interfase, en donde se encuentra el DNA; y 3. la fase acuosa en la superficie, la cual contiene el RNA. Por lo que a continuación se tomó la fase acuosa con RNA y, una vez separada en un microtubo nuevo, se le agregó isopropanol para precipitar el RNA, cada muestra se dejó incubando en hielo durante 10 minutos y posteriormente se centrifugó a 14,000 rpm durante 10 minutos a 4°C. A continuación, se retiró el sobrenadante con isopropanol cuidando no tocar el botón de RNA formado al fondo del microtubo.

Por último, a cada muestra se le agregó 1 mL de etanol al 75% y se centrifugó a 14,000 rpm una vez más durante 5 minutos a 4°C. Se retiró el sobrenadante

cuidando no tocar la pastilla y ésta se dejó secando a temperatura ambiente hasta que todo el etanol se evaporara. Una vez seca, la pastilla se eluyó en agua atemperada previamente a 37°C y se incubó a 60°C durante 15 minutos para, inmediatamente después, pasarlas a un baño de hielo y, en caso de ser necesario, congelar las alícuotas a -80°C.

7.3.2.1.3. Cuantificación de RNA

La cuantificación del RNA extraído con anterioridad se llevó a cabo por medio del espectrofotómetro NanoDrop 2000 (*Thermo Fisher Scientific*) en la función de lectura de absorbancia para RNA y utilizando como blanco agua libre de nucleasas. Se colocó en el pedestal 1 µL de RNA correspondiente a ambas líneas celulares. Una vez obtenidos los resultados de la cuantificación, se comprobó que los radios de absorbancia 260/280 y 260/230 fueran cercanos a 2.0 para determinar que no existe contaminación por otras moléculas como DNA y proteínas, y entre 2.0 a 2.2 para corroborar que no haya presencia de fenoles, respectivamente.

7.3.2.2. *Cuantificación del lncRNA H19*

Una vez que el RNA total fue extraído y su calidad y concentración determinada, se procedió a hacer el análisis cuantitativo de la expresión relativa del lncRNA candidato en LNCaP y PrEC por PCR en tiempo real. Los resultados fueron analizados por el método delta-delta Ct ($2^{-\Delta\Delta Ct}$) y la significancia estadística se determinó por el estadístico tipo t-student de una cola, no pareada.

7.3.2.2.1. Diseño de oligonucleótidos

Para llevar a cabo el ensayo de qPCR, se diseñaron dos pares de oligonucleótidos para el gen de cada lncRNA seleccionado. Para ello se buscó la secuencia completa de cada uno en las plataformas en línea de *Ensembl Genome Browser*¹²⁶, NCBI (*National Center of Biothechnology Information*)¹²⁷ y LNCipedia¹²⁸. A partir de ellas se buscó la región conservada de todas las isoformas del transcrito reportadas. Los posibles pares de oligonucleótidos se obtuvieron a partir de la herramienta *Primer-Blast* de NCBI¹²⁹ y los parámetros de diseño fueron los siguientes: el producto de PCR debía ser mayor a 70 nucleótidos y no pasar de 150; una Tm óptima de 60°C

y un rango máximo de 2°C de diferencia respecto a la óptima, en este caso, 58°C y 62°C. El resto de los parámetros se mantuvieron iguales a los predeterminados por el programa.

Posteriormente, se descartaron todos aquellos oligonucleótidos cuyo par no coincidiera en número de nucleótidos y aquellos que tuvieran alta frecuencia de bases G o C. Como segundo filtro, se dio uso a la herramienta *Oligo Analyzer Tool*³⁰ de IDT (*Integrated DNA Technologies*) y se evaluaron los candidatos con el objetivo de elegir uno cuyo contenido de CG se mantuviera dentro de 50-60%, y que no formara estructuras secundarias, heterodímeros o dímeros consigo mismo. Por último, se corroboró que los oligonucleótidos seleccionados generaran un amplicón correspondiente al gen de lncRNA, para ello se utilizó la herramienta en línea *In-Silico PCR*³¹ de *UCSC Genome Browser*. Una vez comprobado, se solicitó su síntesis en IDT (Tabla 3).

Tabla 3. Información de los oligonucleótidos diseñados para los experimentos de PCR en tiempo real.

Nombre	Longitud (bases)	Longitud del amplicón (bases)	Secuencia
H19_132_(APAZ)	19	132	Fw: ACA TGA CAT GGT CCG GTG T
			Rv: CAT AGT GTG CCG ACT CCG T
H19_98_(APAZ)	21	98	Fw: GAG GAA CCA GAC CTC ATC AGC
			Rv: GTG GCC ATG AAG ATG GAG TCG

Fw: *forward* (sentido). Rv: *reverse* (antisentido).

7.3.2.2.2. Síntesis de cDNA

Para poder llevar a cabo la cuantificación de transcritos por la técnica de PCR en tiempo real, fue necesario generar cDNA a partir del RNA extraído de líneas celulares.

I. Tratamiento con DNasa

Para poder asegurar que las muestras estuvieran libres de contaminación por DNA genómico, se trató el RNA extraído con la enzima TURBO DNasa (*Thermo Fisher Scientific*). El procedimiento consistió en agregar 2 µL de solución amortiguadora

10X a 17.5 μL de RNA y posteriormente, agregar 0.5 μL de TURBO DNase (2 U/ μL) para obtener un volumen final de 20 μL , la mezcla se incubó a 37°C durante 1 hora. Posteriormente se le agregaron 2 μL de reactivo inactivador y se incubó a 26°C durante 5 minutos. Por último, cada muestra se centrifugó por, aproximadamente, 2 minutos para precipitar el reactivo inactivador separando el sobrenadante con RNA (Se utilizó el kit TURBO™ DNase, Cat. AM2238 de *Thermo Fisher Scientific*).

II. Tratamiento de transcriptasa reversa

Para llevar a cabo la síntesis de cDNA, se mezclaron 10 μL de RNA tratado con DNasa junto con 3.2 μL de agua libre de nucleasas estéril, 2 μL de Solución Amortiguadora PCR 10X (II), 0.8 μL de dNTPs 10 mM, 2 μL de *Random Primers*, 1 μL de Inhibidor de RNasa y 1 μL de enzima Retro-transcriptasa de alta capacidad (RT) a 50U/ μL . Para cada RNA se preparó un control negativo de la reacción (-RT) en donde se sustituyó el volumen de la enzima por agua libre de nucleasas (Se utilizó el kit *High Capacity cDNA Reverse Transcription*, ref. 4368814, *Applied Biosystems de Thermo Fisher Scientific*). La reacción se llevó a cabo en un termociclador con los parámetros descritos en la tabla 4. Una vez terminada, cada muestra con cDNA se diluyó 1:4 en agua libre de nucleasas estéril.

Tabla 4. Programa de termociclador para la síntesis de cDNA.

Paso	1	2	3	4
Temperatura	25°C	37°C	80°C	4°C
Tiempo	10 min	120 min	5 min	∞

III. Evaluación de la calidad de cDNA por PCR

Para corroborar que la reacción de transcripción reversa se haya realizado correctamente y que la calidad del cDNA sea buena, se llevó a cabo un ensayo de PCR punto final. Para cada reacción de 10 μL se mezclaron los componentes descritos en la tabla 5. Los oligonucleótidos utilizados iban dirigidos al gen

constitutivo RPS28. En el caso del control negativo, se sustituyó el volumen del templado por agua libre de nucleasas.

Los resultados se observaron mediante electroforesis en un gel de agarosa al 1%. Las bandas de DNA debían ser visibles y observarse íntegras.

Tabla 5. Componentes para la reacción de PCR punto final.

Componente	Volumen para una reacción de 10 µL	Concentración final
Solución amortiguadora 10X	1 µL	1X
MgCl ₂ 50 mM	0.3 µL	1.5 mM
dNTPs 10 mM	0.2 µL	0.2 mM
Oligonucleótido sentido 10 µM	0.2 µL	0.2 µM
Oligonucleótido antisentido 10 µM	0.2 µL	0.2 µM
Templado de DNA	1 µL	25 ng
DNA Polimerasa Taq Platinum	0.04 µL	2 U/reacción
Agua libre de nucleasas	Hasta 10 µL	-

7.3.2.2.3. Validación del método de cuantificación de transcritos por PCR en tiempo real

Para cuantificar los niveles de expresión relativa de los transcritos se llevó a cabo un ensayo de PCR en tiempo real.

I. Determinación de la eficiencia de amplificación de los oligonucleótidos

Con el fin de calcular la eficiencia de amplificación de los oligonucleótidos diseñados, se realizó un ensayo de PCR en tiempo real bajo las siguientes condiciones: cada reacción fue preparada mezclando 5 µL de *SYBR Green/ROX*

qPCR Master Mix (ref. K0223, *molecular biology*, de *Thermo Fisher Scientific*), 1 µL de templado, 0.2 µL de oligonucleótidos sentido 10 µM, 0.2 µL de oligonucleótidos antisentido 10 µM y 3.6 µL de agua libre de nucleasas. Para el control negativo (NTC) se sustituyó el volumen del templado por agua. Cada reacción se llevó a cabo por triplicado técnico.

Para poder llevar a cabo la cuantificación de la eficiencia, era necesario realizar una curva estándar con diferentes concentraciones del templado, por lo que se realizaron diluciones seriales 1:5, 1:25, 1:125, 1:625 del cDNA de la línea celular LNCaP a partir de la dilución inicial 1:4.

Para este ensayo se dio uso al termociclador QuantStudio 3 (*applied biosystems de Thermo Fisher Scientific*) y las condiciones para este equipo se programaron de la manera que se describe en la tabla 6. Los resultados de cada dilución se utilizaron para hacer una regresión lineal y la determinación del valor de la pendiente. La eficiencia fue calculada sustituyendo los valores de la siguiente fórmula:

$$E = (10^{\frac{1}{m}} - 1)(100)$$

Se considera que la eficiencia es óptima para utilizarse en futuros experimentos cuando el valor resultante se encuentra entre 95-105%. Fue este criterio en el que se basó la decisión de qué par de oligonucleótidos sería el mejor para el ensayo de expresión relativa.

Tabla 6. Programación de termociclador para PCR tiempo real.

Etapa de desnaturalización		Etapa de amplificación			Etapa de disociación		
50°C	95°C	95°C	60°C	72°C	95°C	60°C	95°C
2 min	10 min	20 s	20 s	20 s	15 s	1 min	1 s

II. Cuantificación de la expresión relativa del lncRNA mediante el método $\Delta\Delta Ct$

La cuantificación relativa de los transcritos del lncRNA candidato, se llevó a cabo mediante PCR en tiempo real de la misma manera que se describe en el apartado *I. Determinación de la eficiencia de amplificación de los oligonucleótidos.*

Para este ensayo se utilizó el par de oligonucleótidos seleccionado previamente como los de mejor eficiencia, además de un par dirigido al gen constitutivo de referencia RPS28. Se realizaron 4 réplicas biológicas por cada línea celular (LNCaP 1-4 y PrEC 1-4) y 4 réplicas técnicas por gen blanco (gen constitutivo y genes problema). Además, se incluyeron los controles NTC y -RT por triplicado técnico para cada condición. El análisis de resultados se realizó mediante el método $\Delta\Delta Ct$:

$$\begin{aligned}\Delta\Delta Ct &= \Delta Ct (\text{muestra problema}) - \Delta Ct (\text{muestra control}) \\ &= (Ct (\text{gen problema}) - Ct (\text{gen constitutivo}))_{\text{muestra problema}} \\ &\quad - ((Ct (\text{gen problema}) - Ct (\text{gen constitutivo}))_{\text{referencia}})\end{aligned}$$

También se determinó la cuantificación relativa de los cuadruplicados y se calculó el error estándar asociado.

7.3.3. Fase de validación experimental C: análisis de expresión diferencial de muestras de pacientes con recurrencia bioquímica vs. pacientes sin recurrencia bioquímica

Una vez analizada la expresión de lncRNA en las muestras de tejido adyacente y tejido tumoral por RNA-Seq y qPCR, así como las líneas celulares LNCaP (tumoral) y PrEC (normal) con el fin de identificar transcritos que distingan a la condición cancerosa de la sana; se continuó con el análisis enfocado en la búsqueda de biomarcadores para recurrencia bioquímica en CaP.

7.3.3.1. Selección de muestras y criterios de inclusión

Las muestras de pacientes utilizadas para esta fase fueron las mismas de la cohorte independiente descrita en la sección 7.3.1.1. Obtención de muestras y criterios de inclusión (cohorte independiente). En este caso, únicamente se tomaron en cuenta las muestras correspondientes a tejido tumoral (n=16) y fueron agrupadas en dos categorías: el grupo control conformado por las muestras de tejido tumoral de pacientes que no han presentado una recidiva posterior a la prostatectomía radical (n=11); y el grupo problema conformado por las muestras de tejido tumoral de pacientes que presentaron recurrencia bioquímica posterior a la cirugía (n=5). Para esta fase no fue necesario repetir la extracción de RNA total, la síntesis de cDNA ni la PCR en tiempo real, ya que todos esos ensayos ya habían sido realizados.

7.3.3.2. Análisis de expresión diferencial

Para el análisis de expresión diferencial se llevó a cabo el mismo procedimiento descrito en la sección 7.3.1.3.4. Análisis cuantitativo de expresión diferencial.

7.3.4. Análisis estadístico

Con el fin de determinar si los datos generados se distribuían de manera normal, se realizaron pruebas de Shapiro-Wilks y Lilliefors con ayuda del Mat. César Augusto Pérez. Posteriormente, al tratarse de un estudio de casos y controles, se determinó si había, o no, diferencia estadística entre el grupo control y el grupo problema mediante los análisis estadísticos t-student y U de Mann-Whitney no pareadas y con intervalo de confianza de 95% para aquellos datos con distribución normal y distribución no normal, respectivamente.

7.4. Caracterización *in silico* de los lncRNA candidatos

Con el fin de inferir la posible implicación de los lncRNA candidatos en la patología del CaP y de la recurrencia bioquímica, se llevó a cabo su caracterización *in silico* mediante uso de herramientas en línea para determinar su localización genómica, longitud del transcrito, identificación de isoformas, determinación del potencial codificante, sus niveles de expresión en diferentes tejidos, la región subcelular en la

que se acumulan, así como las asociaciones predichas con factores transcripcionales, proteínas y otras moléculas. Los parámetros y herramientas utilizadas para cada uno de los conceptos mencionados se muestran a continuación en la tabla 7.

Tabla 7. Características y parámetros para el análisis *in silico* de los lncRNA.

Concepto	Parámetro	Herramienta
Localización genómica (con base en el genoma hg38)	Regiones intra o intergénicas	<i>LNCipedia, Ensembl, GeneCards.</i>
Longitud del transcrito	Mayores a 200 pb	<i>LNCipedia, Ensembl, GeneCards.</i>
Identificación de isoformas	Presencia	<i>LNCipedia, Ensembl.</i>
Potencial codificante	$X \leq 0$	CNIT
Expresión en tejidos	Expresión diferencial en tejido tumoral prostático	TANRIC, <i>MiTranscriptome</i>
Acumulación subcelular	Acumulación en núcleo o citoplasma	<i>IncATLAS</i>
Asociación a moléculas	Nivel de interacción	<i>Rain, RNAinter</i>

8. RESULTADOS

8.1. Análisis bioinformático

8.1.1. Selección de muestras de pacientes con cáncer de próstata localmente avanzado

Con el fin de encontrar lncRNA con expresión específica de tejido tumoral en pacientes con cáncer de próstata localmente avanzado, se llevó a cabo el análisis bioinformático de datos de RNA-Seq. Para ello se utilizó RNA extraído de 3 pacientes con etapas 2Ta, 2Tc, 3Ta, 3Tb y 3Tc, con una concentración de APE mayor a los 7.5 ng/mL y con puntaje de Gleason de 6, 7 y 9, y cuyas características coincidieran con los criterios de inclusión estipulados en el apartado de *Criterios de inclusión* de la sección de *Metodología*. Posteriormente, a partir del RNA obtenido se hizo secuenciación masiva en paralelo de RNA en donde se obtuvieron los transcriptomas que correspondían a tejido neoplásico y tejido adyacente. Estas muestras se obtuvieron del repositorio de muestras biológicas de urología del INCan. Esta información permitió continuar con los análisis dirigidos y enfocados en una población de pacientes mexicanos con CaP localmente avanzado de alto riesgo sin recurrencia bioquímica.

8.1.2. Integridad y calidad de RNA de las muestras para secuenciación

Los análisis para determinar la calidad del RNA total de las 3 muestras correspondientes al tejido adyacente y las 3 muestras correspondientes al tejido tumoral de los pacientes seleccionados muestran que el RNA total obtenido cumplen con un valor de integridad suficientemente alto en el caso de las muestras del paciente 1 (N1, T1) y del paciente 2 (N2, T2) en las que el valor de RIN (número de integridad de RNA) está por arriba de 7. Por otro lado, en el caso de las muestras N3 y T3 correspondientes al paciente 3, el valor de RIN es cercano a 5, sin embargo, a pesar de que su valor de RIN es menor a 7, el electroferograma muestra que la proporción 2:1 entre las bandas correspondientes al RNA ribosomal 28S y 18S se conserva, y se aprecia que los genes ribosomales aún se encuentran íntegros,

principalmente el RNA ribosomal 28s, por lo que estas muestras también fueron seleccionadas para ser secuenciadas (Figura 11).

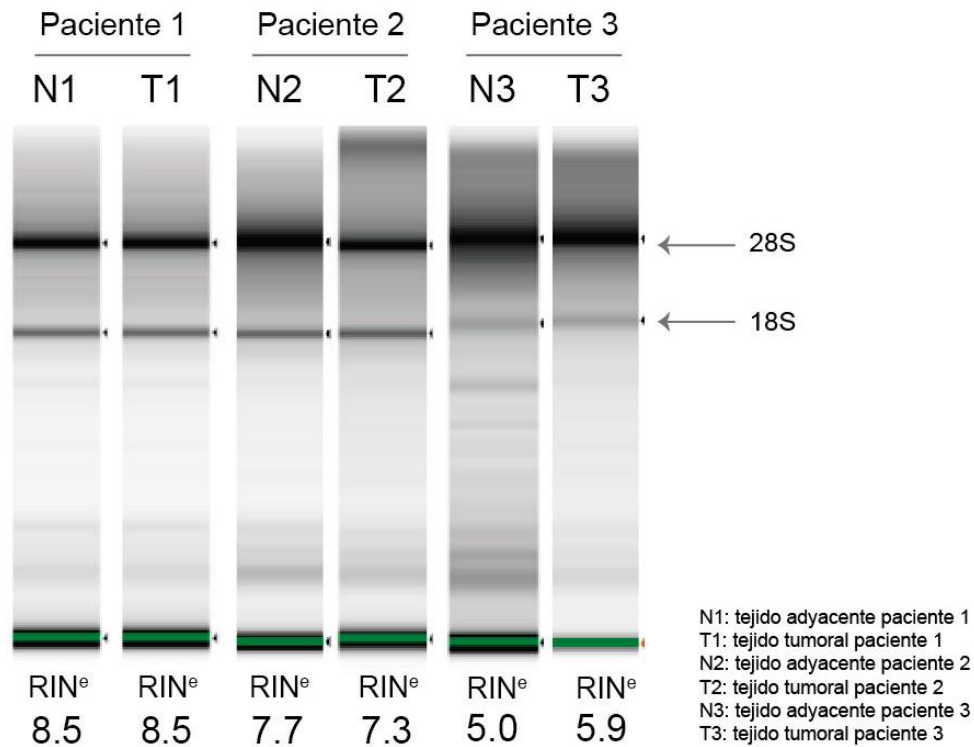


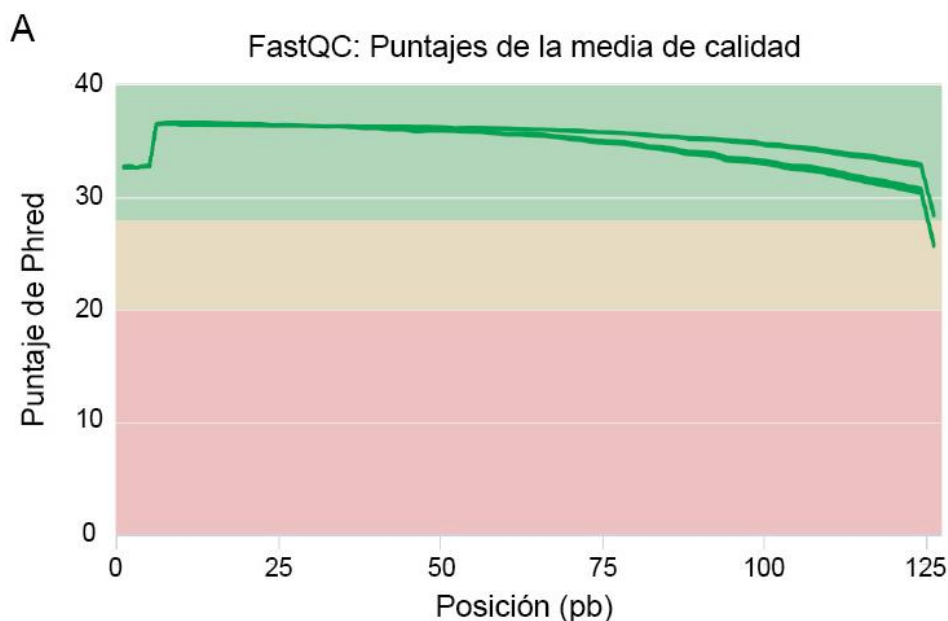
Figura 11. Integridad y calidad del RNA de muestras de pacientes. RNA total extraído a partir de tejido tumoral y adyacente de 3 pacientes. La calidad del RNA fue evaluada en un bioanalizador a partir del cual se obtuvieron los valores de RIN mostrados en la parte inferior de la figura de cada una de las muestras. La letra N corresponde a muestras de tejido adyacente, mientras que la letra T a tejido tumoral. Las muestras N1 y T1 corresponden al primer paciente, N2 junto con T2 y N3 junto con T3, corresponden al segundo y tercer paciente, respectivamente. RIN: valor de integridad del RNA que establece la calidad e integridad de los ácidos nucleicos. La escala tiene un rango del 1 al 10, en donde el 1 indica degradación y 10 indica alta calidad de integridad de RNA.

8.1.3. Selección de datos de RNA-Sec de acuerdo con los criterios de control de calidad para análisis bioinformáticos

Con el fin de continuar con los análisis bioinformáticos de expresión diferencial, se corroboró en primera instancia que la calidad de las lecturas de secuenciación fuera suficiente y cumpliera con los criterios de control para continuar con los análisis bioinformáticos. Para ello, se revisaron los parámetros de calidad evaluados en la plataforma de FastQC así como la evaluación global de los 6 transcriptomas

mediante la herramienta MultiQC (Figura 12). El análisis de calidad global del conjunto conformado por los 6 transcriptomas muestra que cumplen con los criterios de calidad para los análisis bioinformáticos con respecto al parámetro de Phred dado por el valor Q, el cual se relaciona a la probabilidad de error en la detección de las bases secuenciadas. En particular, cuando el valor es mayor o igual a Q20 se considera una calidad *aceptable* (región amarilla) debido a que hubo un error de secuenciación cada 100 bases; por el contrario, cuando se obtiene un valor Q10 o menor (región roja), significa que existe una probabilidad de que haya un error cada 10 bases derivando en que los datos de secuenciación no puedan ser analizados (Figura 12A). De igual manera, se mostró un contenido de N por base aceptable; la presencia de bases N se refiere a la incapacidad del secuenciador de distinguir una base nitrogenada. Cada vez que la base no es reconocida como A, T, C o G es sustituida por una base N, por lo que un alto número de conteos de bases N implica que no se secuenció con certeza ese nucleótido; este parámetro es útil para conocer la precisión con la que se llevó a cabo la secuenciación (Figura 12B).

En resumen, los resultados de la secuenciación de los 6 transcriptomas cumplen con los parámetros de calidad básicos para continuar con su análisis, ya que no se presentaron datos de lecturas con una calidad inadecuada, y todas tuvieron valores por arriba de 20Q, siendo ésta una calidad suficiente y alta para realizar análisis bioinformáticos.



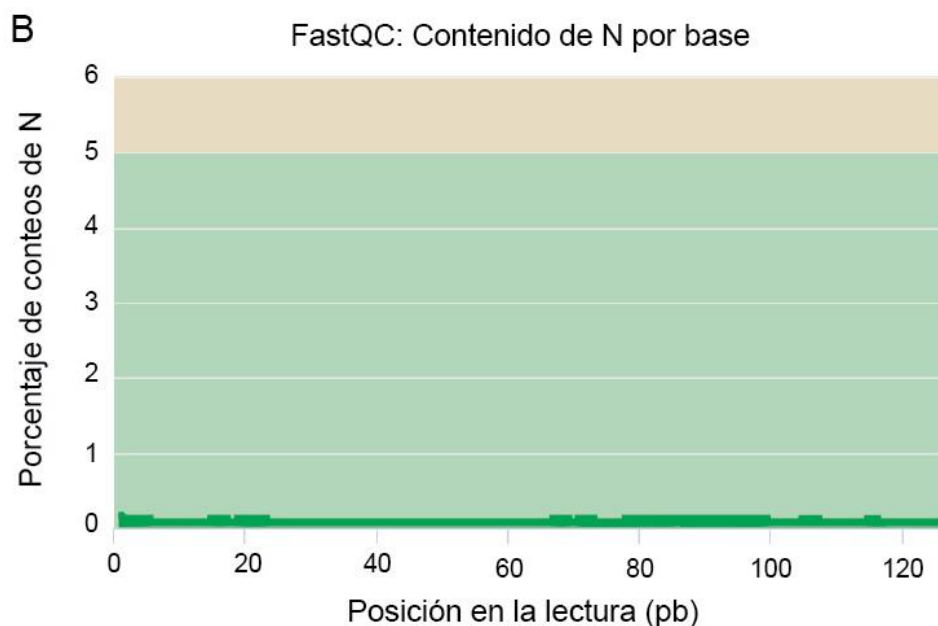


Figura 12. Parámetros de calidad grupales de secuenciación masiva en paralelo de RNA. A) Se muestran los valores promedio del puntaje del parámetro de Phred obtenido de acuerdo a la posición en pares de bases de las lecturas obtenidas. Se observan 3 colores: la zona roja corresponde a valores con calidad inaceptable ya que tiene un 90% de exactitud y una probabilidad de error 1 en 10 correspondiente a la distinción de bases nitrogenadas; la amarilla corresponde a valores con calidad intermedia al contar con un 99% de exactitud y una probabilidad de error de 1 en 100; por último, la zona verde corresponde a aquellos que tienen una calidad considerada como aceptable o alta con un 99.9% de exactitud y una probabilidad de error de 1 en 1,000. **B)** Se muestra el contenido de N por base. En este gráfico se compara el porcentaje de bases N contra la posición en las lecturas secuenciadas, si los valores son cercanos a cero implica que se distinguió con precisión a qué base nitrogenada correspondía la posición en la lectura.

8.1.4. Identificación del perfil de expresión diferencial de lncRNA

Una vez que se evaluó la calidad de los datos de secuenciación, se continuó con el análisis para la determinación de los patrones de expresión diferencial de los transcritos en las muestras de pacientes. Para ello, se compararon los datos del tejido adyacente contra los datos del tejido tumoral mediante el uso de dos flujos de trabajo para análisis bioinformático (*pipelines*) distintos, uno de ellos con la herramienta STAR (por duplicado) y el otro con *Salmon*.

Por otra parte, el PCA muestra la agrupación de los 6 transcriptomas por valores de expresión génica, tomando en cuenta los componentes principales 1 y 2 (Figura 13). Se observó que evaluar el transcriptoma completo (mRNA, lncRNA, miRNA, pseudogenes, etc.) sin hacer distinción entre los diferentes transcritos, no

se generan agrupaciones definidas (Figura 13A). Es por esto que se llevó a cabo un segundo análisis en donde se evaluaron únicamente a los transcritos con expresión diferencial significativa y, además, se separaron en biotipos, es decir, RNA mensajeros y RNA largos no codificantes. Con respecto a estos análisis, el PCA permitió observar una agrupación definida en ambos paneles, es decir, en el panel de mRNA (Figura 13B) y principalmente en el panel de lncRNA (Figura 13C). En consecuencia, se concluye que el perfil de expresión de lncRNA permite la distinción entre muestras de tejido tumoral y muestras de tejido adyacente.

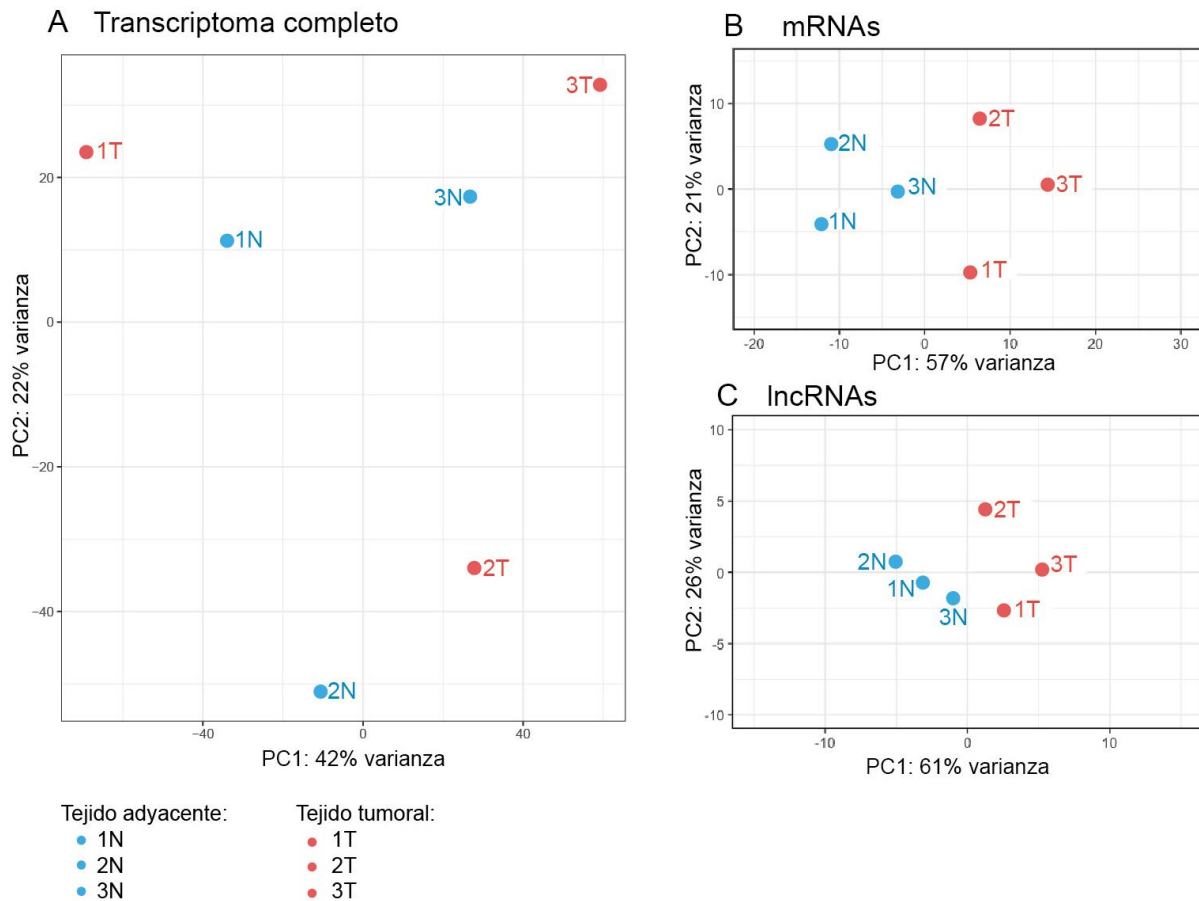


Figura 13. Análisis de componentes principales (PCA) de los resultados obtenidos por RNA-Seq. A) Gráfico de PCA correspondiente al transcriptoma completo. Se observa que los niveles de expresión de todos los transcritos no distingue los grupos de tejido adyacente y de tejido tumoral con la varianza de PC1 y PC2. **B)** Gráfico de PCA correspondiente a los transcritos codificantes con expresión diferencial significativa. Se observa que la varianza en los niveles de expresión de los mRNA distingue a los grupos de tejido adyacente y de tejido tumoral con los porcentajes de PC1 y PC2. **C)** Gráfico de PCA correspondiente a los lncRNA con expresión diferencial significativa. Se observa que la varianza en los niveles de expresión de los lncRNA distingue a los grupos de tejido adyacente y de tejido tumoral con los porcentajes de PC1 y PC2. En color azul se resaltan las muestras de tejido adyacente y en rojo las muestras de tejido tumoral; n=3.

Con el fin de identificar lncRNA no codificantes con expresión específica de CaP, se llevó a cabo un análisis de expresión diferencial en el que se compararon muestras de tejido adyacente y de tejido tumoral. Para ello se siguió un *pipeline* bioinformático mediante las herramientas STAR y DESeq2, del cual se obtuvieron un gráfico de volcán (Figura 14A) y de pastel (Figura 14B). A partir de este análisis se identificaron 8 lncRNA con expresión diferencial estadísticamente significativa. Dentro de éstos, 5 se encontraban sobreexpresados y 3 subexpresados de acuerdo al punto de corte establecido (\log_2FC -1.5 y 1.5; $FDR \leq 0.05$). Esto indica la existencia de un perfil de expresión de lncRNA específico y diferencial en el grupo de muestras de tejido tumoral de próstata.

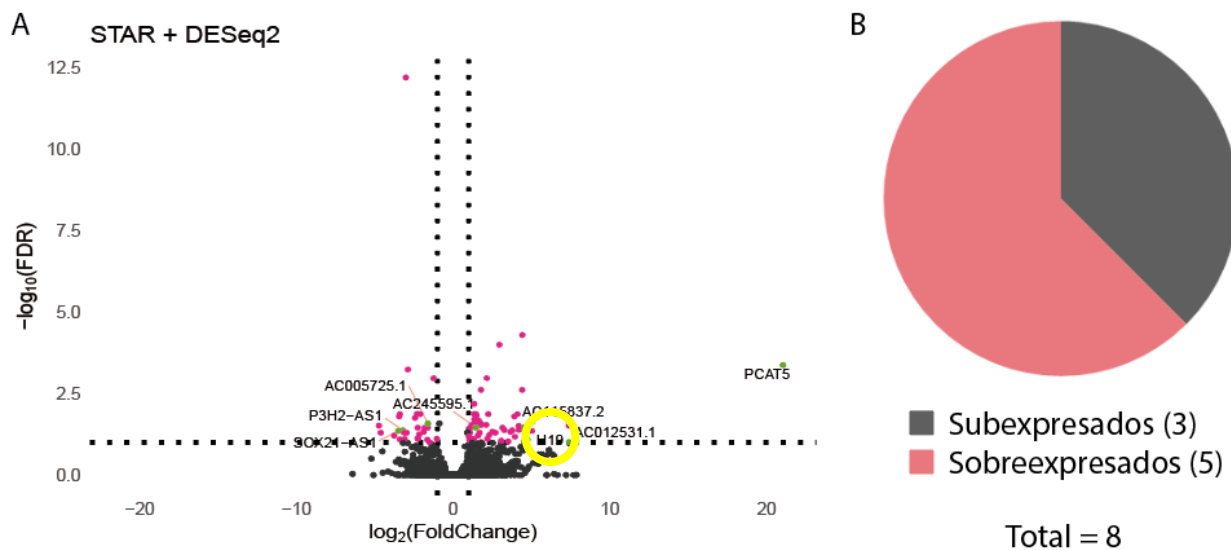


Figura 14. Identificación de lncRNA con expresión diferencial. A) Gráfico de volcán en donde se muestran los lncRNA con expresión diferencial identificados en el tejido tumoral distinguidos en color verde en los cuadrantes superiores derecho e izquierdo. **B)** Gráfico de pastel que representa la parte proporcional de los lncRNA subexpresados y sobreexpresados de los 8 totales. Punto de corte: -1.5 y 1.5; $n=3$; $FDR < 0.05$.

Posteriormente, se generó un mapa de calor con el fin de visualizar los patrones de expresión de mRNA y de transcritos no codificantes mediante un análisis de agrupación jerárquica (Figura 15). Cada columna representa el tipo de muestra evaluada, ya sea tejido tumoral o adyacente; mientras que las filas representan cada uno de los transcritos. En este gráfico se observa la clara distinción entre los dos tipos de muestras, ya que en el mapa de calor se separan

el tejido adyacente y el tejido tumoral en dos grupos diferentes, lo que sugiere que existe un panel de lncRNA que es específico de CaP y que, por lo tanto, distingue al tejido tumoral del tejido normal.

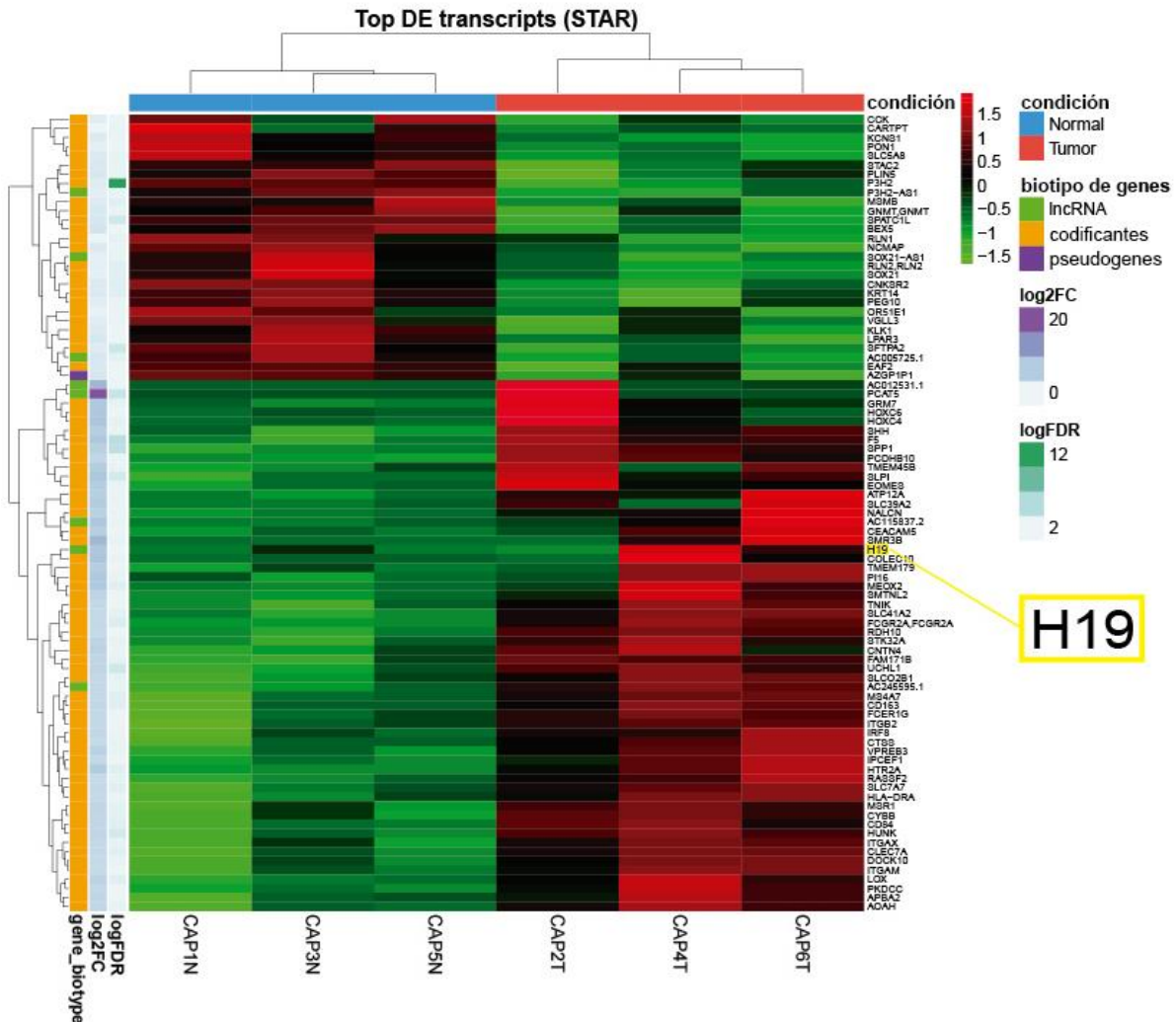


Figura 15. La agrupación jerárquica de los transcritos distingue al tejido tumoral del tejido normal en muestras de pacientes con cáncer de próstata. El mapa de calor generado a partir del análisis de expresión diferencial permite observar la agrupación jerárquica de los transcritos secuenciados con expresión diferencial y observar la distinción entre tejido normal (azul) y tejido tumoral (rosa) colocados en columnas. Este gráfico representa a todo el transcriptoma, incluyendo lncRNA (verde claro), transcritos codificantes (amarillo) y pseudogenes (morado) colocados en las filas. Punto de corte: 1.5 y -1.5; n=3; FDR<0.05.

A partir del gráfico de volcán (Figura 14A) y del mapa de calor (Figura 15) se pudieron distinguir los lncRNA que conforman al perfil de expresión diferencial y que, por lo tanto, son específicos de la condición tumoral de CaP. En la tabla 8 se

muestra la lista de lncRNA, su localización genómica, la tasa de cambio en su expresión y el valor de *p* obtenido a partir del análisis bioinformático de los datos de RNA-Seq.

Tabla 8. Panel de lncRNA sobreexpresados y subexpresados obtenido por RNA-Seq.

Símbolo	Localización genómica	Log2FoldChange	P value
PCAT5	chr10: 35,778,302-35,800,920	21.04476082	7.3287E-08
AC012531.1	chr12: 54,019,910-54,022,589	7.413918943	0.00037518
H19	chr11:1,995,130-2,001,710	4.737912466	0.00028065
AC115837.2	chr8: 74,609,698-74,633,320	4.526410272	5.9364E-05
AC245595.1	chr1: 144,276,484-144,305,180	1.457637593	6.0488E-05
AC005725.1	chr17: 11,288,205-11,290,811	-1.60469137	3.2817E-05
P3H2-AS1	chr3:190,120,952-190,150,133	-3.20140632	9.0329E-05
SOX21-AS1	chr13:94,703,454-94,803,430	-3.48759722	9.5497E-05

Por último, se realizó un diagrama de Venn con el fin de observar los lncRNA con expresión diferencial significativa presente en el primer y segundo análisis con la herramienta STAR, y en el análisis con la herramienta Salmon. Basado en esto, se seleccionaron a los transcrito de H19 y SOX21-AS1 como candidatos para continuar con la validación en tejido tumoral de pacientes con CaP localmente avanzado, ya que son los únicos lncRNA representado en los dos *pipelines* bioinformáticos realizados con STAR y Salmon (Figura 16).

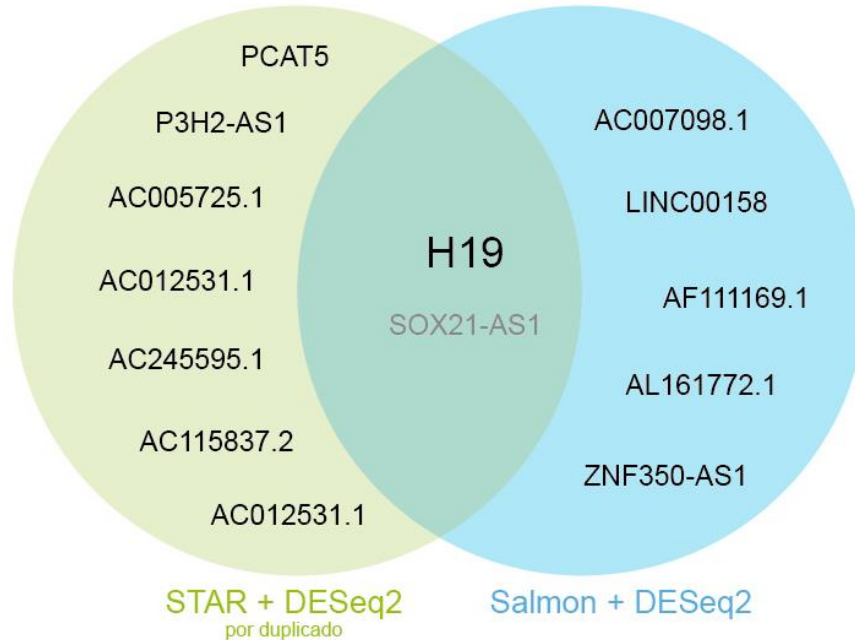


Figura 16. Análisis de convergencia de genes entre 2 *pipelines* bioinformáticos independientes. Diagrama de Venn mostrando la convergencia entre transcritos con expresión diferencial significativa derivados del análisis de expresión diferencial de los datos de RNA-Sec. En color azul se muestra la lista de genes obtenidos por dos análisis independientes mediante la herramienta STAR y DESeq2, mientras que en color gris se muestra la lista obtenida mediante Salmon y DESeq2. El lncRNA H19 está presente en los 2 grupos abordados.

8.2. Validación experimental

Con el objetivo de validar experimentalmente los datos observados en el análisis de RNA-Sec, se realizaron ensayos cuantitativos por reverso transcriptasa acoplada a PCR en tiempo real tanto en líneas celulares de tejido prostático y CaP, así como en muestras de tejido de pacientes con CaP localmente avanzado de una cohorte de 16 pacientes independiente a la estudiada en la fase de descubrimiento pero con las mismas características clínico-patológicas, adicionalmente se incluyeron muestras de pacientes que presentaban recurrencia bioquímica.

8.2.1. Análisis de integridad y calidad del RNA total para la fase de validación

El RNA total correspondiente a las 16 muestras de tejido adyacente y a las 16 de tumoral de los pacientes seleccionados cumplen con los criterios de calidad para los análisis de expresión, ya que el valor de RIN (número de integridad de RNA) es

mayor o igual a 7 (Figura 17). A partir de este análisis se concluye que las muestras de RNA extraídas eran adecuadas para el análisis del arreglo de expresión de lncRNA.

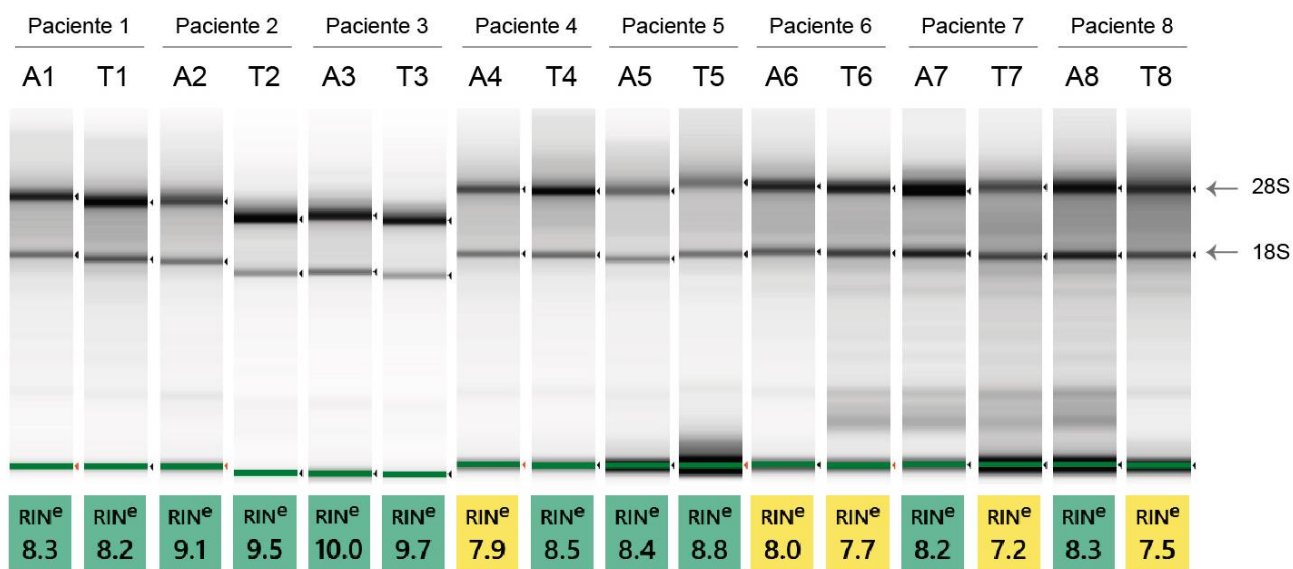


Figura 17. Integridad y calidad del RNA total obtenido de muestras de tejido de pacientes. Imagen representativa del RNA total extraído a partir de tejido tumoral y adyacente de 16 pacientes con CaP localmente avanzado. La calidad del RNA total fue evaluada en un bioanalizador a partir del cual se obtuvieron los valores de RIN mostrados en la parte inferior de la figura. RIN: valor de integridad del RNA que establece la calidad e integridad de los ácidos nucleicos. La escala va del 1 al 10, donde 1 indica degradación y 10 indica alta calidad e integridad. A: tejido adyacente. T: tejido tumoral.

8.2.2. Determinación de un panel de lncRNA con expresión diferencial en muestras de pacientes con cáncer de próstata localmente avanzado

Para continuar con los ensayos de validación experimental, así como corroborar los análisis bioinformáticos y obtener un conjunto de lncRNA con expresión diferencial en pacientes con CaP localmente avanzado de alto riesgo, se realizó un ensayo cuantitativo por RT-qPCR mediante el arreglo de expresión de lncRNA de la marca comercial QIAGEN *RT² lncRNA qPCR Assay (Human Cancer PathwayFinder)*. Este arreglo contiene los 84 principales lncRNA reportados en diferentes tipos de cáncer, incluido a H19, el cual se encontró sobreexpresado en el análisis de expresión diferencial por RNA-Seq anteriormente descrito.

La plataforma de QIAGEN utilizada para el análisis de este arreglo de expresión se encuentra disponible en línea (<https://dataanalysis2.qiagen.com/lncRNA>). En esta plataforma se genera un reporte para cada lncRNA analizado junto con un conjunto de gráficos que incluye un mapa de agrupación jerárquica, un gráfico de volcán y un gráfico de barras. El perfil de expresión de lncRNA se muestran en la figura 18. A partir de los resultados globales del ensayo por RT-qPCR del arreglo de expresión de RT², se observó que existe un conjunto de lncRNA que se expresa de manera diferencial en muestras de pacientes con CaP localmente avanzado cuando se compara contra el tejido adyacente, además de que este panel se conforma por 22 genes subexpresados y 15 sobreexpresados (Figura 18).

Conformando al panel de expresión diferencial, se encuentra el gen específico de CaP PCA3 sobreexpresado. Como se mencionó con anterioridad, PCA3 es un lncRNA con sobreexpresión específica de CaP utilizado actualmente en la clínica como biomarcador, lo cual indica que los datos observados son confiables, que corresponden a un perfil específico de CaP, y que las muestras analizadas corresponden a tumores.

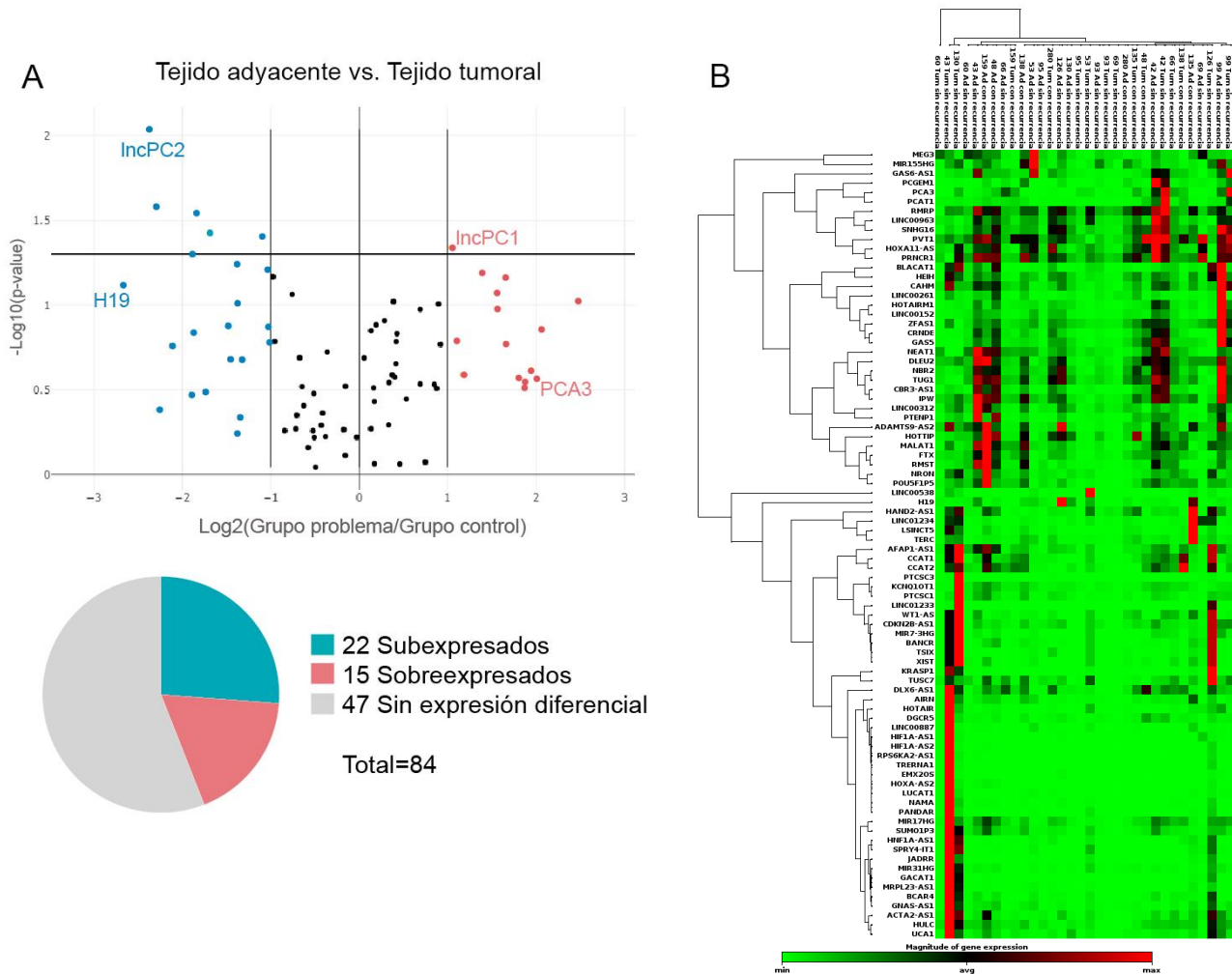


Figura 18. Panel de expresión diferencial de lncRNA en muestras de pacientes con CaP localmente avanzado comparando tejido adyacente vs. tejido tumoral. Análisis de expresión diferencial llevado a cabo mediante ensayos de RT-qPCR en el arreglo de expresión RT² para lncRNA. **A)** En la parte superior se muestra un gráfico de volcán en donde se señalan los lncRNA subexpresados en color azul y los sobreexpresados en color rosa. Todos los transcritos representados en el cuadrante superior izquierdo y derecho presentan una expresión diferencial estadísticamente significativa, mientras que los representados en los cuadrantes inferior izquierdo y derecho únicamente presentaron una tasa de cambio alta respecto al grupo control (tejido adyacente). En la parte inferior se muestra un gráfico de pastel representativo de la proporción de lncRNA subexpresados (azul), sobreexpresados (rosa) y sin expresión diferencial (gris) en el análisis. **B)** Mapa de agrupación jerárquica en el que se presenta la relación de lncRNA de las muestras por similitud en sus perfiles de expresión (columnas), así como el nivel de expresión de cada lncRNA (filas); las tonalidades cercanas a verde representan una subexpresión, mientras que las tonalidades cercanas a rojo representan una sobreexpresión. Puntos de corte Log2FC: 2 y -2; n=16; *pvalue*<0.05.

Asimismo, los análisis de expresión diferencial muestran que el lncRNA H19 se subexpresa al comparar tejido neoplásico y tejido adyacente, contrario a lo observado en los análisis de expresión diferencial de RNA-Sec, donde H19 se encuentra sobreexpresado. Sin embargo, en esta fase de validación H19 se encuentra encabezando la lista de los lncRNA subexpresados (Figura 19) y no sobreexpresados como se observó en la cohorte inicial (Tabla 8). En conjunto, estos resultados indican que existe un conjunto de lncRNA con expresión diferencial específica de las muestras de CaP tumorales, dentro de éste se encuentran PCA3, LNCPC1, LNCPC2 y H19, el cual fue validado bioinformáticamente y experimentalmente.

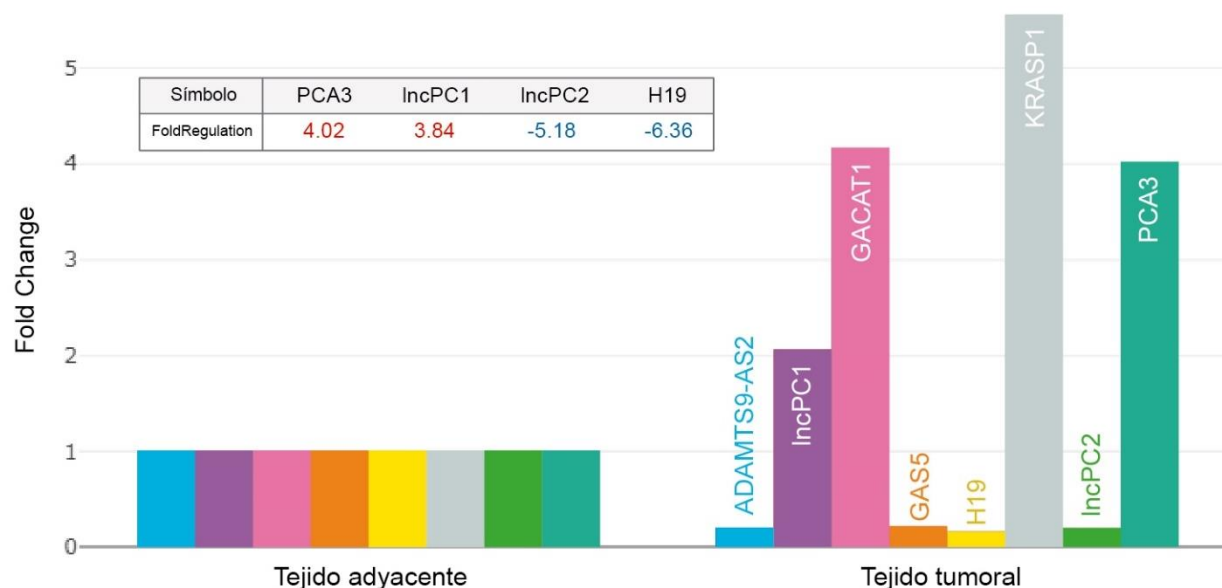


Figura 19. El panel de lncRNA distingue al tejido tumoral del tejido adyacente en muestras de cáncer de próstata localmente avanzado. Gráfico de barras en el que se compara la expresión relativa (*Fold Change**) de los 8 lncRNA con mayor tasa de cambio, la cual fue normalizada con el gen constitutivo control RNS7K en tejido adyacente y tejido tumoral. La tabla de la esquina superior izquierda muestra los valores de tasa de cambio (*Fold Regulation***) de los lncRNA que conforman el panel. **Fold Change*: valor calculado a partir de la tasa de expresión relativa entre el grupo control de tejido adyacente y el grupo problema de tejido tumoral. Los números mayores a 1 indican sobreexpresión; entre 0 y 1, subexpresión; y 1 que no hubo cambio. ***Fold Regulation*: inverso negativo del valor *Fold Chage*.

Posteriormente, se identificó que los lncRNA ADAMTS9-AS2, GACAT1, GAS5, H19, KRASP1, LNCPC2, PANDAR y PCA3 se encuentran dentro de los 10 genes con mayor sobreexpresión o subexpresión en las pruebas realizadas con el arreglo de expresión RT² (Figura 19). Estos datos confirman al panel de lncRNA

específico de tejido tumoral mencionado anteriormente. Finalmente, se llevó a cabo un análisis de conjuntos mediante diagramas de Venn con el fin de observar de manera sintetizada las coincidencias de genes no codificantes con expresión diferencial entre los análisis bioinformáticos hechos a partir de los datos de RNA-Sec y los ensayos experimentales del arreglo de expresión RT² (Figura 20A). En el diagrama de Venn se observó que únicamente existe una coincidencia entre los transcritos de los datos de RNA-Sec y las qPCRs individuales. Los resultados muestran que H19 está presente en ambos ensayos, el análisis por secuenciación masiva en paralelo de RNA y el análisis por RT-qPCR mediante el arreglo de expresión RT² (Figura 20B).

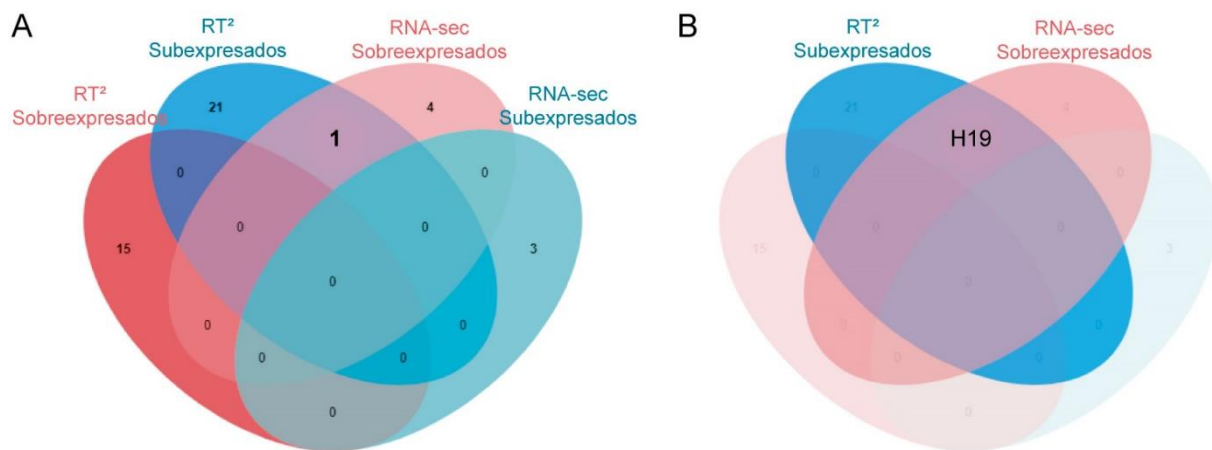


Figura 20. El lncRNA H19 presenta expresión diferencial en el análisis de RNA-Sec y en el arreglo de expresión RT² de PCR en tiempo real. A) Diagrama de Venn representativo de los lncRNA sobreexpresados (rojo) y subexpresados (azul) obtenidos a partir del análisis de RNA-Sec (derecha) y del arreglo de expresión por qPCR (izquierda). **B)** Diagrama de Venn que representa los lncRNA sobreexpresados (rojo) del análisis de RNA-Sec y su intersección con los lncRNA subexpresados (azul) del análisis del arreglo de expresión por qPCR RT². H19 es el único lncRNA representado en ambos ensayos. Diagrama realizado en jvenn¹³².

A partir del análisis de expresión diferencial se puede concluir que existe un conjunto de lncRNA que no ha sido reportado previamente, que se expresa diferencialmente en muestras de tejido tumoral y que las distingue del tejido

adyacente. PCA3 y LNCPC1 se sobreexpresan en muestras de pacientes con CaP localmente avanzado, mientras que H19 y LNCPC2 se subexpresan.

8.2.3. Validación de la expresión del lncRNA H19 en líneas celulares de cáncer de próstata por RT-qPCR

Continuando con la validación experimental, se cuantificaron los niveles de expresión relativa del transcrito H19 en las líneas celulares de próstata PrEC (no neoplásica, tejido calibrador) y LNCaP (neoplásica, tejido problema) con el objetivo de establecer si en el modelo biológico de líneas celulares de próstata se presentaba el mismo patrón de expresión observado en los resultados de RNA-Seq provenientes de muestras de pacientes con CaP localmente avanzado. Para ello se realizó un ensayo de RT-qPCR (Figura 21). De acuerdo a la expresión cuantificada en la línea celular calibradora PrEC, el lncRNA H19 presenta una subexpresión en la línea celular neoplásica LNCaP.

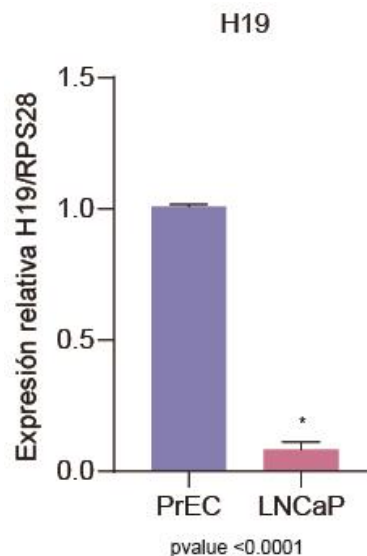


Figura 21. Expresión relativa del lncRNA candidato H19 en las líneas celulares PrEC y LNCaP. Gráfico de barras generada a partir del análisis de expresión relativa de H19 por el método $\Delta\Delta Ct$. En ella se observa que H19 se subexpresa en LNCaP ($p < 0.0001$) a comparación de su expresión en PrEC. En color azul se presentan los resultados obtenidos a partir de PrEC mientras que en color rosa se presentan los resultados obtenidos a partir de LNCaP. Para la RT-qPCR se realizaron cuadruplicados técnicos y cuadruplicados biológicos ($n=4$). Los datos se analizaron mediante la prueba t-student no pareada y se utilizó al gen constitutivo RPS28 como gen control para la normalización de los niveles de expresión. Gráfica realizada en *GraphPad Prism*¹³³.

En resumen, los resultados corroboran que H19 no sólo se subexpresa en muestras de tejido obtenido a partir de prostatectomía radical, sino que también presenta la misma tendencia en el modelo celular neoplásico de próstata, ya que, al comparar sus niveles de expresión contra la línea celular sana PrEC, se observó que también se subexpresa en la línea celular neoplásica LNCaP. Esto sugiere que el lncRNA H19 se subexpresa tanto en tejidos tumorales de pacientes como en las líneas celulares neoplásicas.

8.2.4. Determinación de un panel de lncRNA asociado al pronóstico de recurrencia bioquímica en muestras de pacientes con cáncer de próstata localmente avanzado de alto riesgo

Una vez que se encontraron lncRNA con expresión específica de tejido tumoral de pacientes con CaP (PCA3, LNCPC1, H19 y LNCPC2), se buscó determinar si existían lncRNA asociados específicamente al pronóstico de los pacientes; en particular, se evaluaron pacientes con CaP localmente avanzado de alto riesgo que presentaron recurrencia bioquímica. Para ello, se analizaron los datos de expresión diferencial únicamente de las muestras correspondientes a tejido tumoral, para así comparar las muestras de tejido tumoral de pacientes con recurrencia bioquímica y las de aquellos pacientes que no presentaron recurrencia bioquímica.

A partir del análisis cuantitativo, se identificó un conjunto de lncRNA que se expresa diferencialmente en muestras de tejido tumoral con recurrencia bioquímica. Dentro de los resultados, se observó un patrón de expresión característico entre las muestras de tejido de pacientes con recurrencia bioquímica, además de que la proporción de lncRNA sobreexpresados es cercana a la de lncRNA subexpresados, contando con 31 transcritos subexpresados, 23 sobreexpresados y 30 sin expresión diferencial (Figura 22).

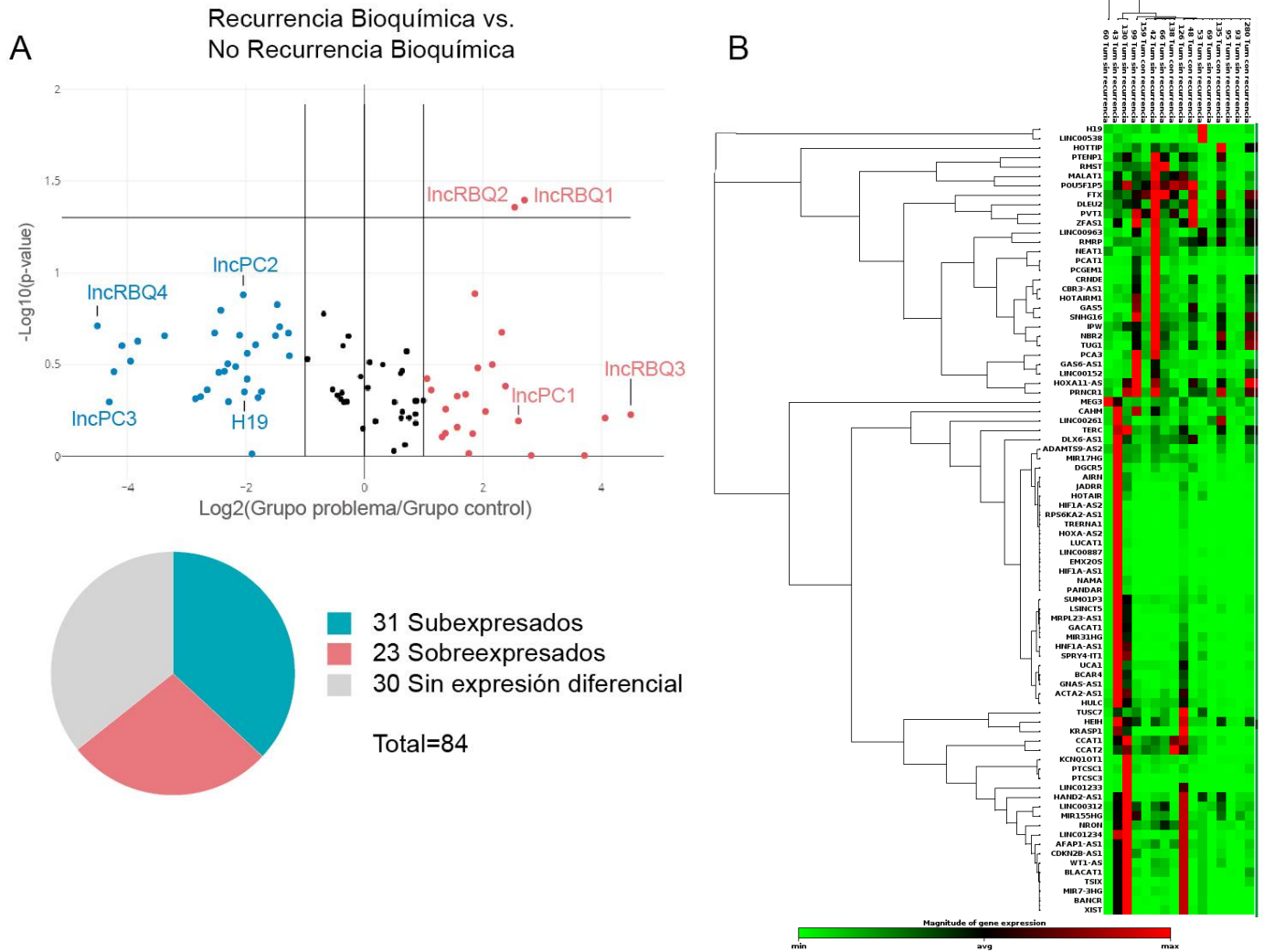


Figura 22. Panel de expresión diferencial de lncRNA en muestras de tejido tumoral de pacientes con CaP localmente avanzado que presentan recurrencia bioquímica. Análisis de expresión diferencial llevado a cabo mediante ensayos de RT-qPCR en el arreglo de expresión RT² de lncRNA. **A)** En la parte superior se muestra un gráfico de volcán en donde se señalan los lncRNA subexpresados en color azul y los sobreexpresados en color rosa. Todos los transcritos representados en el cuadrante superior izquierdo y derecho presentan una expresión diferencial estadísticamente significativa, mientras que los representados en los cuadrantes inferior izquierdo y derecho únicamente presentaron una tasa de cambio alta respecto al grupo control (tejido tumoral de pacientes sin recurrencia bioquímica). En la parte inferior se muestra un gráfico de pastel representativo de la proporción de lncRNA subexpresados (azul), sobreexpresados (rosa) y sin expresión diferencial (gris) en el análisis. **B)** Mapa de agrupación jerárquica en el que se presenta la relación de lncRNA de las muestras por similitud de perfiles de expresión (columnas), así como el nivel de expresión de cada lncRNA (filas); las tonalidades cercanas a verde representan una subexpresión, mientras que las cercanas a rojo representan una sobreexpresión. Puntos de corte Log₂FC: 2 y -2; n=11 sin recurrencia, n=5 con recurrencia; *pvalue*<0.05.

Dentro de los lncRNA que conforman a este panel se encuentran LNCRBQ1, LNCRBQ2, LNCRBQ3, LNCRBQ4, LNCPC3, y, por segunda vez, LNCPC1, LNCPC2 y H19. Para ello, se diseñó un gráfico de barras que representara la expresión relativa de los transcritos LNCRBQ1, LNCPC1, LNCPC2, LNCRBQ2, H19, LNCRBQ3, LNCPC3 y LNCRBQ4, los cuales se encuentran dentro de los 10 genes con mayor sobreexpresión o subexpresión dentro de los resultados obtenidos (Figura 23).

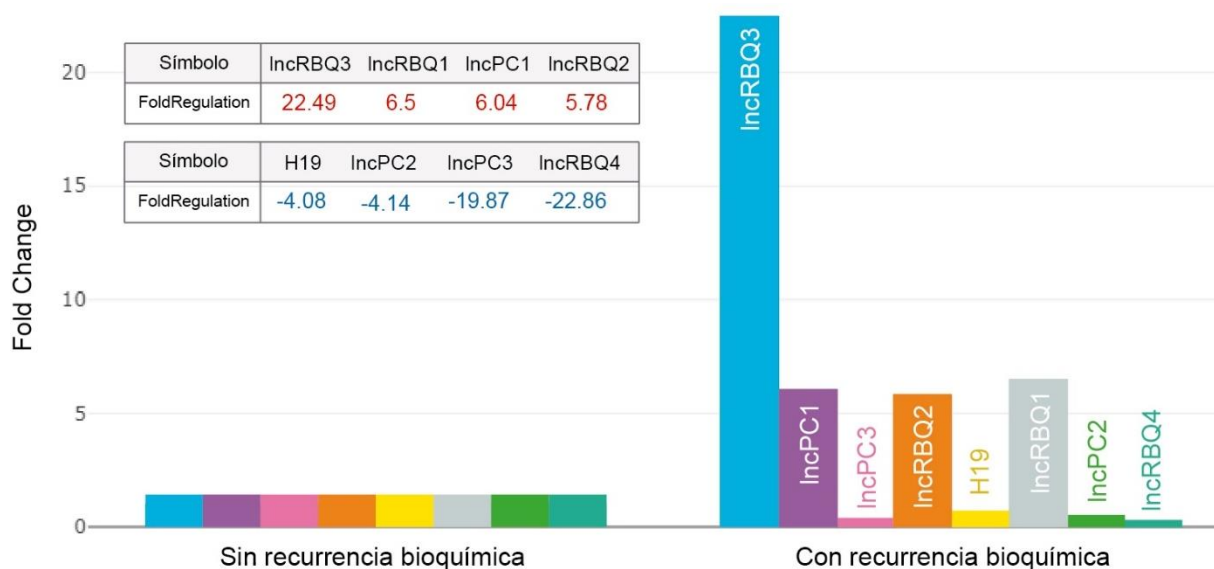


Figura 23. El panel de lncRNA define al tejido tumoral de pacientes con recurrencia bioquímica del tejido de pacientes sin recurrencia bioquímica. Gráfico de barras en el que se compara la expresión relativa (*Fold Change**) de los 8 lncRNA con mayor tasa de cambio, la cual fue normalizada con el gen constitutivo control RN7SK en tejido tumoral de pacientes con y sin recurrencia bioquímica. La tabla de la esquina superior izquierda muestra los valores de tasa de cambio (*Fold Regulation*** de los lncRNA que conforman el panel. **Fold Change*: valor calculado a partir de la tasa de expresión relativa entre el grupo control y el grupo problema, los números mayores a 1 indican sobreexpresión; entre 0 y 1, subexpresión; y 1 que no hubo cambio. ***Fold Regulation*: inverso negativo del valor *Fold Change*.

Posteriormente, se procedió a analizar los datos individuales de cada lncRNA que integra el panel de expresión diferencial, esto se realizó mediante el análisis de expresión relativa por ΔCt (Figura 24). A partir de estos resultados puede observarse la tendencia de cada uno de los lncRNA a sobreexpresarse, en el caso de LNCRBQ3, LNCPC1, LNCRBQ1 y LNCRBQ2, y a subexpresarse en el caso de LNCRBQ4, LNCPC3, H19 y LNCPC2 dentro de las muestras de tejido tumoral de pacientes con recurrencia bioquímica.

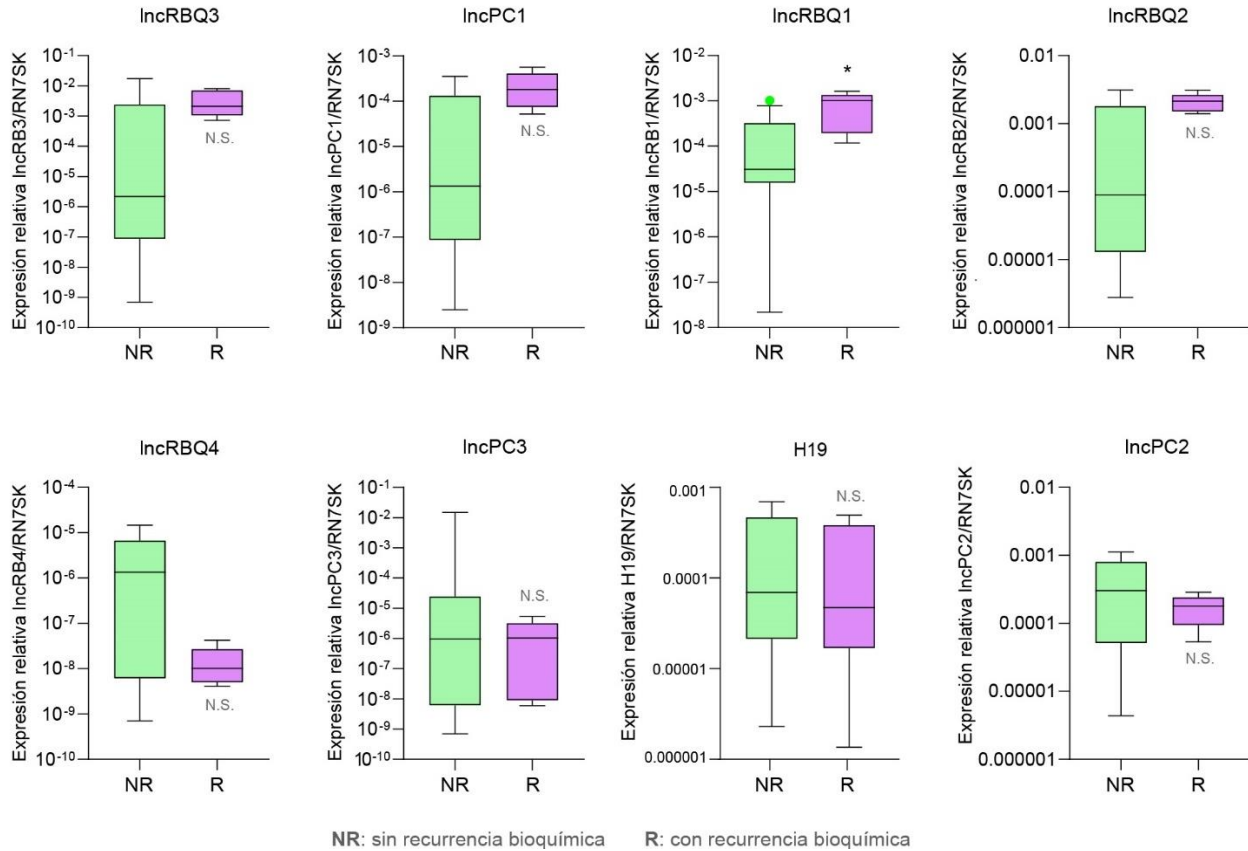


Figura 24. Los lncRNA candidatos que conforman al panel tienden a la sobreexpresión y subexpresión en muestras de pacientes con recurrencia bioquímica. Gráficas de caja y bigotes generadas a partir del análisis de expresión relativa de los lncRNA del panel de recurrencia bioquímica por el método ΔCt . En ella se observa la subexpresión o sobreexpresión de los lncRNA en muestras de tejido de pacientes con recurrencia bioquímica (R) a comparación de su expresión en tejido de pacientes sin recurrencia bioquímica (NR). En color verde se presentan los resultados obtenidos a partir del tejido de pacientes sin recurrencia, mientras que en color morado se presentan los resultados obtenidos a partir del tejido de pacientes con recurrencia. Los datos se analizaron mediante la prueba no paramétrica U de Mann-Whitney no pareada y se utilizó al gen constitutivo RN7SK como gen calibrador para la normalización; n=11 sin recurrencia, n=5 con recurrencia; $pvalue < 0.05$; N.S.: no significativo. Gráficas realizadas en *GraphPad Prism*¹³³.

En conclusión, estos resultados y los obtenidos anteriormente, indican que existe una firma de lncRNA con expresión diferencial en muestras de tejido tumoral de pacientes con recurrencia bioquímica que no se ha reportado con anterioridad y podría tener potencial como biomarcador de pronóstico de utilidad clínica. Este grupo de lncRNA se conforma por LNCRBQ1, LNCPC1, LNCRBQ2, LNCRBQ3, LNCRBQ4, LNCPC2, LNCPC3 y H19.

8.2.5. Análisis de conjuntos

Con el objetivo de distinguir cuáles de los lncRNA pertenecientes al panel de expresión diferencial tienen expresión tumor-específica y cuáles tienen expresión específica de recurrencia bioquímica, se llevó a cabo un análisis de conjuntos comparando todos los lncRNA con tasas de cambio altas dentro de los dos enfoques evaluados en este proyecto: 1) tejido adyacente vs. tejido tumoral y 2) tejido tumoral de pacientes con recurrencia bioquímica vs. tejido tumoral de pacientes sin recurrencia bioquímica. Para ello se hizo un diagrama de Venn y se identificaron las disyuntivas entre lncRNA (Figura 25A). El diagrama de Venn muestra que hay 4 lncRNA que tienen expresión tumor-específica: LNCPC1 sobreexpresado, y H19, LNCPC2 y LNCPC3 subexpresados (Figura 25B). Este conjunto de genes no-codificantes no se había reportado antes como una firma de lncRNA con expresión específica en tejido tumoral de CaP localmente avanzado de alto riesgo. Por el contrario, se identificaron 4 lncRNA con expresión específica de recurrencia bioquímica: LNCRBQ1, LNCRBQ2 y LNCRBQ3 sobreexpresados y LNCRBQ4 subexpresado (Figura 25C); de igual manera, este panel tampoco se ha reportado como una firma de lncRNA con expresión específica de tejido tumoral de pacientes con recurrencia bioquímica.

Este enfoque permitió la identificación de aquellos lncRNA que podrían tener importancia clínica, específicamente, en el estudio de recurrencia bioquímica, la cual puede presentarse en pacientes con CaP que hayan sido sometidos a prostatectomía radical. A partir de este análisis se definieron dos paneles de lncRNA nunca antes reportados, el primero de ellos con expresión específica de CaP localmente avanzado de alto riesgo (LNCPC1, H19, LNCPC2 Y LNCPC3) y el segundo con expresión específica de recurrencia bioquímica (LNCRBQ1, LNCRBQ2, LNCRBQ3 Y LNCRBQ4).

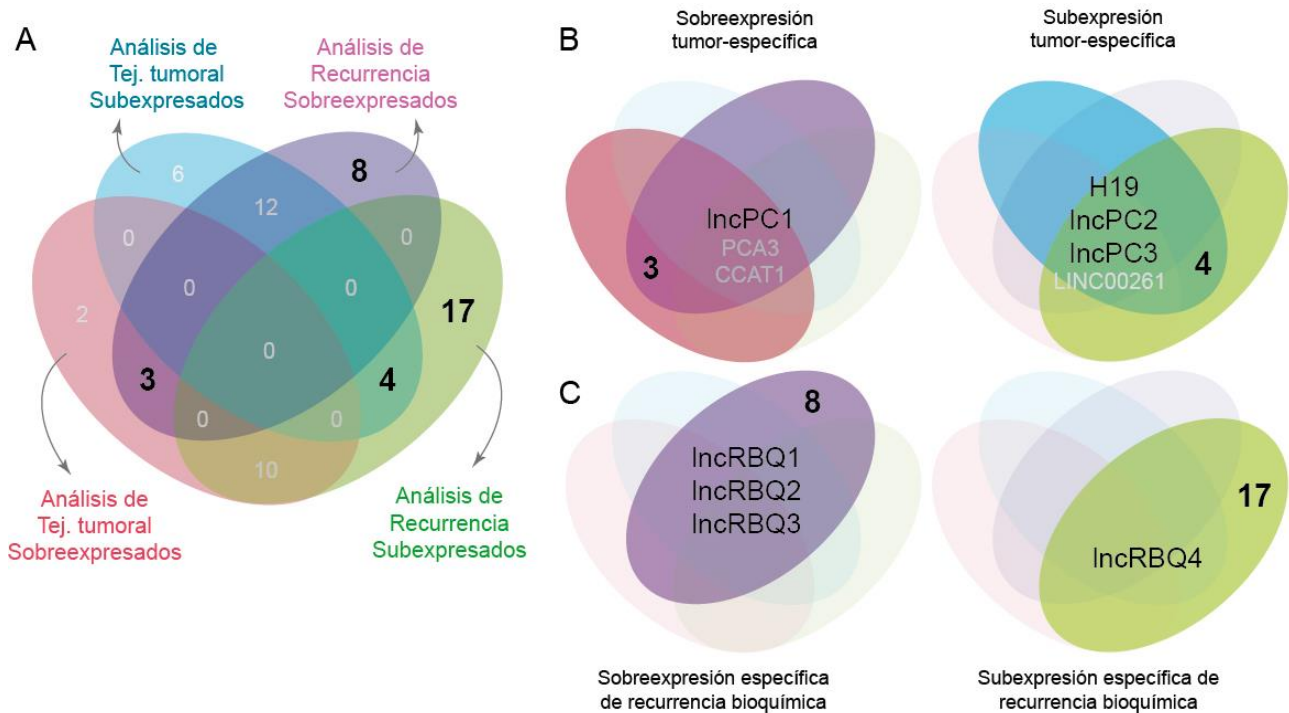


Figura 25. Existen coincidencias y divergencias en los lncRNA con expresión diferencial entre el análisis de tej. adyacente vs. tej. tumoral al compararlo con el análisis de recurrencia bioquímica vs. no recurrencia bioquímica. **A)** Diagrama de Venn representativo de los transcritos sobreexpresados (rosa y morado) y subexpresados (azul y verde) obtenidos a partir del análisis de tejido tumoral con CaP (derecha) y del análisis de recurrencia bioquímica (izquierda). **B)** Diagrama de Venn representativo de los transcritos sobreexpresados y subexpresados que convergen entre ambos análisis. **C)** Diagrama de Venn representativo de los transcritos sobreexpresados y subexpresados únicamente en el análisis de recurrencia bioquímica. Diagrama realizado en jvenn¹³².

8.3. Caracterización *in silico* de los lncRNA candidatos como biomarcadores moleculares en cáncer de próstata localmente avanzado de alto riesgo

Una vez identificadas las firmas de lncRNA con expresión diferencial mediante la validación experimental, se decidió caracterizar de manera *in silico* a cada uno de los lncRNA que las integran. Esto con el fin de conocer sus características genéticas principales, su localización celular, su expresión en tejido sano y tumoral de próstata, y su posible interacción con otras moléculas como DNA, RNA y proteínas de unión a RNA.

8.3.1. Características genéticas generales y determinación del potencial codificante de lncRNA

Como primer acercamiento, se realizó una búsqueda bibliográfica de las características principales de los lncRNA candidatos, la cual se compone de su localización genómica, su longitud en nucleótidos, la hebra de DNA a partir de la cual se sintetiza el lncRNA, a qué clase de lncRNA pertenecen, cuántas isoformas se han descrito, su función o asociación a CaP y, por último, su nivel de potencial codificante.

Tabla 9. Principales características de la firma de lncRNA con expresión diferencial en el grupo de casos de recurrencia bioquímica.

Símbolo	Localización genómica	Longitud (nt)	Cadena	Clase	Isoformas descritas	Función
H19	Chr.11	798	-	intergénico	12	Proliferación, involucrado en la vía de señalización del receptor de andrógenos.
LNCP2	Chr.14	1,351	+	antisentido	50	Inhibe la progresión de CaP, unión a EZH2, interacción con miRNA.
LNCRBQ1	Chr.7	4,665	+	antisentido	4	Regula a Hoxa9, promueve metástasis en CaP.
LNCRBQ2	Chr.X	2,356	-	intergénico	72	Asociado a progresión de cáncer colorrectal, esponja de miRNA.
LNCRBQ3	Chr.16	8,903	-	antisentido	24	Promueve proliferación, migración e invasión en CaP, decoy para miR-101.
LNCP1	Chr.7	565	-	antisentido	11	Promueve metástasis en nodos linfáticos en CaP.
LNCRBQ4	Chr.19	865	+	intergénico	49	Asociado a progresión en retinoblastoma.
LNCP3	Chr.10	7,282	-	antisentido	9	Asociado a proliferación, migración e invasión en CaP.

Información recopilada de las herramientas y bases de datos de *LNCipedia*¹²⁸, *Ensembl*¹³⁴ y *GeneCards*¹³⁵.

A partir de la información recolectada, se observó que todos los transcritos cumplen con las dos características principales que distinguen a un lncRNA, es decir, todos presentan una longitud mayor a 200 nucleótidos (Tabla 9) y tienen un potencial codificante negativo (Figura 26); al igual que los controles no codificantes PCA3 y HOTAIR, y al contrario de los controles codificantes KLK3 (mRNA del APE o PSA por

sus siglas en inglés), el RA (AR por sus siglas en inglés) y el gen de la actina (Figura 26). Estos datos corroboran que los paneles conformados por transcritos de RNA propuestos carecen de potencial para codificar proteínas, que se clasifican dentro de los lncRNA intergénicos y antisentido, y que por lo menos 6 de ellos (H19, LNCPC2, LNCRBQ1, LNCRBQ3, LNCPC1 y LNCPC3) se han visto asociados a CaP.

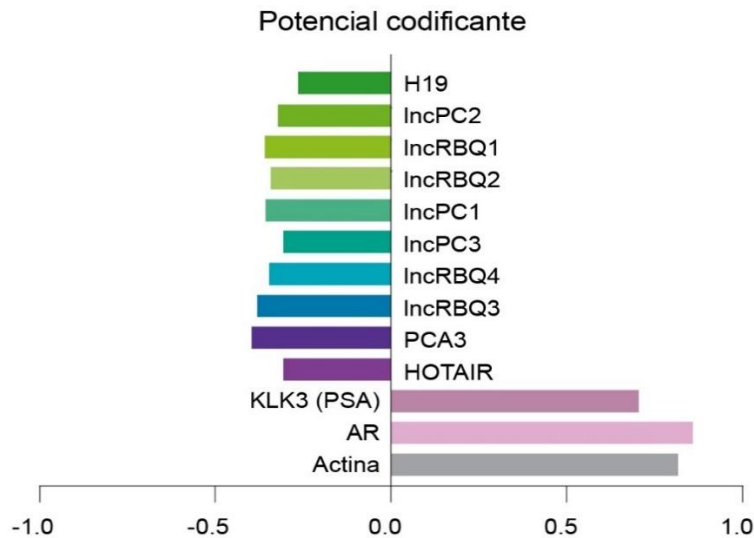


Figura 26. Potencial codificante de los lncRNA candidatos. Gráfica de barras que representa el valor del potencial codificante de cada lncRNA analizado. Se utilizaron a los genes no codificantes PCA3 y HOTAIR, y a los genes codificantes del APE, el receptor de andrógenos y el gen de la actina como controles positivos y negativos, respectivamente. Valores obtenidos a partir de la herramienta CNIT¹³⁶.

8.3.2. Determinación de la expresión diferencial de los lncRNA en tejido prostático sano y tumoral basada en datos de acceso público

Para poder determinar los niveles de expresión de los lncRNA candidatos en una cohorte independiente y con un mayor número de pacientes con CaP, se utilizó la base de datos de acceso público TANRIC¹³⁷, la cual cuenta con información de la expresión de lncRNA en 20 tipos diferentes de cáncer, entre ellos el CaP, incluyendo información de *The Cancer Genome Atlas* (TCGA).

En los análisis realizados con los datos públicos se comparó y validó lo que se había encontrado en nuestros ensayos experimentales, ya que el panel con expresión tumor-específica conformado por los lncRNA LNCPC1, H19, LNCPC2 y LNCPC3, efectivamente presenta expresión diferencial con significancia estadística en tejido tumoral de pacientes caucásicos con CaP (Figura 27B). Esto, tomando

como referencia al lncRNA PCA3 cuya sobreexpresión está reportada como específica de CaP (Figura 27A).

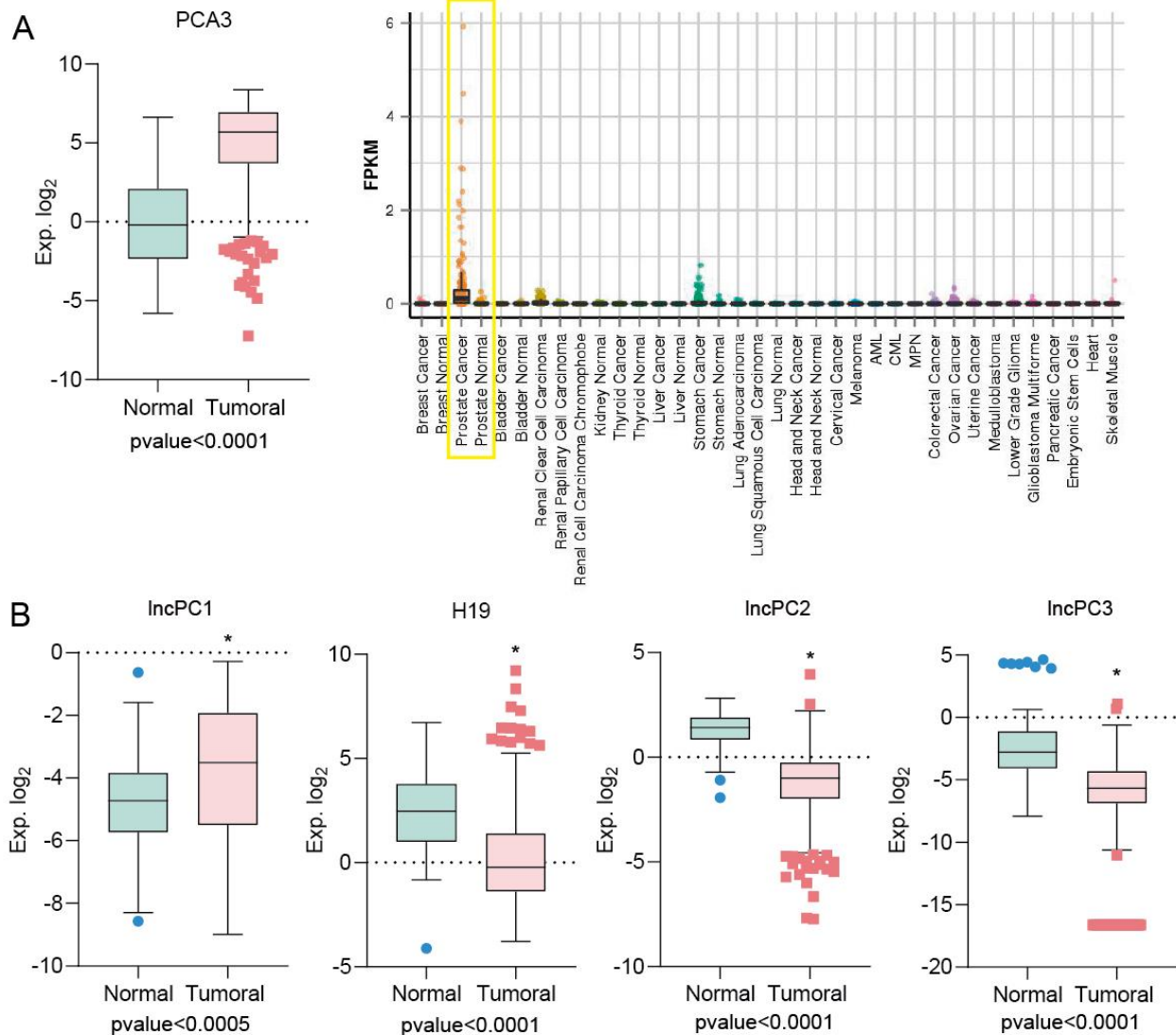


Figura 27. Validación *in silico* de la expresión diferencial del panel lncRNA con expresión tumor-específica en tejido tumoral y normal. A) Gráfico de caja y bigotes correspondientes al lncRNA control PCA3 específico de CaP. Se observa la sobreexpresión en datos colectados de TANRIC (izquierda) (N=52; T=374) y la gráfica generada por la herramienta *MiTranscriptome*⁷⁹ de la expresión de PCA3 en diferentes tipos de cáncer, el recuadro amarillo remarca la comparación de muestras de cáncer de próstata y de próstata sana (derecha) (n=304). B) Gráficos de caja y bigotes correspondientes al resto de los lncRNA que conforman al panel con expresión específica de tejido tumoral: LNCPC1, H19, LNCPC2 y LNCPC3. Se observa la expresión diferencial en datos colectados de TANRIC al comparar tejido tumoral y tejido sano. Los datos se analizaron mediante la prueba no paramétrica U de Mann-Whitney no pareada. Exp. Log₂: RPKM; H19: N=52 y T=373, LNCPC2: N=105 y T=837, LNCPC1: N=52 y T=373; $p\text{value} < 0.05$. Gráficas realizadas en *GraphPad Prism*¹³³.

Continuando con el análisis de datos de la base TANRIC, se evaluó al resto de los lncRNA candidatos, esta vez a los lncRNA que integran al panel con expresión específica de recurrencia bioquímica. Los lncRNA LNCRBQ3, LNCRBQ1 y LNCRBQ2 presentaron expresión diferencial significativa al compararse con el tejido sano. Los anteriormente mencionados están sobreexpresados en tejido tumoral al igual que en el ensayo experimental realizado por nosotros, sin embargo, el lncRNA LNCRBQ4 no presentó diferencia significativa entre tejido sano y tumoral. Esto puede deberse a la heterogeneidad en las muestras, ya que en los datos depositados en TANRIC de tejido tumoral podrían encontrarse más de un tipo de CaP (Figura 28).

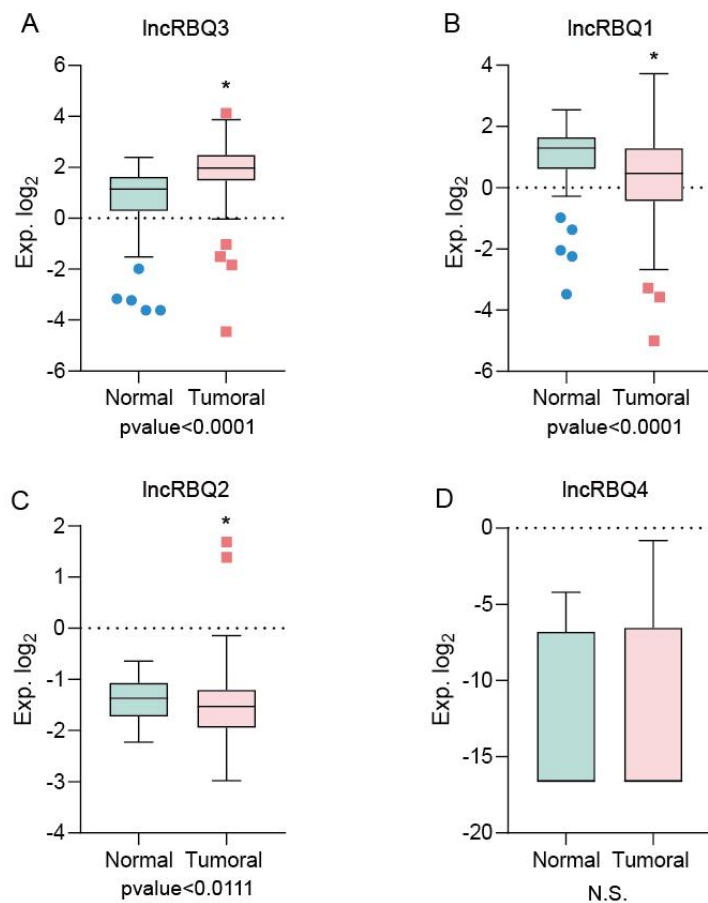


Figura 28. Análisis *in silico* de la expresión diferencial del panel específico de recurrencia bioquímica en tejido tumoral y normal de próstata. Gráficas de caja y bigotes correspondientes al panel de lncRNA con expresión específica de recurrencia bioquímica. Se observa la expresión diferencial en datos colectados de TANRIC al comparar tejido tumoral y tejido sano de los lncRNA **A)** LNCRBQ3, **B)** LNCRBQ1, **C)** LNCRBQ2 y **D)** LNCRBQ4. Los datos se analizaron mediante la prueba no paramétrica U de Mann-Whitney no pareada. Exp. Log₂: RPKM; N=52 y T=374; *pvalue*<0.05; N.S.: no significativo. Gráficas realizadas en *GraphPad Prism*¹³³.

8.3.3. Determinación de la localización y acumulación subcelular de los lncRNA analizados

Las evidencias experimentales han establecido que los lncRNA pueden encontrarse acumulados tanto en el núcleo, ya sea en la región de cromatina o en el nucleoplasma, como en el citoplasma. Esto varía según la función que estén desempeñando dentro de la célula, por ejemplo, si un lncRNA participa en la regulación de la expresión génica, éste se encontrará acumulado principalmente en el núcleo¹³⁸. Es por ello que se llevó a cabo una búsqueda en la base de datos lncATLAS¹³⁹ para conocer la localización celular de los lncRNA candidatos. En las gráficas obtenidas, se muestra la acumulación del transcrito en núcleo y citoplasma de acuerdo a ensayos de RNA-Sec posterior a ensayos de fraccionamiento celular. Esta herramienta se basa en los datos de líneas celulares de diferentes tipos de cáncer (Figura 29), lo cual es una aproximación para saber más respecto a los lncRNA caracterizados a pesar de que no son datos específicos de tejido de pacientes con CaP. En la Figura 29 se muestran las gráficas de 7 de los lncRNA encontrados en la fase experimental. En ellas se observa que todos los lncRNA presentan una mayor acumulación nuclear a excepción de H19, que se encuentra localizado principalmente en citoplasma. Cabe mencionar que la gráfica del lncRNA LNCRBQ4 no se muestra debido a que no hay registro de este lncRNA en la base de datos. En conclusión, estos resultados indican que los lncRNA analizados se localizan principalmente en núcleo a excepción de H19 presente en citoplasma.

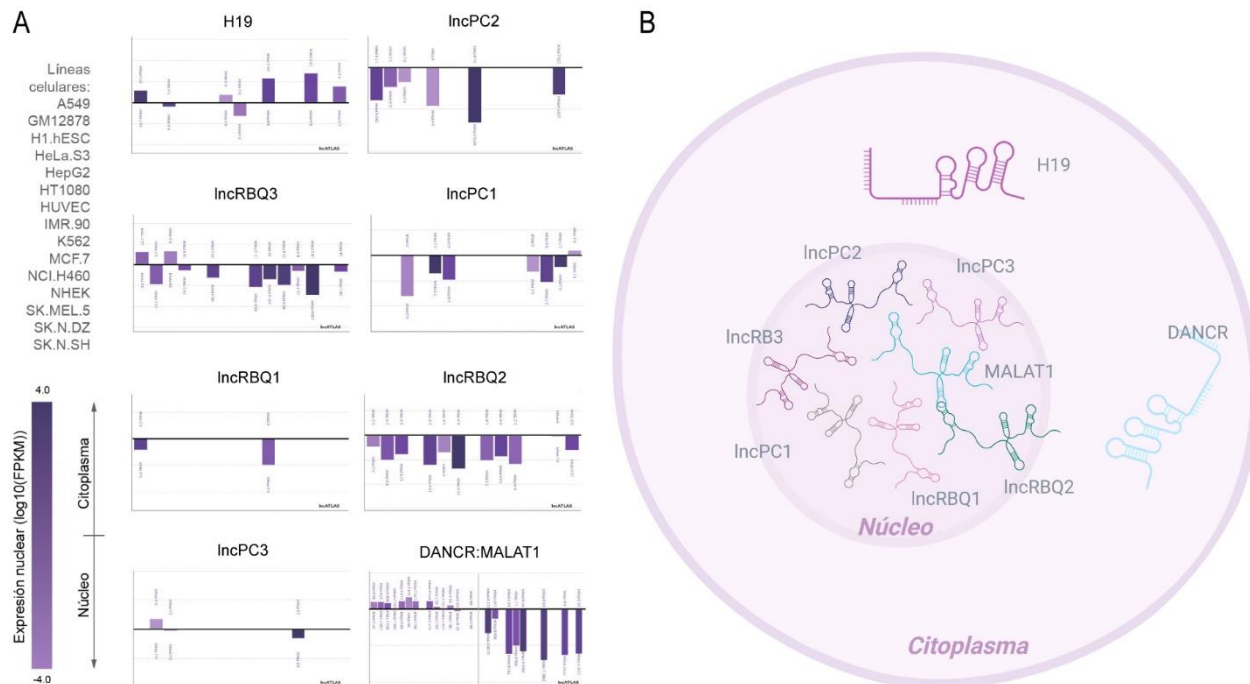


Figura 29. Localización celular de los lncRNA candidatos. **A)** Gráficos de barras obtenidos de la plataforma en línea lncATLAS correspondientes a la acumulación nuclear o citoplásmica de H19, LNCPC2, LNCRBQ3, LNCPC1, LNCRBQ1, LNCRBQ2 y LNCPC3. El valor de la concentración relativa núcleo/citoplasma graficado en el eje de las ordenadas indica la localización, si el valor es negativo significa que se encuentra en núcleo, mientras que, si el valor es positivo, indica que se encuentra en citoplasma. La intensidad del color indica el nivel de expresión en FPKM. En la parte superior izquierda se mencionan las líneas celulares del eje de las abscisas. Como controles se utilizaron al lncRNA nuclear MALAT1 y al lncRNA citoplásmico DANCR. **B)** Imagen representativa de la localización celular (nuclear/citoplásmica) de los lncRNA caracterizados en el inciso A.

8.3.4. Determinación de la interacción entre los lncRNA caracterizados y otras biomoléculas (DNA, proteínas, microRNA, etc.)

Con el fin de identificar biomoléculas relevantes en CaP que interactuaran con los lncRNA descubiertos en la fase experimental, se evaluaron las predicciones y evidencias que se tienen sobre las interacciones de los lncRNA candidatos con diferentes biomoléculas, ya sea, DNA, RNA, proteínas e incluso fármacos. Las redes de interacciones generadas a partir de las herramientas RNAinter¹⁴⁰ y RAIN¹⁴¹ (Figura 30) muestran las interacciones predichas por minería de textos, por modelos predictivos, por homología, por ensayos experimentales y de coexpresión; así como por lo reportado en la literatura en la actualidad y por ensayos de RNA-Sec. En la figura se muestran únicamente las gráficas correspondientes a los lncRNA H19 y

LNCRBQ1, el primero de ellos representando al panel de lncRNA con expresión tumor-específica, y el segundo representando al panel de lncRNA con expresión específica de recurrencia bioquímica (Figura 30).

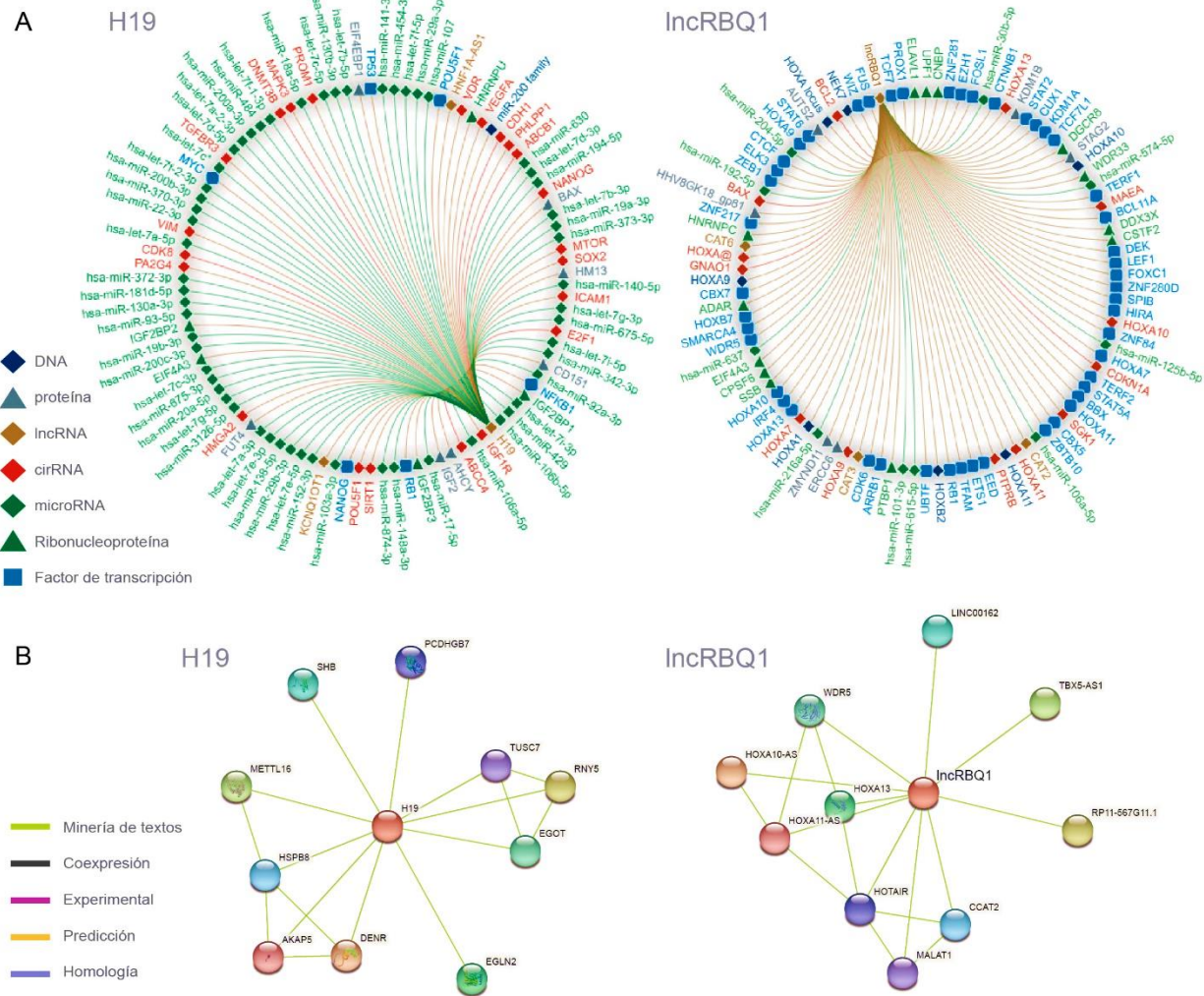


Figura 30. Red de interacción de los lncRNA que conforman los paneles de expresión diferencial en CaP y pacientes con recurrencia bioquímica. Imagen representativa de las interacciones de los lncRNA candidatos. **A)** Gráficos de circo representativos de las interacciones predichas y reportadas para los lncRNA H19 y LNCRBQ1 con moléculas de miRNA (rombo verde), circRNA (rojo), lncRNA (dorado), DNA (azul marino), proteínas (gris), ribonucleoproteínas (triángulo verde) y factores de transcripción (azul claro). **B)** Redes de interacción de los lncRNA H19 y LNCRBQ1. Cada línea corresponde a la fuente de la interacción graficada. Verde: minería de textos; negro: coexpresión; morado: experimental; naranja: predicción; azul: homología.

Después de haber evaluado cada una de las redes de interacción de los lncRNA, se identificaron las biomoléculas más importantes que estuvieran asociadas a CaP. En la tabla 10 se resumen estos hallazgos y se resaltan algunas de las interacciones predichas más importantes como la interacción con RA y las cinasas AKT1, mTOR y ERK1, así como con los genes supresores de tumores PTEN y P53, entre otras. Se ha reportado que todas estas moléculas están involucradas en vías de señalización representadas en cáncer y, principalmente, en CaP (Tabla 10).

Tabla 10. Principales interacciones predichas para los lncRNA caracterizados y su asociación con CaP.

Símbolo	Vías de señalización asociadas a cáncer de próstata
AR	Vía de señalización del receptor de andrógenos .
AKT1	Vía de señalización PI3K-AKT-mTOR , facilita la tumorigénesis, progresión y resistencia a terapia.
EZH2	Modificación transcripcional, supresión transcripcional.
E2F1	Vía de señalización IKKa-E2F1-BMI1, asociado a progresión de CRPC después de ADT.
MAPK3 (ERK1)	Proliferación celular, diferenciación celular, ciclo celular. Vía de FGFR (desarrollo de la próstata por GF) → activación de vías PI3K , MAPK, AR .
miR-200	Transición epitelio-mesenquimal, adhesión e invasión en CaP.
mTOR	Vía de señalización PI3K-AKT-mTOR , facilita la tumorigénesis, progresión y resistencia a terapia.
Myc	Transcripción, sobreexpresado por la fusión TMPRSS2:ERG , activa la expresión de EZH2.
PTEN	Pérdida presente en ~50% de CaP, asociada a progresión, metástasis, resistencia a castración.
P53	Ciclo celular, apoptosis.
RB1	Ciclo celular, señalización de AR y progresión a CRPC.
STAT	Proliferación, apoptosis, diferenciación celular. Vía de señalización JAK-STAT.
SUZ12	Modificación transcripcional, supresión transcripcional.

CRPC: Cáncer de próstata resistente a castración. ADT: Terapia de privación de andrógenos

9. DISCUSIÓN

El cáncer de próstata es un problema de salud pública a nivel mundial, ya que representa el 29.3% de casos nuevos y el 17.1% de muertes por cáncer en México²⁶. Recientemente se ha determinado que cerca del 30% de los pacientes con CaP sometidos a prostatectomía radical, presentan recurrencia bioquímica posterior a la cirugía, y que el 40% de ellos progresan a CaP metastásico⁶⁴. Además, son pocos los biomarcadores de pronóstico aprobados por la FDA que predigan la recurrencia bioquímica en pacientes con CaP posterior a la cirugía. Es por ello que este proyecto se enfocó en la búsqueda de nuevos biomarcadores de pronóstico de recurrencia bioquímica en pacientes mexicanos con CaP localmente avanzado de alto riesgo mediante una fase de descubrimiento basada en secuenciación masiva en paralelo de RNA y una fase de validación experimental. En este caso, se analizó la proporción no codificante del transcriptoma, específicamente, el biotipo conformado por los lncRNA presentes en muestras de tejido tumoral de los pacientes.

El proyecto se basó en un estudio de casos y controles en donde se compararon muestras de tejido tumoral y de tejido adyacente. Como primer resultado, se determinó que la expresión diferencial del transcriptoma no codificante, específicamente la de los lncRNA, permite la separación y diferenciación de muestras. Es decir, en el análisis de PCA las 3 muestras de tejido adyacente se encuentran más cercanas entre sí debido a que su perfil de expresión transcripcional es parecido y, por lo tanto, se encuentran alejadas de las 3 muestras de tejido tumoral. De igual manera, las 3 muestras de tejido tumoral se encuentran más cercanas entre sí por la misma razón. Por el contrario, el análisis de componentes principales basado en los niveles de expresión del transcriptoma completo no muestra la misma agrupación, además de que los porcentajes de varianza de PC1 y PC2 son considerablemente menores a los del análisis de lncRNA, indicando que la variación entre tejido tumoral y tejido adyacente es baja. A partir del análisis de expresión diferencial de RNA-Sec, se determinó que existe

un panel de lncRNA con expresión diferencial en CaP que no ha sido reportado anteriormente, el cual se conforma por 8 lncRNA, de los cuales 3 están subexpresados y 5 sobreexpresados. Con el fin de corroborar los resultados obtenidos, se repitió este análisis bioinformático utilizando el mismo flujo de herramientas: STAR junto con DESeq2 y un tercer análisis con la combinación de *Salmon* y DESeq2. Particularmente, el lncRNA H19 se encontró significativamente sobreexpresado en los resultados de los 2 *pipelines* independientes, además de haber sido el único transcrito presente en ambos análisis junto con SOX21-AS1, sin embargo, este último no fue tomado en cuenta para los siguientes ensayos ya que no se observó representado en la fase de validación. Debido a que el tamaño de muestra (n muestral) de este estudio fue muy pequeña, los resultados obtenidos nos muestran sólo un panorama de la población que estamos estudiando, por lo que una perspectiva del proyecto es incrementar el tamaño de la n muestral para corroborar que el perfil de expresión diferencial de los lncRNA en muestras de CaP coincide con el que encontramos en la etapa de descubrimiento.

Continuando con la segunda etapa, enfocada en la validación experimental de lo observado en el análisis de RNA-Sec, se llevaron a cabo ensayos de PCR cuantitativa (qPCR) acompañados de una reacción de transcriptasa reversa (RT-PCR) previa. Todo ello mediante el kit *RT² lncRNA qPCR Assay (Human Cancer PathwayFinder)* de QIAGEN que consiste en un arreglo de expresión de los 84 principales lncRNA representados en cáncer, entre ellos HOTAIR, XIST, PCA3 y H19. La etapa de validación experimental se dividió a su vez en tres subetapas, la primera de ellas se enfocó directamente en la comprobación de los resultados preliminares de RNA-Sec, ya que se analizó una cohorte independiente de muestras de tejido tumoral y adyacente de pacientes, pero esta vez con un número mayor de pacientes, en la que se colectaron 16 muestras control (tejido adyacente) y 16 muestras problema (tejido tumoral). A partir de este ensayo se obtuvo un perfil de expresión diferencial que distingue a las muestras tumorales de las de tejido adyacente, conformado por 15 genes sobreexpresados y 22 genes subexpresados,

entre ellos PCA3 y LNCPC1 con sobreexpresión, y H19, LNCPC2 y LNCPC3 con subexpresión. Cabe mencionar que en el mapa de agrupación jerárquica generado se puede observar la heterogeneidad de las muestras analizadas, característica particular del cáncer de próstata¹⁴². El conjunto de lncRNA identificado en esta segunda fase de validación no coincide con el que se encontró inicialmente, sin embargo, el lncRNA H19 nuevamente fue el único gen representado en ambos análisis, con la particularidad de que en el análisis de RNA-Seq inicial está sobreexpresado y en el ensayo por RT-qPCR está subexpresado. Nuevamente, para poder explicar la discrepancia observada, era necesario evaluar un número mayor de muestras, por lo que se recurrió a los recursos *in silico* y se consultaron los valores de expresión en RPKMs de la base de datos de TANRIC correspondientes a los 5 lncRNA con expresión diferencial. A partir de ellos se pudo corroborar que este panel era específico de CaP debido a que PCA3, un lncRNA con aplicación clínica en el diagnóstico de CaP, está sobreexpresado en las muestras de tejido tumoral de nuestro análisis y en los datos públicos de TANRIC. Los resultados de los lncRNA que conforman el panel (LNCPC1, H19, LNCPC2 y LNCPC3) también coincidieron con los resultados que se obtuvieron en la fase experimental, LNCPC1 se encuentra sobreexpresado en muestras de tejido tumoral de CaP mientras que H19, LNCPC2 y LNCPC3 se encuentran subexpresados. Cabe mencionar que la gráfica de H19 permite observar algunos valores atípicos u *outliers* que representan a los pacientes que sobreexpresan este gen. Esto podría explicar el resultado preliminar obtenido en un inicio, en el que H19 se sobreexpresaba en las 3 muestras de tejido tumoral iniciales. La segunda subetapa, como complemento de la primera, se destinó a la validación de los datos del gen H19 observados en un modelo de línea celular. Para ello se eligieron a PrEC y LNCaP como modelos de tejido epitelial sano y neoplásico, respectivamente. Como se observa en los resultados, H19 nuevamente presenta una subexpresión significativa en el modelo neoplásico, corroborando que la subexpresión de H19 es específica de CaP, al igual que la expresión de LNCPC1, LNCPC2 y LNCPC3.

La tercera y última subetapa consistió en comparar únicamente los niveles de expresión de los lncRNA de las muestras tumorales de pacientes sin recurrencia bioquímica (grupo control) y con recurrencia bioquímica (grupo problema), esto con el fin de encontrar un conjunto de lncRNA específico de recurrencia bioquímica en pacientes con CaP localmente avanzado de alto riesgo. Como resultado se observó un grupo de lncRNA con expresión diferencial conformado por 23 transcritos sobreexpresados y 31 subexpresados, entre ellos, LNCRBQ3, LNCRBQ1 y LNCRBQ2 sobreexpresados, y LNCRBQ4 subexpresado. Dentro de este panel también se encontró a LNCPC1 sobreexpresado y a H19, LNCPC2 y LNCPC3 subexpresados. Aun cuando la base de datos pública de TANRIC no se especializa en recurrencia bioquímica de CaP, se decidió evaluar los datos de cada uno de los lncRNA restantes en los datos de tejido tumoral y normal. Evaluando estos datos, pudo observarse que todo el panel de genes, a excepción de LNCRBQ4, presenta una expresión diferencial significativa en tejido tumoral. Esto, a pesar de no ser un resultado con información directa sobre recurrencia en CaP, sí muestra el impacto que podrían tener en esta enfermedad, ya que lo reportado en la actualidad nos indica que un gran porcentaje de la población de pacientes puede tener predisposición a recurrir. Posteriormente, se buscó distinguir cuáles de los lncRNA con expresión diferencial que se encontraron en la fase de validación eran específicos de muestras de tejido tumoral y cuáles eran específicos de muestras de pacientes con recurrencia bioquímica. Para ello se realizó un diagrama de Venn en donde se observaron las convergencias y divergencias de los genes sobreexpresados y subexpresados resultantes de los análisis (análisis 1: tejido adyacente contra tejido tumoral; análisis 2: pacientes sin recurrencia bioquímica contra pacientes con recurrencia bioquímica). Dentro del panel de lncRNA con expresión tumor-específica se encuentran LNCPC1, H19, LNCPC2 y LNCPC3, mientras que dentro del panel de lncRNA con expresión específica de recurrencia bioquímica se encuentran LNCRBQ1, LNCRBQ2, LNCRBQ3 y LNCRBQ4. Ninguno de estos paneles ha sido reportado anteriormente como biomarcador pronóstico en estudios clínicos y moleculares de CaP.

Por último, se continuó la consulta de datos públicos para realizar la caracterización *in silico* de los lncRNA. A partir de ésta, se corroboró que el panel se conformara únicamente por RNA que no tuvieran potencial codificante. Posteriormente, se determinaron sus características principales, dentro de las funciones que han sido reportadas, se encontró que al menos H19, LNCPC2, LNCRBQ1, LNCRBQ3, LNCPC1 y LNCPC3 han sido asociados a CaP, lo cual indica que podrían estar involucrados en la recurrencia bioquímica de pacientes. Además, se determinó que la mayoría de los lncRNA del panel se acumula en núcleo y que, por el contrario, H19 se acumula principalmente en citoplasma. Por último, se analizaron las redes de interacción que se generaron en herramientas en línea y se concluyó lo siguiente: las interacciones predichas y reportadas experimentalmente sugieren que cada uno de los lncRNA caracterizados posiblemente interactúan con moléculas involucradas en cáncer, como lo son proteínas, RNA, regiones de DNA, factores de transcripción e incluso agentes utilizados para el tratamiento contra el cáncer. Por ejemplo, LNCRBQ4 y LNCPC3 interactúan con RNA circulares (circRNA) del receptor de andrógenos, los cuales se expresan diferencialmente y han sido reportados como factores relevantes en CaP metastásico resistente a castración¹⁴³, mientras que LNCRBQ3 interactúa directamente con la proteína del RA, también considerado como un factor transcripcional de suma importancia por estar involucrado en CaP. En el caso de LNCPC2, se reporta una interacción con la proteína mTOR y el factor transcripcional *enhancer of zeste homolog 2* (EZH2), ambos involucrados en cascadas de señalización que favorecen la proliferación y supervivencia celular cuando su regulación se ve alterada; EZH2 se caracteriza por ser la subunidad catalítica del complejo de regulación transcripcional Polycomb 2, cuyo papel en cáncer se ha visto asociado a la desregulación del silenciamiento transcripcional de genes supresores de tumores que conlleva a la tumorigénesis¹⁴⁴. Aunado a esto, LNCPC1, LNCRBQ1 y LNCRBQ4 interactúan con *enhancer of zeste homolog 1* (EZH1), mientras que LNCPC3 con SUZ12, las cuales también son subunidades que conforman al complejo *Polycomb* y se han visto involucradas en el mismo fenómeno descrito anteriormente¹⁴⁴. Asimismo, LNCRBQ2 se relaciona con más de un factor

transcripcional como FUS, componente del complejo hnRNP involucrado en el procesamiento de miRNA, en la exportación de mRNA maduros al citoplasma y en la regulación de la expresión génica¹⁴⁵. De igual manera se observan una gran cantidad de asociaciones con miRNA, como es el caso de miR-215, el cual es un RNA blanco de LNCRBQ2, quien funciona como esponja inhibiendo su actividad y promoviendo la progresión en cáncer colorrectal¹⁴⁶. Por último, H19 presentó una gran cantidad de interacciones con moléculas involucradas en cáncer, por ejemplo, con el supresor de tumores TP53; el circRNA de MAPK3 involucrado en proliferación y diferenciación celular; el factor transcripcional del proto-oncogén MYC; el circ-RNA de mTOR, CDK8 y SOX2, así como con los factores transcripcionales NFKB1 y RB1, cuyo papel en cáncer de próstata ya ha sido reportado debido a que promueve la supervivencia celular, la invasión tumoral, metástasis y quimio-resistencia¹⁴⁷. En conjunto, estos resultados indican que los lncRNA caracterizados podrían tener una función relevante en CaP y en la recurrencia de la enfermedad ya que interactúan con factores que no sólo están representados en vías de señalización importantes en cáncer, sino en vías de señalización involucradas en CaP. Si bien se han descrito ya en otros tipos de cáncer, las interacciones encontradas en este análisis dan pauta a entender la probable influencia que tienen en la recurrencia bioquímica del CaP.

A partir de los análisis realizados y los resultados obtenidos, en este proyecto se propone al conjunto de lncRNA conformado por LNCPC1, H19, LNCPC2 y LNCPC3 como un panel con expresión diferencial específica de muestras de tejido tumoral de pacientes con CaP localmente avanzado, y al conjunto de lncRNA conformado por LNCRBQ1, LNCRBQ2, LNCRBQ3 y LNCRBQ4 como un panel con expresión diferencial específico de muestras de pacientes con recurrencia bioquímica en cáncer de próstata, haciendo énfasis en los lncRNA LNCPC1, LNCRBQ1 y LNCRBQ2, los cuales fueron los únicos con expresión diferencial estadísticamente significativa. Esta aseveración se da como resultado preliminar y antecedente de futuros estudios enfocados en la validación y reproducción del ensayo, así como en la búsqueda de vías en las que se pueden ver involucrados y

la determinación de la utilidad clínica como biomarcador, ya sea como genes individuales o como una firma génica. Dentro de las perspectivas del proyecto se tiene evaluar la sensibilidad y especificidad del panel, así como la comparación de éste con otros biomarcadores basados en mRNA y proteínas con el fin de dirigirse hacia la aplicación en la medicina de precisión.

Finalmente, este es un trabajo que se basa en la identificación de lncRNA mediante la herramienta de RNA-Seq y los arreglos de expresión transcriptómica, enfatizando la utilidad e importancia que tiene la RNA-Seq en el descubrimiento de genes cuando no se tiene un antecedente directo de lo que se está buscando. De igual manera, propone dos paneles de lncRNA que no se han identificado con anterioridad en la literatura biomédica y que pueden ser un preámbulo para su investigación a profundidad en un futuro enfocado en la aplicación del manejo personalizado del paciente oncológico con CaP.

10. CONCLUSIONES

A partir de los resultados obtenidos en este proyecto se concluye lo siguiente:

- Existe un panel de lncRNA con expresión diferencial que podría distinguir a las muestras de tejido adyacente de las del tejido tumoral en cáncer de próstata localmente avanzado que no ha sido reportado. El lncRNA LNCPC1 se sobreexpresa, mientras que los lncRNA H19, LNCPC2 y LNCPC3 se subexpresan.
- Existe un panel de lncRNA con expresión diferencial que podría distinguir a las muestras de pacientes con recurrencia bioquímica de las muestras de pacientes sin recurrencia bioquímica que no ha sido reportado. Los lncRNA LNCRBQ1, LNCRBQ2 y LNCRBQ3 se sobreexpresan, mientras que el lncRNA LNCRBQ4 se subexpresa.
- Los lncRNA LNCRBQ1 y LNCRBQ2 presentan una sobreexpresión estadísticamente significativa en muestras de pacientes con recurrencia bioquímica.
- Los lncRNA caracterizados podrían interactuar con moléculas asociadas a cáncer de próstata como lo son el receptor de andrógenos, PTEN, P53 y con componentes de la vía de señalización PI3K-AKT-mTOR.
- El panel de lncRNA conformado por LNCPC1, H19, LNCPC2 y LNCPC3 presenta expresión tumor-específica en CaP.
- El panel de lncRNA conformado por LNCRBQ1, LNCRBQ2, LNCRBQ3 y LNCRBQ4 es un potencial biomarcador de recurrencia bioquímica en CaP.

En conclusión, este estudio demuestra que existe un conjunto de lncRNA asociado a muestras de pacientes con recurrencia bioquímica en pacientes mexicanos con CaP localmente avanzado de alto riesgo, por lo que se propone como un posible biomarcador pronóstico de CaP para el manejo clínico de los pacientes.

11. PERSPECTIVAS

Las propuestas para continuar la dirección de este proyecto son las siguientes:

- Ampliar la cohorte a 100 pacientes para darle robustez a los datos de RNA-Sec, ya que en este estudio únicamente se trabajó con los datos de secuenciación de 3 pacientes con CaP localmente avanzado.
- Llevar a cabo un segundo análisis bioinformático enfocado en la comparación de pacientes con recurrencia bioquímica y pacientes sin recurrencia bioquímica.
- Llevar a cabo la caracterización molecular mediante ensayos experimentales de los lncRNA analizados en este estudio, evaluándolos como biomarcadores moleculares individuales y en conjunto como una firma génica de expresión en CaP localmente avanzado de alto riesgo.
- Determinar la utilidad clínica de los biomarcadores candidatos evaluando el porcentaje de sensibilidad y especificidad de cada uno de los lncRNA de manera individual y en conjunto.

12. REFERENCIAS

1. Schatten, H. Brief Overview of Prostate Cancer Statistics, Grading, Diagnosis and Treatment Strategies. *Adv. Exp. Med. Biol.* **1095**, 1–14 (2018).
2. Torre, L. A. *et al.* Global cancer statistics, 2012. *CA Cancer J Clin* **65**, 87–108 (2015).
3. Wang, G., Zhao, D., Spring, D. J. & DePinho, R. A. Genetics and biology of prostate cancer. *Genes Dev.* **32**, 1105–1140 (2018).
4. Dunn, M. W. & Kazer, M. W. Prostate Cancer Overview. *Seminars in Oncology Nursing* **27**, 241–250 (2011).
5. What Is Prostate Cancer? <https://www.cancer.org/cancer/prostate-cancer/about/what-is-prostate-cancer.html>.
6. Cox, A., Jefferies, M. & Persad, R. Prostate Structure and Function. in *Blandy's Urology* 509–521 (John Wiley & Sons, Ltd, 2019). doi:10.1002/9781118863343.ch25.
7. Amis, E. S. Anatomy and Physiology of the Prostate. in *Radiology of the Lower Urinary Tract* (ed. Lang, E. K.) 167–169 (Springer, 1994). doi:10.1007/978-3-642-84431-7_8.
8. McNeal, J. E. Normal and pathologic anatomy of prostate. *Urology* **17**, 11–16 (1981).
9. McNeal, J. E. Normal histology of the prostate. *Am. J. Surg. Pathol.* **12**, 619–633 (1988).
10. Haffner, J. *et al.* Peripheral zone prostate cancers: Location and intraprostatic patterns of spread at histopathology. *The Prostate* **69**, 276–282 (2009).
11. Shappell, S. B. *et al.* Prostate pathology of genetically engineered mice: definitions and classification. The consensus report from the Bar Harbor meeting of the Mouse Models of Human Cancer Consortium Prostate Pathology Committee. *Cancer Res.* **64**, 2270–2305 (2004).
12. Applewhite, J. C., Matlaga, B. R., McCullough, D. L. & Hall, M. C. Transrectal ultrasound and biopsy in the early diagnosis of prostate cancer. *Cancer Control* **8**, 141–150 (2001).

13. Hudson, D. L. Epithelial stem cells in human prostate growth and disease. *Prostate Cancer and Prostatic Diseases* **7**, 188–194 (2004).
14. Shen, M. M. & Abate-Shen, C. Molecular genetics of prostate cancer: new prospects for old challenges. *Genes Dev.* **24**, 1967–2000 (2010).
15. Peehl, D. M. Primary cell cultures as models of prostate cancer development. *Endocr. Relat. Cancer* **12**, 19–47 (2005).
16. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* **10**, 57–63 (2009).
17. Wang, Z. A. *et al.* Lineage analysis of basal epithelial cells reveals their unexpected plasticity and supports a cell-of-origin model for prostate cancer heterogeneity. *Nat. Cell Biol.* **15**, 274–283 (2013).
18. Wang, Z. A., Toivanen, R., Bergren, S. K., Chambon, P. & Shen, M. M. Luminal cells are favored as the cell of origin for prostate cancer. *Cell Rep* **8**, 1339–1346 (2014).
19. Henry, G. H. *et al.* A Cellular Anatomy of the Normal Adult Human Prostate and Prostatic Urethra. *Cell Rep* **25**, 3530-3542.e5 (2018).
20. Lilja, H., Oldbring, J., Rannevik, G. & Laurell, C. B. Seminal vesicle-secreted proteins and their reactions during gelation and liquefaction of human semen. *J. Clin. Invest.* **80**, 281–285 (1987).
21. Udager, A. M. & Tomlins, S. A. Molecular Biomarkers in the Clinical Management of Prostate Cancer. *Cold Spring Harb Perspect Med* **8**, (2018).
22. Brawer, M. K., Rennels, M. A., Nagle, R. B., Schifman, R. & Gaines, J. A. Serum Prostate-Specific Antigen and Prostate Pathology in Men Having Simple Prostatectomy. *Am J Clin Pathol* **92**, 760–764 (1989).
23. Bostwick, D. G. Prostate-specific antigen. Current role in diagnostic pathology of prostate cancer. *Am. J. Clin. Pathol.* **102**, S31-37 (1994).

24. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians* **71**, 209–249 (2021).
25. Ferlay, J. *et al.* Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *Int. J. Cancer* **144**, 1941–1953 (2019).
26. Cancer today. <http://gco.iarc.fr/today/home>.
27. Leitzmann, M. F. & Rohrmann, S. Risk factors for the onset of prostatic cancer: age, location, and behavioral correlates. *Clin Epidemiol* **4**, 1–11 (2012).
28. Rawla, P. Epidemiology of Prostate Cancer. *World J Oncol* **10**, 63–89 (2019).
29. Cancer of the Prostate - Cancer Stat Facts. *SEER*
<https://seer.cancer.gov/statfacts/html/prost.html>.
30. Kheirandish, P. & Chinegwundoh, F. Ethnic differences in prostate cancer. *British Journal of Cancer* **105**, 481–485 (2011).
31. Okobia, M. N., Zmuda, J. M., Ferrell, R. E., Patrick, A. L. & Bunker, C. H. Chromosome 8q24 variants are associated with prostate cancer risk in a high risk population of African ancestry. *Prostate* **71**, 1054–1063 (2011).
32. Gallagher, R. P. & Fleshner, N. Prostate cancer: 3. Individual risk factors. *CMAJ* **159**, 807–813 (1998).
33. Gann, P. H. Risk Factors for Prostate Cancer. *Rev Urol* **4**, S3–S10 (2002).
34. Zhou, A. *et al.* Interferon action and apoptosis are defective in mice devoid of 2',5'-oligoadenylate-dependent RNase L. *EMBO J.* **16**, 6355–6363 (1997).
35. Noda, D. *et al.* ELAC2, a putative prostate cancer susceptibility gene product, potentiates TGF-beta/Smad-induced growth arrest of prostate cells. *Oncogene* **25**, 5591–5600 (2006).
36. Gallagher, D. J. *et al.* Germline BRCA mutations denote a clinicopathologic subset of prostate cancer. *Clin. Cancer Res.* **16**, 2115–2121 (2010).

37. Descotes, J.-L. Diagnosis of prostate cancer. *Asian J Urol* **6**, 129–136 (2019).
38. Saini, S. PSA and beyond: alternative prostate cancer biomarkers. *Cell Oncol (Dordr)* **39**, 97–106 (2016).
39. Loeb, S. & Catalona, W. J. What is the role of digital rectal examination in men undergoing serial screening of serum PSA levels? *Nature Clinical Practice Urology* **6**, 68–69 (2009).
40. Draisma, G. *et al.* Lead time and overdiagnosis in prostate-specific antigen screening: importance of methods and context. *J. Natl. Cancer Inst.* **101**, 374–383 (2009).
41. Prostate-Specific Antigen (PSA) Test - National Cancer Institute. <https://www.cancer.gov/types/prostate/psa-fact-sheet> (2017).
42. Wolf, A. M. D. *et al.* American Cancer Society guideline for the early detection of prostate cancer: update 2010. *CA Cancer J Clin* **60**, 70–98 (2010).
43. Kearns, J. T. & Lin, D. W. Improving the Specificity of PSA Screening with Serum and Urine Markers. *Curr Urol Rep* **19**, 80 (2018).
44. Prostate Cancer - Stages and Grades. *Cancer.Net* <https://www.cancer.net/cancer-types/prostate-cancer/stages-and-grades> (2012).
45. AJCC - What is Cancer Staging? <https://cancerstaging.org/references-tools/Pages/What-is-Cancer-Staging.aspx>.
46. Paner, G. P. *et al.* Updates in the Eighth Edition of the Tumor-Node-Metastasis Staging Classification for Urologic Cancers. *Eur Urol* **73**, 560–569 (2018).
47. Epstein, J. I. *et al.* The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma: Definition of Grading Patterns and Proposal for a New Grading System. *Am. J. Surg. Pathol.* **40**, 244–252 (2016).

48. Harnden, P., Shelley, M. D., Coles, B., Staffurth, J. & Mason, M. D. Should the Gleason grading system for prostate cancer be modified to account for high-grade tertiary components? A systematic review and meta-analysis. *The Lancet Oncology* **8**, 411–419 (2007).
49. Martin, N. E., Mucci, L. A., Loda, M. & DePinho, R. A. Prognostic Determinants in Prostate Cancer. *Cancer J* **17**, 429–437 (2011).
50. Shukla-Dave, A. *et al.* Prediction of prostate cancer recurrence using magnetic resonance imaging and molecular profiles. *Clin. Cancer Res.* **15**, 3842–3849 (2009).
51. Demichelis, F. *et al.* TMPRSS2:ERG gene fusion associated with lethal prostate cancer in a watchful waiting cohort. *Oncogene* **26**, 4596–4599 (2007).
52. Berg, K. D. The prognostic and predictive value of TMPRSS2-ERG gene fusion and ERG protein expression in prostate cancer biopsies. *Dan Med J* **63**, B5319 (2016).
53. Halvorsen, O. J., Haukaas, S. A. & Akslen, L. A. Combined loss of PTEN and p27 expression is associated with tumor cell proliferation by Ki-67 and increased risk of recurrent disease in localized prostate cancer. *Clin. Cancer Res.* **9**, 1474–1479 (2003).
54. Klein, E. A. *et al.* A 17-gene assay to predict prostate cancer aggressiveness in the context of Gleason grade heterogeneity, tumor multifocality, and biopsy undersampling. *Eur Urol* **66**, 550–560 (2014).
55. Freedland, S. J. *et al.* Prognostic utility of cell cycle progression score in men with prostate cancer after primary external beam radiation therapy. *Int J Radiat Oncol Biol Phys* **86**, 848–853 (2013).
56. Ebell, M. H. Prolaris Test for Prostate Cancer Risk Assessment. *Am Fam Physician* **100**, 311–312 (2019).
57. Erho, N. *et al.* Discovery and validation of a prostate cancer genomic classifier that predicts early metastasis following radical prostatectomy. *PLoS One* **8**, e66855 (2013).

58. Karnes, R. J. *et al.* Validation of a genomic classifier that predicts metastasis following radical prostatectomy in an at risk patient population. *J. Urol.* **190**, 2047–2053 (2013).
59. Kretschmer, A. & Tilki, D. Biomarkers in prostate cancer - Current clinical utility and future perspectives. *Crit Rev Oncol Hematol* **120**, 180–193 (2017).
60. Kornberg, Z., Cooperberg, M. R., Spratt, D. E. & Feng, F. Y. Genomic biomarkers in prostate cancer. *Transl Androl Urol* **7**, 459–471 (2018).
61. Neoplasias prostáticas | Smith y Tanagho. Urología general, 18e | AccessMedicina | McGraw-Hill Medical.
<https://accessmedicina.mhmedical.com/content.aspx?bookid=1487§ionid=96874955#1119287628>.
62. Zaorsky, N. G., Raj, G. V., Trabulsi, E. J., Lin, J. & Den, R. B. The dilemma of a rising prostate-specific antigen level after local therapy: What are our options? *Seminars in oncology* **40**, 322–336 (2013).
63. Cookson, M. S. *et al.* Variation in the definition of biochemical recurrence in patients treated for localized prostate cancer: the American Urological Association Prostate Guidelines for Localized Prostate Cancer Update Panel report and recommendations for a standard in the reporting of surgical outcomes. *J. Urol.* **177**, 540–545 (2007).
64. Gu, Y., Lin, X., Kapoor, A., Mei, W. & Tang, D. Attempt to predict early recurrence of prostate cancer following prostatectomy through machine learning. *AME Medical Journal* **3**, (2018).
65. MA, D. *et al.* Association of molecular biomarkers expression with biochemical recurrence in prostate cancer through tissue microarray immunostaining. *Oncol Lett* **10**, 2185–2191 (2015).

66. Sharma, S. & Watabe, K. Biomarkers and mechanisms associated with recurrent prostate cancer. *Front Biosci (Landmark Ed)* **19**, 339–351 (2014).
67. Agoulnik, I. U. *et al.* Androgens modulate expression of transcription intermediary factor 2, an androgen receptor coactivator whose expression level correlates with early biochemical recurrence in prostate cancer. *Cancer Res* **66**, 10594–10602 (2006).
68. Kahl, P. *et al.* Androgen receptor coactivators lysine-specific histone demethylase 1 and four and a half LIM domain protein 2 predict risk of prostate cancer recurrence. *Cancer Res* **66**, 11341–11347 (2006).
69. Danila, D. C., Fleisher, M. & Scher, H. I. Circulating tumor cells as biomarkers in prostate cancer. *Clin Cancer Res* **17**, 3903–3912 (2011).
70. Lin, X. *et al.* Assessment of biochemical recurrence of prostate cancer (Review). *Int. J. Oncol.* **55**, 1194–1212 (2019).
71. Gittelman, M. C. *et al.* PCA3 molecular urine test as a predictor of repeat prostate biopsy outcome in men with previous negative biopsies: a prospective multicenter clinical study. *J. Urol.* **190**, 64–69 (2013).
72. Prensner, J. R. *et al.* The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. *Nat. Genet.* **45**, 1392–1398 (2013).
73. Hajjari, M. & Salavaty, A. HOTAIR: an oncogenic long non-coding RNA in different cancers. *Cancer Biol Med* **12**, 1–9 (2015).
74. Prensner, J. R. *et al.* Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. *Nat. Biotechnol.* **29**, 742–749 (2011).
75. Yoshimoto, R., Mayeda, A., Yoshida, M. & Nakagawa, S. MALAT1 long non-coding RNA in cancer. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1859**, 192–199 (2016).

76. Bussemakers, M. J. G. *et al.* DD3::A New Prostate-specific Gene, Highly Overexpressed in Prostate Cancer. *Cancer Res* **59**, 5975–5979 (1999).
77. Hangauer, M. J., Vaughn, I. W. & McManus, M. T. Pervasive Transcription of the Human Genome Produces Thousands of Previously Unidentified Long Intergenic Noncoding RNAs. *PLOS Genetics* **9**, e1003569 (2013).
78. Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* **489**, 101–108 (2012).
79. Iyer, M. K. *et al.* The landscape of long noncoding RNAs in the human transcriptome. *Nat. Genet.* **47**, 199–208 (2015).
80. Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Res* **22**, 1775–1789 (2012).
81. Schmitz, K.-M., Mayer, C., Postepska, A. & Grummt, I. Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. *Genes Dev.* **24**, 2264–2269 (2010).
82. Geisler, S. & Coller, J. RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nat. Rev. Mol. Cell Biol.* **14**, 699–712 (2013).
83. Mao, Y. S., Sunwoo, H., Zhang, B. & Spector, D. L. Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs. *Nat. Cell Biol.* **13**, 95–101 (2011).
84. Sun, M. & Kraus, W. L. From discovery to function: the expanding roles of long noncoding RNAs in physiology and disease. *Endocr. Rev.* **36**, 25–64 (2015).
85. Rinn, J. L. & Chang, H. Y. Genome Regulation by Long Noncoding RNAs. *Annual Review of Biochemistry* **81**, 145–166 (2012).
86. Ulitsky, I. & Bartel, D. P. lincRNAs: Genomics, Evolution, and Mechanisms. *Cell* **154**, 26–46 (2013).

87. Camacho, C. V., Choudhari, R. & Gadad, S. S. Long noncoding RNAs and cancer, an overview. *Steroids* **133**, 93–95 (2018).
88. Qiu, M.-T., Hu, J.-W., Yin, R. & Xu, L. Long noncoding RNA: an emerging paradigm of cancer research. *Tumour Biol.* **34**, 613–620 (2013).
89. Pennisi, E. Cell biology. Lengthy RNAs earn respect as cellular players. *Science* **344**, 1072 (2014).
90. Ponting, C. P., Oliver, P. L. & Reik, W. Evolution and functions of long noncoding RNAs. *Cell* **136**, 629–641 (2009).
91. Shi, T., Gao, G. & Cao, Y. Long Noncoding RNAs as Novel Biomarkers Have a Promising Future in Cancer Diagnostics. *Dis Markers* **2016**, 9085195 (2016).
92. Jarroux, J., Morillon, A. & Pinskaya, M. History, Discovery, and Classification of lncRNAs. *Adv Exp Med Biol* **1008**, 1–46 (2017).
93. Quinn, J. J. & Chang, H. Y. Unique features of long non-coding RNA biogenesis and function. *Nature Reviews Genetics* **17**, 47–62 (2016).
94. Ransohoff, J. D., Wei, Y. & Khavari, P. A. The functions and unique features of long intergenic non-coding RNA. *Nat Rev Mol Cell Biol* **19**, 143–157 (2018).
95. Ma, L., Bajic, V. B. & Zhang, Z. On the classification of long non-coding RNAs. *RNA Biol* **10**, 924–933 (2013).
96. Chen, L.-L. Linking Long Noncoding RNA Localization and Function. *Trends Biochem Sci* **41**, 761–772 (2016).
97. Salviano-Silva, A., Lobo-Alves, S. C., de Almeida, R. C., Malheiros, D. & Petzl-Erler, M. L. Besides Pathology: Long Non-Coding RNA in Cell and Tissue Homeostasis. *Noncoding RNA* **4**, 3 (2018).
98. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **144**, 646–674 (2011).

99. Schmitt, A. M. & Chang, H. Y. Long Noncoding RNAs in Cancer Pathways. *Cancer Cell* **29**, 452–463 (2016).
100. Luo, G. *et al.* Long Non-Coding RNA MEG3 Inhibits Cell Proliferation and Induces Apoptosis in Prostate Cancer. *Cell Physiol Biochem* **37**, 2209–2220 (2015).
101. Smolle, M. A., Bauernhofer, T., Pummer, K., Calin, G. A. & Pichler, M. Current Insights into Long Non-Coding RNAs (LncRNAs) in Prostate Cancer. *Int J Mol Sci* **18**, 473 (2017).
102. Lemos, A. E. G. *et al.* PCA3 long noncoding RNA modulates the expression of key cancer-related genes in LNCaP prostate cancer cells. *Tumour Biol.* **37**, 11339–11348 (2016).
103. Salameh, A. *et al.* PRUNE2 is a human prostate cancer suppressor regulated by the intronic long noncoding RNA PCA3. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 8403–8408 (2015).
104. Mitobe, Y., Takayama, K.-I., Horie-Inoue, K. & Inoue, S. Prostate cancer-associated lncRNAs. *Cancer Lett* **418**, 159–166 (2018).
105. Zhang, G., He, X., Ren, C., Lin, J. & Wang, Q. Long noncoding RNA PCA3 regulates prostate cancer through sponging miR-218-5p and modulating high mobility group box 1. *Journal of Cellular Physiology* **234**, 13097–13109 (2019).
106. Haese, A. *et al.* Clinical utility of the PCA3 urine assay in European men scheduled for repeat biopsy. *Eur. Urol.* **54**, 1081–1088 (2008).
107. Hessels, D. *et al.* DD3(PCA3)-based molecular urine analysis for the diagnosis of prostate cancer. *Eur. Urol.* **44**, 8–15; discussion 15-16 (2003).
108. Xue, W.-J., Ying, X.-L., Jiang, J.-H. & Xu, Y.-H. Prostate cancer antigen 3 as a biomarker in the urine for prostate cancer diagnosis: a meta-analysis. *J Cancer Res Ther* **10 Suppl**, C218-221 (2014).

109. Durand, X., Moutereau, S., Xylinas, E. & de la Taille, A. Progenesa™ PCA3 test for prostate cancer. *Expert Rev. Mol. Diagn.* **11**, 137–144 (2011).
110. Sartori, D. A. & Chan, D. W. Biomarkers in prostate cancer: what's new? *Curr Opin Oncol* **26**, 259–264 (2014).
111. Mehra, R. *et al.* A novel RNA in situ hybridization assay for the long noncoding RNA SChLAP1 predicts poor clinical outcome after radical prostatectomy in clinically localized prostate cancer. *Neoplasia* **16**, 1121–1127 (2014).
112. Cimadamore, A. *et al.* Long Non-coding RNAs in Prostate Cancer with Emphasis on Second Chromosome Locus Associated with Prostate-1 Expression. *Front Oncol* **7**, 305 (2017).
113. Prensner, J. R. *et al.* Nomination and validation of the long noncoding RNA SChLAP1 as a risk factor for metastatic prostate cancer progression: a multi-institutional high-throughput analysis. *Lancet Oncol* **15**, 1469–1480 (2014).
114. Xiong, T., Li, J., Chen, F. & Zhang, F. PCAT-1: A Novel Oncogenic Long Non-Coding RNA in Human Cancers. *Int J Biol Sci* **15**, 847–856 (2019).
115. Prensner, J. R. *et al.* PCAT-1, a long noncoding RNA, regulates BRCA2 and controls homologous recombination in cancer. *Cancer Res.* **74**, 1651–1660 (2014).
116. Prensner, J. R. *et al.* The long non-coding RNA PCAT-1 promotes prostate cancer cell proliferation through cMyc. *Neoplasia* **16**, 900–908 (2014).
117. Xue, M., Chen, W. & Li, X. Urothelial cancer associated 1: a long noncoding RNA with a crucial role in cancer. *J. Cancer Res. Clin. Oncol.* **142**, 1407–1419 (2016).
118. Yu, Y., Gao, F., He, Q., Li, G. & Ding, G. lncRNA UCA1 Functions as a ceRNA to Promote Prostate Cancer Progression via Sponging miR143. *Mol Ther Nucleic Acids* **19**, 751–758 (2019).

119. Fotouhi Ghiam, A. *et al.* Long non-coding RNA urothelial carcinoma associated 1 (UCA1) mediates radiation response in prostate cancer. *Oncotarget* **8**, 4668–4689 (2017).
120. Stark, R., Grzelak, M. & Hadfield, J. RNA sequencing: the teenage years. *Nature Reviews Genetics* **20**, 631–656 (2019).
121. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).
122. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
123. Peña-Llopis, S. & Brugarolas, J. Simultaneous isolation of high-quality DNA, RNA, miRNA and proteins from tissues for genomic applications. *Nat Protoc* **8**, 2240–2255 (2013).
124. Galaxy. <https://usegalaxy.org/root>.
125. GENCODE - Human Release 34. <https://www.gencodegenes.org/human/>.
126. Ensembl genome browser 100. <https://www.ensembl.org/index.html>.
127. Home - Gene - NCBI. <https://www.ncbi.nlm.nih.gov/gene/>.
128. LNCipedia. <https://lncipedia.org/>.
129. Information, N. C. for B., Pike, U. S. N. L. of M. 8600 R., MD, B. & Usa, 20894. National Center for Biotechnology Information. <https://www.ncbi.nlm.nih.gov/>.
130. Oligo Analyzer. <https://www.idtdna.com/calc/analyzer>.
131. UCSC In-Silico PCR. https://genome.ucsc.edu/cgi-bin/hgPcr?hgsid=856009311_7ZfB39FI9728Z6cU27fzGfX8zX2j.
132. Bardou, P., Mariette, J., Escudié, F., Djemiel, C. & Klopp, C. jvenn: an interactive Venn diagram viewer. *BMC Bioinformatics* **15**, 293 (2014).
133. Home - GraphPad. <https://www.graphpad.com/>.

134. Homo_sapiens - Ensembl genome browser 103.
https://www.ensembl.org/Homo_sapiens/Info/Index.
135. GeneCards - Human Genes | Gene Database | Gene Search.
<https://www.genecards.org/>.
136. CNIT: a fast and accurate web tool for identifying protein-coding and long non-coding transcripts based on intrinsic sequence composition | Nucleic Acids Research | Oxford Academic. <https://academic.oup.com/nar/article/47/W1/W516/5506859>.
137. Li, J. *et al.* TANRIC: An Interactive Open Platform to Explore the Function of lncRNAs in Cancer. *Cancer Res* **75**, 3728–3737 (2015).
138. Carlevaro-Fita, J. & Johnson, R. Global Positioning System: Understanding Long Noncoding RNAs through Subcellular Localization. *Mol Cell* **73**, 869–883 (2019).
139. Mas-Ponte, D. *et al.* LncATLAS database for subcellular localization of long noncoding RNAs. *RNA* **23**, 1080–1087 (2017).
140. Lin, Y. *et al.* RNAInter in 2020: RNA interactome repository with increased coverage and annotation. *Nucleic Acids Res* **48**, D189–D197 (2020).
141. Junge, A. *et al.* RAIN: RNA-protein Association and Interaction Networks. *Database (Oxford)* **2017**, (2017).
142. Berglund, E. *et al.* Spatial maps of prostate cancer transcriptomes reveal an unexplored landscape of heterogeneity. *Nature Communications* **9**, 2419 (2018).
143. Cao, S. *et al.* Circular RNAs add diversity to androgen receptor isoform repertoire in castration-resistant prostate cancer. *Oncogene* **38**, 7060–7072 (2019).
144. Veneti, Z., Gkouskou, K. K. & Eliopoulos, A. G. Polycomb Repressor Complex 2 in Genomic Instability and Cancer. *Int J Mol Sci* **18**, 1657 (2017).
145. Geuens, T., Bouhy, D. & Timmerman, V. The hnRNP family: insights into their role in health and disease. *Hum Genet* **135**, 851–867 (2016).

146. Yang, Y. *et al.* LncRNA FTX sponges miR-215 and inhibits phosphorylation of vimentin for promoting colorectal cancer progression. *Gene Ther* **25**, 321–330 (2018).
147. Verzella, D. *et al.* Targeting the NF- κ B pathway in prostate cancer: a promising therapeutic approach? *Curr Drug Targets* **17**, 311–320 (2016).

Prostate Cancer: The promising role of new molecular biomarkers, from the patient management to artificial intelligence approaches.

Ana Paula Alarcón-Zendejas¹, Anna Scavuzzo², Miguel A. Jiménez-Ríos², Rosa M. Álvarez-Gómez³, Rogelio Montiel-Manríquez¹, Castro-Hernández Clementina¹, Miguel A. Jiménez-Dávila², Delia Pérez-Montiel⁴, Rodrigo González-Barrios¹, Francisco Jiménez-Trejo⁵, Cristian Arriaga-Canon^{1*} and Luis A. Herrera^{1,6*}.

Affiliations:

¹Unidad de Investigación Biomédica en Cáncer, Instituto Nacional de Cancerología-Instituto de Investigaciones Biomédicas, UNAM. Avenida San Fernando No. 22 Col. Sección XVI, Tlalpan. C. P. 14080, CDMX-México.

²Departamento de Urología, Instituto Nacional de Cancerología. Avenida San Fernando No. 22 Col. Sección XVI, Tlalpan. C.P. 14080, CDMX-México.

³Clínica de Cáncer Hereditario, Instituto Nacional de Cancerología. Avenida San Fernando No. 22 Col. Sección XVI, Tlalpan. C.P. 14080, CDMX-México.

⁴Departamento de Patología, Instituto Nacional de Cancerología. Avenida San Fernando No. 22 Col. Sección XVI, Tlalpan. C.P. 14080, CDMX-México.

⁵Instituto Nacional de Pediatría, Insurgentes Sur No. 3700-C. Coyoacán. C. P. 04530, CDMX-México.

⁶Instituto Nacional de Medicina Genómica. Periférico Sur 4809. Arenal Tepepan, Tlalpan. CP 14610. Mexico City, Mexico.

Contact Information

*Correspondence:

Luis A. Herrera, herreram@biomedicas.unam.mx

Tel. 52 5350-1900 ext. 1901.

Cristian Arriaga, carriagac@incan.edu.mx.

Tel. 52 3693-5200 Ext. 236

Keywords: Prostate cancer, biomarkers, transcriptomics, genomics, artificial intelligence.

Conflict of interests:

None

Abstract

Prostate cancer is one of the leading causes of cancer death in men worldwide. Risk stratification for recurrence or progression in prostate cancer is performed with the support of clinical-pathological data such as the clinical stage, which is determined by digital-rectal exploration, the sum of the Gleason score from prostate biopsies, and serum levels of prostate-specific antigen (PSA). There are several risk factors for prostate cancer, including age, prostate cancer family history, and ethnicity, among others. For several decades, methods aimed at the early detection of prostate cancer have included the determination of PSA serum levels. This and other recent advances, including the search for changes in circulating cell-free DNA, microRNAs, lncRNAs and extracellular vesicles, are expected to contribute to an earlier and accurate diagnosis and therefore an improvement in disease prognosis and better therapeutic strategies. In this review, we analyze several aspects that are relevant for prostate cancer, including its epidemiology, risk factors, importance of genetics, diagnosis, stage classification, prognostic factors, molecular pathogenesis, and bioinformatic approaches for its analysis such as Machine Learning. We also include current techniques that will allow an earlier diagnosis, such as Spatial Transcriptomics, Exome Sequencing, and Whole Genome Sequencing. An understanding of the biology of this neoplasm will contribute to its early detection and the selection of appropriate treatments for each patient in the context of precision medicine, thus improving the prognosis and quality of life of patients with prostate cancer.

Introduction

Prostate cancer (PCa) refers to the abnormal proliferation of cells in the prostate gland, which is located in the male reproductive tract¹. Although the proliferation of PCa cells initially occurs in the prostate gland, they can reach other tissues, mainly bones and lymph nodes². PCa is a biologically heterogeneous disease that can be painless, progressive, and lethal³. PCa includes a wide spectrum of clinical manifestations, and thus, its natural history can vary⁴. There are cases in which the patient goes through several phases. These stages can include an initial phase in which the disease has a localized-regional extension and the patient shows no symptoms, to a very late phase in which the patient is resistant to hormonal therapy, and invasion of adjacent tissues; alternately in some cases, the disease does not evolve and remains latent⁵.

The risk factors associated with PCa include age (men older than 60 years of age), positive history of PCa, and ethnicity, among others^{6,7}. Detection of PCa includes the measurement of prostate-specific antigen (PSA) serum levels and the digital rectal exam (DRE), contributing with the detection of PCa in early stages⁸. The localized disease, when it is confined to the prostate, is treated with radical prostatectomy (RP)⁸. In contrast, localized advanced PCa is treated with surgery (RP), adjuvants, radiotherapy such as external beam radiation or brachytherapy; and hormone therapy such as Luteinizing hormone-releasing hormone (LHRH) agonists and antagonists, abiraterone, and enzalutamide; while metastatic disease is usually treated with hormone therapy such as apalutamide, and chemotherapy⁹. If PCa is diagnosed on time, the treatment can be effective and with minimal morbidity¹⁰.

Hereditary prostate cancer.

It is recognized that hereditary cancer represents the etiology of 5 to 10% of all neoplasms¹¹. PCa has been considered to be the tumor associated with the greatest family history of cancer¹². Paradoxically, few high-susceptibility genes consistently related to the hereditary of PCa have been identified, presenting a pattern of dominant autosomal inheritance¹³. This phenomenon has been linked to a wide phenotypic variation and genetic heterogeneity, which has complicated the search for genes that are highly associated with high PCa predisposition¹⁴.

Currently, there is a particular interest in analyze Pathogenic Variants (PV) in high predisposition cancer genes associated with defects in homologous recombination repair and mismatch repair¹⁵. This is due to targeted therapeutic options, that include the use of PARP1 inhibitors; chemotherapies with platinum compounds, and pembrolizumab (with pooled evidence of microsatellite instability in tumor), particularly in patients with metastatic and castration-resistant disease^{16,17}. The use of multi-gene panels in germline diagnosis has identified PV in 7% to 12% of prostate cancer patients^{18,19}, highlighting *BRCA1*, *BRCA2*, *ATM*, *BRIP1*, *CHEK2*, *NBN*, *BARD1*, *RAD51C*, *MRE11A* and *PALB2* (homologous recombination repair); *MLH1*, *MSH2*, *MSH6* and *PMS2* (mismatch repair). Some of these high risk genes have clinical guidelines; option for risk reduction surgeries, and personalized treatment, which benefits not only the patient where PV is identified, but also the family²⁰. The main predictors for finding a PV in a multipanel study include high Gleason score; personal history of breast or pancreatic cancer, as well as a family history suggestive of Hereditary Breast and Ovarian Cancer (HBOC) and / or Lynch syndrome (LS)¹².

Regarding the genotype-phenotype correlations described, those attributable to HBOC stand out. The HBOC-*BRCA2* gene, a tumor-suppressing gene located on the long arm of chromosome 13, has been repeatedly associated with a high risk of PCa, particularly in a subgroup of increased risk of subclinical, high-grade disease,

increased mortality and earlier age at diagnosis²¹. Patients carrying heterozygous germline mutations in *BRCA2* have a 25% risk of suffering from PCa during their lifetime²². Close urological follow-up measures, with semestral detection of PSA and urological checks, are recommended for these male beginning at the age of 40 years²³. It is estimated that germline mutations in *BRCA2* could be the cause of up to 1% of PCa globally^{21,22}. Although the *BRCA2* mutation frequency varies according to the disease status²⁴.

Another highly predisposed gene outside the DNA repair pathways, *HOXB13*, is a member of the homeobox gene family whose protein product is essential in urogenital development and maintenance of the prostate in adult life. The *HOXB13* missense mutation -G84E- (*HOXB13* recurrent mutation) has been demonstrated to contribute has been shown to correlate with an early age of onset of PCa²⁵. Although what is recognized of hereditary prostate cancer does not yet imply a generalized molecular diagnosis to all PCa patients, it does entail the identification of a subgroup of patients with metastatic disease; early age of onset, and cancer family history, who will have a greater benefit for the therapeutic options and family prevention as a result of the molecular approach^{26,27}.

Diagnosis

The detection of PCa is regularly performed using the digital rectal exam, transrectal ultrasound, and serum determination of PSA. The need for prostate biopsy is based on PSA levels and/or changes in the digital rectal exam. PSA is a glycoprotein that is mainly synthesized in the prostate; therefore, its detection is paramount in PCa diagnosis⁸. Even though, clinical studies have been performed to determine the efficiency of PSA due to the false-positive results and the current discrepancies concerning the efficacy of this type of test²⁸. In order to cover the proportion of indeterminate findings, novel diagnostic biomarkers have been developed such as Prostate Health Index, 4K score, SelectMDx, ConfirmMDx and PCA3²⁹. In summary,

there is an intense debate regarding PSA as a diagnostic and screening tool in PCa, and therefore it is especially important to focus on other types of markers that can support diagnostic procedures and decision making for therapy, such as the diagnostic biomarker PCA3.

Stage classification.

Stage classification in PCa is performed using the TNM system (Tumor, Node, Metastasis), which considers three factors: extension of the primary tumor, infiltration of lymph nodes, and metastasis³⁰. For the specific case of PCa, the most used stage classification system was proposed by the American Joint Committee on Cancer (AJCC) and the Union for International Cancer Control (UICC) and is applied to adenocarcinomas³¹. Likewise, using the Gleason Score, it is possible to attribute a histological grade to PCa cases. Thus, 5 different histological patterns observed in prostate tumors were catalogued. Each one of them corresponds to a different level of differentiation, with 1 being the highest (more differentiated; less aggressive) and 5 the lowest (less differentiated; more aggressive). The International Society of Urologic Pathology has recommended a new prognosis grading system that has simplified the 9 possible Gleason Scores into 5 prognostic grade groups (Table 1)³². This new classification called as Epstein System can better define cancer aggressiveness and significantly reduce upgrading events³³.

Prognostic factors in prostate cancer.

Although several studies have been performed to analyze different molecular biomarkers, such as variant V7 of the androgen receptor³⁴ or inactivation of the *PTEN* or *c-MYC* gene³⁵, to date, none of them have been approved as a prognostic biomarker for use in clinical settings^{36,37}. Therefore, clinical-pathological factors such as age, Gleason Score, presence of extracapsular tumor extension, invasion to seminal vesicles, clinical stage, and PSA scores are still tools that allow the treatment to be personalized according to the patient's needs²⁹. Currently, there are

a few molecular prognostic biomarkers in clinical use such as OncotypeDX Genomic Prostate Score³⁸, Prolaris³⁹, ProMark⁴⁰, and Decipher⁴¹, based on cancer-associated gene panels²⁹. These molecular tests guide the urologist to establish the appropriate treatment and predict recurrence and progression risk after localized treatment. However, it is quite important to keep searching for molecular markers that can aid in the prognosis of the patient as well as the establishment of patient response to different treatments, such as gene fusions, AR variants and non-coding RNAs.

Molecular pathogenesis of prostate cancer.

The molecular pathogenesis of PCa involves several factors such as changes in signal transduction pathways, for example the MAPK and PI3K-AKT pathways, the constant production of angiogenesis inducers, and the activation of the WNT pathway; affecting angiogenesis, adhesion molecule expression, and cell cycle control⁴². Molecular studies have established that in many cases, different changes are caused by genetic predisposition, since several chromosomal regions have been involved in the predisposition of developing PCa⁴³. It has also been observed that PCa predisposition is probably multigenic and caused by different mendelian hereditary models, incomplete penetrance, and ethnic influence⁴⁴. With the advent of massive parallel sequencing RNA and DNA, several genomic changes that contribute to PCa pathogenesis have been deciphered⁴⁵. Besides, loss of a copy of the tumor-suppressor gene *PTEN* has been found in 60% of men suffering from PCa⁴⁶, and other mutations in the genes *p53*, *BRCA1*, and *BRCA2* and loss of the gene *RB* have been reported in smaller proportions^{47,48}. Likewise, PCa progression is mainly dependent on the activation of AR⁴⁹.

Novel prostate cancer biomarkers.

The National Cancer Institute of the United States of America defines a biomarker as a biological molecule that can be detected in blood, tissue or bodily fluids that can be measured and whose quantified values allow the identification of a normal or abnormal process, as well as a disease⁵⁰. From the large variety of molecular markers currently in existence, they can be classified according to the clinical context for which they will be used. For example, there are diagnostic, prognostic, and predictive molecular biomarkers⁵¹. This process has led to the era of Precision Medicine where the selection of treatment is based on the molecular characteristics of the tumor of each patient⁵² (Figure 1). In the following paragraph, we will mention and describe some molecular biomarkers that have recently been reported as useful in PCa patient management.

***TMPRSS2-ERG* gene fusion.**

The *TMPRSS2* gene is located on chromosome 21q22.3 and encodes a transmembrane receptor of the TSP family (type II transmembrane serine protease). It has been reported that the *TMPRSS2* gene is regulated by androgens and expressed in prostate tissue. In contrast, the *ERG* gene is located on chromosome 21q22.2 and encodes a nuclear protein that specifically binds DNA in purine-rich regions and acts as a transcription factor⁵³. Experimental evidence has established that the *TMPRSS2-ERG* gene fusion has important functions in PCa⁵³. The importance of this fusion lies in that it has been identified in approximately 50-79% of PCa cases⁵³⁻⁵⁵.

Some phase I clinical studies, including patients with PCa showing the *TMPRSS2-ERG* fusion, have used PARP1, HDAC1, and DNAPK inhibitors (Table 2). Conversely, phase II clinical studies have focused on inhibiting PARP1 in PCa patients who were positive for ETS gene fusion. Other PARP1 inhibitors that have

been evaluated as monotherapy in CRPC have included olaparib and niraparib⁵⁶. Although the clinical studies are in their initial phases, it is expected that the results will be promising. While there are currently several clinical studies determining the efficiency of PARP1, HDAC1, and DNAPK inhibitors in PCa patients^{57,58}, it is still unknown if these treatments will be effective, therefore, it is important to study these types of gene fusions as therapeutic targets in personalized treatments for PCa patients.

TP53

The *TP53* gene is located on chromosome 17p13 and encodes a nuclear transcriptional factor. It is a tumor-suppressor gene that plays an important role in maintaining genomic stability, in addition to preventing tumorigenesis⁵⁹. It has been established that approximately 3 to 47% of PCa tumors carry mutations in the *TP53* gene⁶⁰, and 2 to 15% have a homozygous deletion in this gene⁶¹. Preliminary studies in mice have established the importance of the *TP53* gene in PCa, in which the R270H mutation was demonstrated to be sufficient to induce PCa in mice⁶², suggesting the importance of mutations in the p53 gene in PCa. Recent studies have established the role of *TP53* as a predictive biomarker, where it has been suggested that patients carrying *TP53* mutations with advanced PCa (metastatic and locally recurrent disease) and who have been treated with humanized monoclonal antibodies, such as bevacizumab, have a higher global survival (HR: 2.50) compared with those that were not treated with bevacizumab⁶³ (Table 3). These clinical studies suggest that *TP53* is a predictive biomarker can be used successfully in the treatment of advanced PCa.

Androgen receptor.

Androgen receptor (AR) is a nuclear transcription factor that binds to the promoters of male sexual genes such as *NKX3-1*, *TMPRSS2*, *FKBP5*, *KLK3*, and *KLK2* and regulates the biological effects of these hormones by the activation of androgen-dependent gene transcription⁶⁴. This gene is located on chromosome X, and sequence analysis has established that it contains a series of CAG repeats⁶⁵. The length of CAG repeats varies among different individuals, and it is known that polymorphisms in these repeats are related to the transcriptional activity of the *AR* gene and a lower number of CAG repeats are associated with an increase in the risk of developing tumors, as well as the aggressiveness of the PCa⁶⁵.

Genomic analyses have determined that approximately 2-18% of PCa tumors contain mutations in androgen receptor^{45,60}, of which 5-52% of tumors have been demonstrated to have AR amplifications⁶¹. In contrast, it has been observed that PCa patients initially respond to androgen deprivation therapy but eventually become resistant to androgen castration therapy. Several mechanisms associated with resistance have been proposed in these types of cases. For example, one resistance mechanism has been associated with the F876L mutation, which was identified in a enzalutamide-resistant clone⁶⁶. Another resistance mechanism occurs through splice variant-7 (AR-V7), which has been associated with resistance to abiraterone and enzalutamide⁶⁷. In addition to the therapies mentioned above, new therapies targeting *AR* are currently being developed and in early clinical study phases (Table 3). This suggests that AR is a predictive biomarker that can be used to determine whether patients with advanced PCa will respond or not to treatment.

BRCA2

One of the genes that has become important in PCa recently is the *BRCA2* gene, which is involved in the maintenance of genomic stability using the double-stranded DNA damage homologous recombination repair pathway⁶⁸. Its importance lies in its mutation in 9% of tumors in PCa. *BRCA2* mutations can serve as prognostic factor in all stages of PCa including localized PCa, underscoring the importance of this gene in PCa⁴⁷. Currently, treatment can be targeted to patients with PCa attributed to mutations in the *BRCA* gene by using novel PARP inhibitors⁶⁹. Evidence on new therapeutical approaches such as the use of carboplatin⁷⁰ in men with castration-resistant prostate cancer, proposes *BRCA2* as a gene with potential clinical utility (PFS HR: 1.425 p-value 0.064, OS HR: 1.706 p-value 0.058 univariate analysis).

Furthermore, whole-exome sequencing and transcriptome studies of tumors before treatment revealed that approximately one-third of patients showed aberrations in genes related to DNA repair mechanisms, while specific changes in the *BRCA2* gene were prevalent in the other 7 patients. In 88% of patients with metastatic castration-resistant PCa who responded to olaparib, the response showed a positive correlation in the analyzed biomarkers, which included *BRCA2*. In this group, Progression-Free Survival (PFS) was higher than in patients with negative values for such biomarkers (median, 9.8 vs. 2.7 months; $P < 0.001$ by the log-rank test). In this case, the findings suggest that tumor sequencing can help identify more efficient treatments for patients, such as olaparib⁷¹, and that *BRCA2* can be a predictive biomarker that can be used in PCa to determine the most suitable treatment, suggesting the importance of predictive markers in clinical practice.

***PTEN/P13K/AKT/mTOR* signaling.**

Signaling pathways involving the *PTEN*, *P13K*, and *AKT* genes are determinants in several growth and cell survival aspects⁷². It has been observed that several genetic changes, such as amplifications, mutations, and chromosomal rearrangements, can compromise the PI3K pathway, causing its permanent activation. The abnormal activation of the PI3K pathway results in changes in growth and cell survival mechanisms, which favors competitive growth, metastatic capacity, and, frequently, increased resistance to treatments⁷³.

In the particular case of PCa, it has been observed that approximately 20% of primary tumors and ~50% castration-resistant tumors contain deletions or mutations in the *PTEN* gene⁷⁴. The importance of mutations in *PTEN* lies in its demonstrated role regulating the *PI3K/AKT* pathway⁷⁵. This function is vital because patients with castration-resistant prostate cancer (CRPC) with changes in the *PI3K* pathway as a consequence of mutations or genetic changes in *PTEN* or *PIK3CA* can be treated with *mTOR* inhibitors such as everolimus (HR: 2.58, 95%, p-value 0.07)⁷⁶. Moreover, clinical studies have demonstrated that everolimus can be administered in patients with CRPC combined with other drugs, such as bicalutamide and docetaxel⁷⁷ (Table 2), which suggests that genes involved in the *PTEN/P13K/AKT/mTOR* pathway can be used as predictive biomarkers to establish the treatment that will be effective in patients with advanced PCa (locally advanced and metastatic).

miRNA

One of the most studied small ncRNAs are microRNAs (miRNAs), these are single stranded RNAs of 21-25 nucleotides in length that regulate the post-transcriptional degradation of messenger RNAs and inhibit their translation into proteins. Because of their high stability in body fluids⁷⁸ as well as to changes of physical and chemical conditions⁷⁹, miRNAs are interesting molecules to be used as biomarkers in cancer.

Free miRNAs can be found in several bodily fluids, such as blood, urine, semen, among others⁸⁰ and their expression levels are tissue-specific and have been found to be deregulated in cancer⁸¹. Moreover, they exhibit differential expression between tumor and normal tissues and are useful for tumor classification according to the lineage of origin, differentiation stage, and tumor aggressiveness⁸². It has been reported that circulating miRNAs can be packed in extracellular vesicles (EV) or in association with proteins such as Argonaute2 or lipoproteins in bio-fluids including blood and urine⁸³⁻⁸⁵. A rich source of miRNAs encapsulated in EV provides a better source of information in comparison to free miRNAs in plasma⁸⁴. Specifically in PCa patients, upregulated and downregulated miRNAs associated with cancer have been identified in peripheral blood^{86,87}. Some miRNAs, such as miR-21, miR-221, miR-1290, and miR-375, have been over-expressed and associated with prognosis in CRPC patients^{84,88}. Yaman *et al.* quantified the levels of miR-21, miR-142, and miR-221 in PCa patients and reported that overexpression of these three miRNAs were associated with an advanced PCa stage⁸⁹. Other groups have identified miRNAs in plasma and serum of patients with locally advanced and metastatic PCa, with BPH and in healthy individuals, showing that differences between each group (i.e., higher levels of miRNAs in patients with locally advanced and metastatic PCa) highlight the role of miRNAs as diagnostic biomarkers⁹⁰. Several groups have studied the diagnostic, prognostic, and predictive characteristics of miRNAs circulating in the plasma and serum of PCa patients finding differentially expressed miRNAs according to the Gleason index⁹¹, response to treatment with docetaxel⁹², and high blood PSA values⁹³. In another study, a panel consisting of four miRNAs was proposed as a biomarker for the diagnosis of prostate cancer⁸⁶. The four miRNAs (miR-4289, miR-326, miR-152-3p and miR-98-5p) were upregulated in plasma of prostate cancer patients compared to healthy controls and was able to differentiate between PCa patients and control individuals with an area under the ROC curve (AUC) of 0.88, proving their diagnostic accuracy. In the study conducted by Sharova *et al.*⁷⁸, a circulating miRNA test consisting of measuring the level of 3 circulating miRNAs (miR-106a, miR-130b and miR-223) was proposed to differentiate between localized PCa and BPH patients. In this test two ratios are calculated: miR-106a/miR-

130b and miR-106a/miR-223 ratios, the results showed a better performance (specificity: 0.806, sensitivity: 0.833, accuracy: 0.821) in comparison to PSA (specificity: 0.065, sensitivity: 0.889, accuracy: 0.507), the area under the ROC curve for miRNA test was 0.84 while for PSA was 0.56. This test could be helpful for PCa screening to avoid unnecessary biopsies and assessment of PCa risk.

Finally, the cellular origin of miRNAs also represents a target of study. The first studies suggest that most miRNAs present in the blood of healthy individuals are released by normal blood cells, while miRNAs associated with diseases such as PCa show the same “mutation signatures” as the neoplastic tissue wherein they originated^{94,95}. Furthermore, the molecular similarities between the components of liquid biopsies (CTCs, EV, miRNAs, and cfDNA) with the tumor tissue from which they originated lend a new perspective for determining the role of each biomolecule as a potential PCa biomarker, however, liquid biopsy information can be reviewed in articles that address this topic in depth from the viewpoint to PCa^{96–98}. Indeed, the use of miRNAs as biomarkers in prostate cancer have shown promising results for risk assessment, diagnosis and prognosis. Implementation of miRNA-based tests could improve the clinical management of prostate cancer patients.

Long non-coding RNAs

As mentioned above, RNA molecules seem to have a critical role in cancer pathways including those within PCa. Similar to microRNAs, long non-coding RNAs (lncRNAs) represent a large subgroup of the human transcriptome⁹⁹. They are known to be RNA transcripts longer than 200 nucleotides with no protein-coding potential¹⁰⁰, these two major differences distinguish them from mRNA transcripts and any other non-coding RNA. lncRNAs have been implicated in several biological processes such as chromatin-reprogramming, imprinting genomic loci, transcriptional regulation in *cis* and *trans* and post-transcriptional regulation of mRNAs^{101–104}.

Among some pathological features in which lncRNAs are involved are cell proliferation, tumorigenesis and malignant transformation¹⁰⁵, this is why several studies have proposed lncRNAs as tumor-suppressor genes and oncogenes^{106,107}. Lately, lncRNAs have drawn the attention of the biomedical community not only because of their critical role in cancer, but because of their potential as molecular biomarkers due to their tissue-specific and tumor-specific expression^{108,105}. Some lncRNAs, such as PCA3, SChLAP1, and PCAT1 have been proposed as good candidates for biomarkers mainly due to their differential expression in prostate cancer patients¹⁰⁹. PCA3 (Prostate Cancer gene 3) is an overexpressed PCa-specific oncogene discovered in 1999 by Bussemakers¹¹⁰. PCA3 is already considered a PCa biomarker and it is measured by the commercial test PROGENSA approved in 2012 by the FDA^{111–113} since its detection in urine samples has reduced ~40% of unnecessary biopsies providing a great utility in urological diagnosis¹¹⁴. PROGENSA PCA3 test has a sensitivity of 62% and a specificity of 75%¹¹⁵ demonstrating why lncRNAs can be one of the molecular markers with clinical utility. Similarly, SChLAP1 (Second Chromosome Locus Associated with Prostate Cancer 1) is known for its high expression levels in PCa. This lncRNA antagonizes the SWI/SNF complex promoting aggressiveness and metastasis of the tumor¹¹⁶. Its effectiveness as a biomarker has been proved by assays such as RNA *in situ* hybridization leading to the development of several tests based on the detection of SChLAP1 expression levels and linking them with the patient's clinical-stage¹¹⁷. Therefore, SChLAP1 is considered as a promising biomarker of clinical utility and one of the best genes for prediction of metastasis and biochemical recurrence in PCa patients^{116,118}. Along with these, Luo and collaborators reported that lncRNA-p21 is overexpressed in neuroendocrine prostate cancer (NEPC) and that a treatment based upon enzalutamide increases its expression, and thus, the neuroendocrine differentiation; all of this is caused by the alteration of the Enz/AR/lncRNA-p21/EZH2/STAT3 axis¹¹⁹. Another example is PCAT1 (Prostate Cancer Associated lncRNA Transcript 1) is an upregulated oncogenic RNA originally identified in prostate cancer by RNA-sequencing analysis¹²⁰. It is related to cell proliferation, apoptosis, migration, and invasion as well as epithelial-mesenchymal

transition and cancer progression via the *Wnt/β-catenin* signaling pathway¹²¹. Finally, PCAT1 negatively regulates BRCA2 tumor suppressor protein, positively regulates Myc oncoprotein¹²² and it might be also acting as a miRNA sponge involved in cell growth¹²⁰. Hence, PCAT1 is considered as a potential biomarker for prostate cancer prognosis and prediction, supporting the statement that lncRNAs represent potential molecular biomarkers in the management of prostate cancer. Most of these candidates and a large number of transcriptional units were found due to the breakthrough of the high-throughput massive sequencing technology, specifically, RNA-seq.

Importance of high-throughput massive sequencing in Prostate Cancer.

DNA and RNA massive parallel sequencing has had a large impact on the generation of new knowledge concerning molecular markers in cancer because it explores the whole genome and transcriptome, allowing the detection of global point mutations, insertions, deletions, variations in copy number, translocations, fusion genes, novel-transcript discovery, transcript abundance estimation, differential gene expression and differential splicing of mRNAs¹²³ (Figure 2). The exome corresponds to the coding regions of the genome (exons), which are the sequences carrying the genetic information for protein synthesis. It represents approximately 1.5% of the genome and it is estimated that 85% of the mutations related to diseases are located in the exome, which highlights the importance of whole-exome sequencing for achieving a general understanding of the molecular pathogenesis of diseases by identifying clinically relevant alterations^{124,125}. On another hand, RNA-seq as a tool in the clinic has led to transcriptome profiling of diseases such as cancer by generating quantitative data of the transcriptional rates and differential splicing in tumor cells. It uses a high-throughput technology and applies the Next Generation Sequencing (NGS) system to sequence the whole transcriptome including not only mRNAs, but also small and long non-coding RNAs¹²⁶. The application of this technology provides a quantitative pattern of genes and non-coding genes with

transcriptional aberrations within the cell in a disease. RNA-seq as an emerging sequencing technology has a promising future in disease diagnosis, prognosis, prediction and treatment¹²⁷.

Among some studies based on RNA sequencing as a potential tool for finding prostate cancer biomarkers and drug targets, Berglund *et al.* analyzed the heterogeneity of prostate cancer through a spatial-transcriptomic study in which several expression profiles were identified within a tissue region obtained after radical prostatectomy (Gs 3+4, pT3b, PSA=7.1). These expression profiles allowed the stratification of the tissue regions into cancer components or groups such as cancer, stroma, reactive stroma, normal glands and Prostatic Intraepithelial Neoplasia (PIN). They also found specific genes as potential biomarkers within the results, for example, *SPINK1*, *PGC*, and *CPP* as specific markers of PCa (Gs 3+3), *NR4A1* as a specific marker of reactive stroma, and *NPY* as a specific marker of PIN. The fact that these markers are expressed in specific and different locations, demonstrates the level of heterogeneity in prostatic tumors and that studies based on sequencing technologies can open the door to the discovery of novel molecular biomarkers¹²⁸.

As is well known, there is an urgent need to classify patients according to the most appropriate and effective therapy to increase the efficacy of treatment and reduce unnecessary interventions that have no effect on the patients (Figure 3). An example of this characteristic is a study supporting the use of exomes in precision medicine has been reported by Robinson *et al.*, who demonstrated that actionable mutations detected with the aid of exomes in castration-resistant PCa patients can help determine the best treatment to use and responses of the patients. The results established a mean rate of 4.4 mutations per Mb, in addition to a gain and loss of chromosome regions, with gains in *AR* and losses in the genes *CHD1*, *PTEN*, *RB1*, and *TP53*. The relevance of this study is that the molecular changes are actionable in 90% of the samples of CRPC patients, and in particular, patients with mutations

in genes such as *BRCA2* (12% of cases) and *ATM* (22% of cases) benefited from treatment with PARP inhibitors (olaparib)¹²⁹. In another study, Armenia *et al.* identified 70 significantly mutated genes that had not been previously associated with PCa, some of them are *CUL3* (a ubiquitin ligase that function as a scaffold in the proteasome system), *SPEN* (a transcription factor involved in repression of gene expression), and *KMT2C* and *KMT2D* (epigenetic regulators with histone-lysine N-methyltransferase activity) by analyzing exome sequencing data from 1013 PCa samples¹³⁰. The markers found in this and other studies could be used as part of gene signatures aimed to stratifying patients with localized and metastatic PCa. Furthermore, recent whole-genome studies have identified mechanisms that generate complex chromosome rearrangements in PCa. Baca *et al.* sequenced the whole genome of 57 prostate tumors and identified several DNA translocations and deletions that arose independently during oncogenesis and progression. They called this phenomenon “chromoplexy” referring to the coordinated and considerable dysregulation of multiple cancer genes supporting a model of punctuated cancer evolution⁶⁰. Therefore, studies based on genomics generate information that could help oncologists to predict the response to treatment, allowing more personalized and effective management of patients with advanced PCa.

This experimental and clinical approach provides information about the emerging responses that current therapies, such as androgen deprivation, and their effect in prostate cancer patients. Sequencing analysis can also provide the necessary data for a more specific and enriched molecular classification of prostate cancer and could provide delineated subtypes among patients for better management¹³¹.

Machine Learning in Prostate Cancer Research

Machine learning is a discipline that teaches computers how to build models from the massive data sets that they are assigned with and learn from them. This

technologic approach is based on statistic algorithms, most of these algorithms are mathematical models that map the variables (features) of a data sample into a set of outcomes^{132,133}. Then, these algorithms go through a process of training to be able to predict the labels by analyzing the features¹³⁴ (Figure 4). It has been proven that machine learning could improve some aspects of biomedicine such as disease diagnosis, monitoring, anatomical imaging of organs, tissue biopsies and personalized treatment by using a collection of molecular and phenotypic data¹³⁵. It has also been proved as useful for its application in the human genome project and advances in cancer research and management¹³⁶.

The advantageous outcome of machine learning also applies to PCa research by improving diagnostic and prognostic accuracy, treatment, imaging, surgical interventions, genomics and transcriptomics. It has been reported that machines can be trained to recognize complex patterns in sequencing data together with radiographic images (such as those generated from computed tomography scanning and magnetic resonance) by classifying pixels for segmentation and registration¹³⁷. Its techniques can identify specific genes or sets of genes within expression profiles and specific expression rate that can predict a certain clinical outcome such as progression, biochemical recurrence or metastasis in PCa¹³⁸. There are commercial genomic classifiers available, such as Decipher, that use the random forest algorithm for prediction of prostate cancer metastasis based on the expression analysis of 22 RNA biomarkers of aggressive prostate cancer^{134,139}. Likewise, deep learning is a branch derived from machine learning than can be used to recognize and classify tissue structures in digital information corresponding to a pathology¹⁴⁰. Tolkach *et al.* developed a trained model based on the technology of deep that recognized tumor tissue from images of 400 histological slides from different patients, as well as a novel algorithm based on three-dimensional reconstruction of prostate cancer architecture that can improve the Gleason grading¹⁴¹. Similarly, there are other algorithms that have been applied in clinics, for example, a Support Vector Machine (SVM) model was used for the detection of positive and negative biopsies through

dynamic contrast-enhanced and diffusion tensor imaging data¹³⁴. Therefore, the development and application of Artificial Intelligence using Machine Learning in clinical practice could open an infinite landscape of approaches within PCa data analysis improving the patient management in a near future.

Conclusions.

Evidence based on clinical studies that focuses on finding new biomarkers suggests that there is a wide molecular field that lies unexplored and that could be the key for many clinical challenges nowadays. These markers, such as the coding and non-coding genes *AR*, *BRCA2*, *PTEN*, *CUL3*, *SPEN*, *PCA3*, *SChLAP1*, and *miR-21*, the *TMPRSS2-ERG* gene fusion including their derivatives, and the androgen receptor variant 7 can be found using genomic, transcriptomics and Machine Learning approaches. However, these are not the only alterations that can be used for the diagnostic, prognostic and prediction in the management of PCa patients. The clinical evidence mentioned in this review suggests the importance of these types of molecular markers and their roles in the decision-making process for establishing the most adequate treatment for patients suffering from this disease. Thus, it is quite important to establish the types of actionable mutations, or their combinations, in patients with advanced PCa (locally advanced and metastatic), allowing us to determine the type of treatment that will provide a positive response in the cancer patient. In this area, genomic analysis, transcriptome sequencing, and new approaches like spatial transcriptomics, along with the clinical-pathological information, could provide the necessary information. Likewise, the application of Machine Learning algorithms will accelerate the identification and discovery of novel molecular biomarkers and it will lead biomedical investigation towards artificial-intelligence-based precision medicine, so it can improve patient management as well as their quality of life, and, in the near future, allow a scientific revolution in medicine for the management of the prostate cancer patient.

ACKNOWLEDGMENTS

We thank to National Cancer Institute of Mexico (INCan) for their support. Figures were created with BioRender.com.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

Figures and figure legends

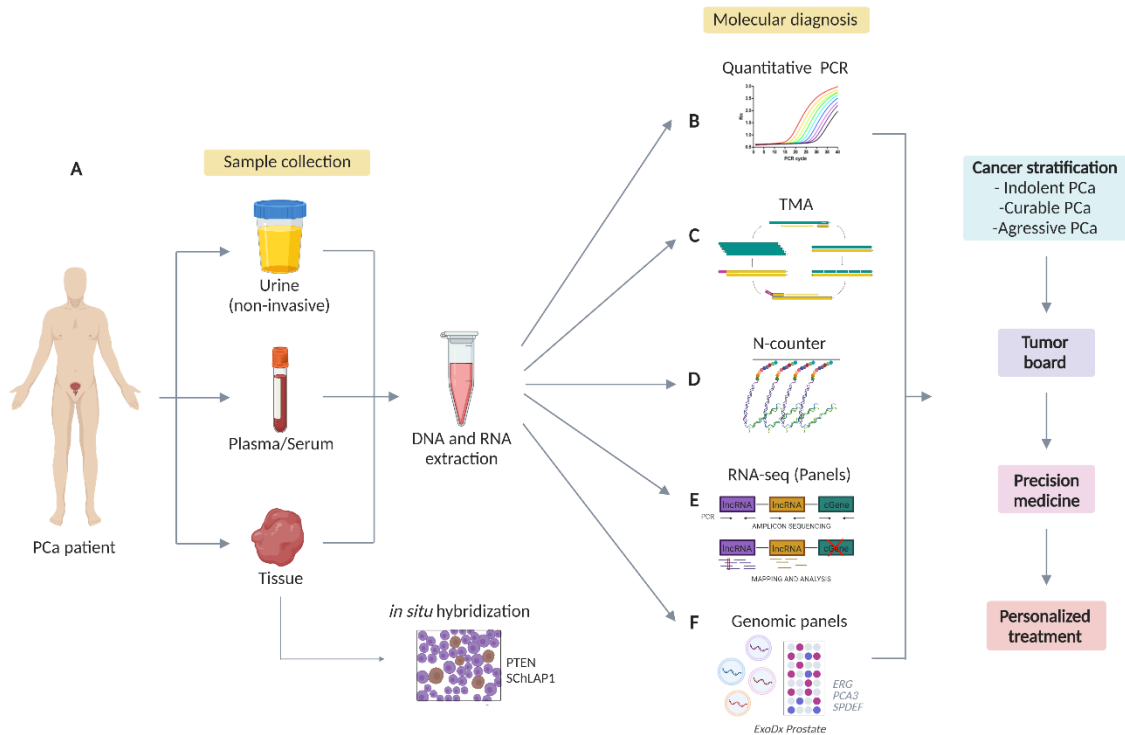


Figure 1. Landscape of precision medicine and molecular tools in prostate cancer. Molecular approaches have a lot of potential as a clinical resource for prevention, detection and treatment of the disease based on the specific molecular profile of the PCa patients. **A)** Sample collection. This can be achieved by sampling tissue, blood or even urine (a non-invasive sampling) from the patient and proceeding with a direct detection of the biomarker by *in situ* hybridization in the tissue sample or a nucleic acid extraction and a molecular assay. **B)** Quantitative PCR (qPCR). This molecular tool can be used to quantify gene expression by determining the amount of a target sequence present in the sample based on fluorescent emission, such as My Prostate Score¹⁴². **C)** Transcription Mediated Amplification (TMA). It is based on the biological process of retroviral replication, in which an RNA target is converted to DNA and then approached and amplified by an RNA polymerase to produce RNA copies from a promoter added to the primer region. PROGENSA is a current test based on a TMA assay. **D)** N-counter. It is a highly multiplexed single molecule counting system where two probes are used to target the RNA molecule of interest, a capture probe and a reporter probe. Once the probes and the RNA molecule hybridize, the complexes are purified and immobilized for the last step of counting the barcodes of the reporter probes, this is an advantageous test due to its immediacy. Dong *et al.* used the NanoString nCounter

assay to target mRNA transcripts in EVs from PCA cell lines¹⁴³. **E**) RNA Sequencing. RNA samples are synthesized into cDNA once it has been fragmented. Then, adaptors are attached to both ends of each fragment so they can be amplified by PCR and subsequently sequenced. The RNA massive sequencing allows analyzing the entire transcriptome, if wanted; and even transcripts yet to be discovered. All the techniques mentioned before could be used in precision medicine so the patients can be treated according to their specific molecular profile. **F**) There are also genomic panels used for the diagnostic of prostate cancer focused on specific biomarkers, such as the commercial test ExoDx Prostate¹⁴⁴ that detects the expression levels of *ERG*, *PCA3* and *SPDEF* by qPCR in exosomes from urine samples.

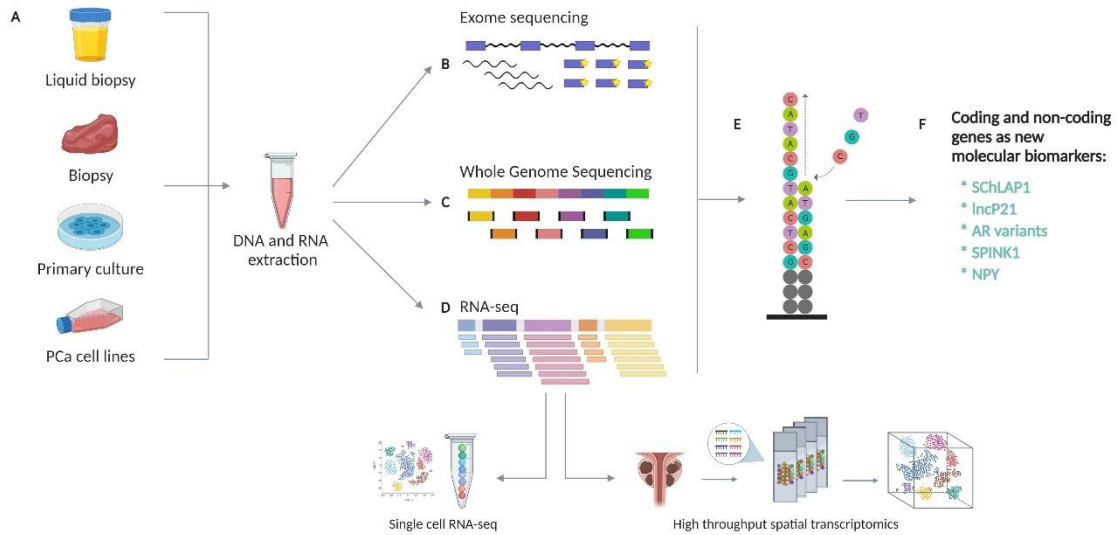


Figure 2. Basic research towards the discovery of new molecular biomarkers.

There are several sources and molecular approaches for the detection of new biomarkers in prostate cancer. **A**) This can be achieved by using an *in vivo* model - for which a prostate biopsy should be taken-, a primary culture or a prostate cancer cell line. Once the nucleic acid extraction has been done, a sequencing technique and the pipeline to be carried out can be chosen, this decision must be made based on the aim of the study. **B**) Exome sequencing. The DNA samples are first fragmented and then biotinylated oligonucleotide probes -also known as *baits*- are used to selectively hybridize to target regions in the genome, in this case, the exonic regions. Later, magnetic streptavidin beads are used to bind to the *baits* and the intronic region is washed away. Finally, PCR assay is used to enriched the DNA target region. **C**) Whole Genome Sequencing. Genomic DNA is randomly fragmented and DNA bar-coding tags are added to each fraction so it can be sequenced later. This sequencing technique allows a uniform coverage across the

complete genome. **D)** RNA-Seq. RNA samples are synthesized into cDNA once it has been fragmented. Then, adaptors are attached to both ends of each fragment so they can be amplified by PCR and subsequently sequenced¹⁴⁵. Within the variants of this technique can be found single-cell RNA-seq, total RNA-seq, targeted RNA-seq, small RNA-seq, spatial transcriptomics, poly-A enrichment, ribosomal RNA depletion, among others. **E)** Illumina next-generation sequencing technology: Individual DNA or cDNA molecules are placed on a flowcell for sequencing by synthesis by using fluorescent labelled nucleotides. As the DNA grow, it is imaged so each of the fluorophores that are being incorporated to the strand can be detected. In recent years new sequencing tools and sequencers have been developed, such as the PacBio sequencer and Nanopore sequencer that can read more than 100 Kb in length of DNA, as well as the disposable sequencer MinION which doesn't need prior installation¹⁴⁶. **F)** After a bioinformatic data analysis the results of the sequencing provide new genes as biomarkers candidates in prostate cancer.

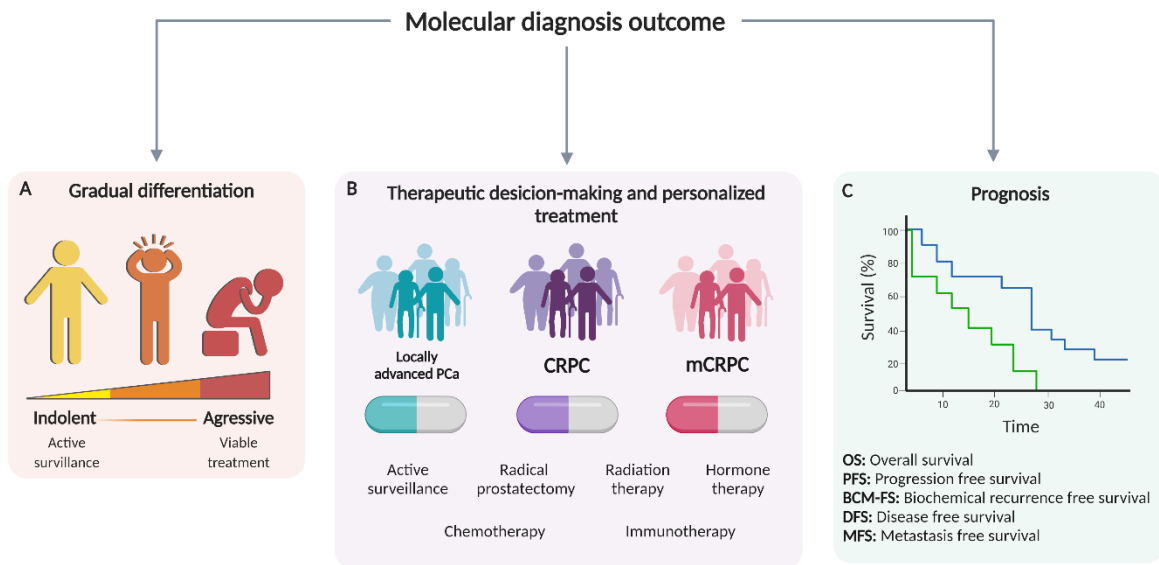


Figure 3. Molecular diagnosis and precision medicine in cancer patient management. The advantages of using approaches that target these signatures for disease diagnosis can be many. **A)** PCa-stage discrimination. Each clinical-pathological profile will be stratified using the emerging techniques mentioned earlier, the results obtained could determine whether the patient has an indolent cancer or if it is an aggressive one. **B)** Personalized treatment. The molecular diagnosis can also determine which specific treatment the patient should receive according to their molecular profile and the type of PCa they have, such as active

surveillance, hormone therapy, surgery, radiation therapy, chemotherapy or immunotherapy (e.g. PD-1 inhibitors, sipuleucel-T vaccine¹⁴⁷). **C)** Prognosis. Finishing by getting a prognostic overview of the length of time that the patient will be alive or how well will the patient respond to the treatment he has been given. The prognosis can include overall-survival, progression-free survival, biochemical recurrence-free survival, disease-free survival, cancer-specific survival, and metastasis-free survival.

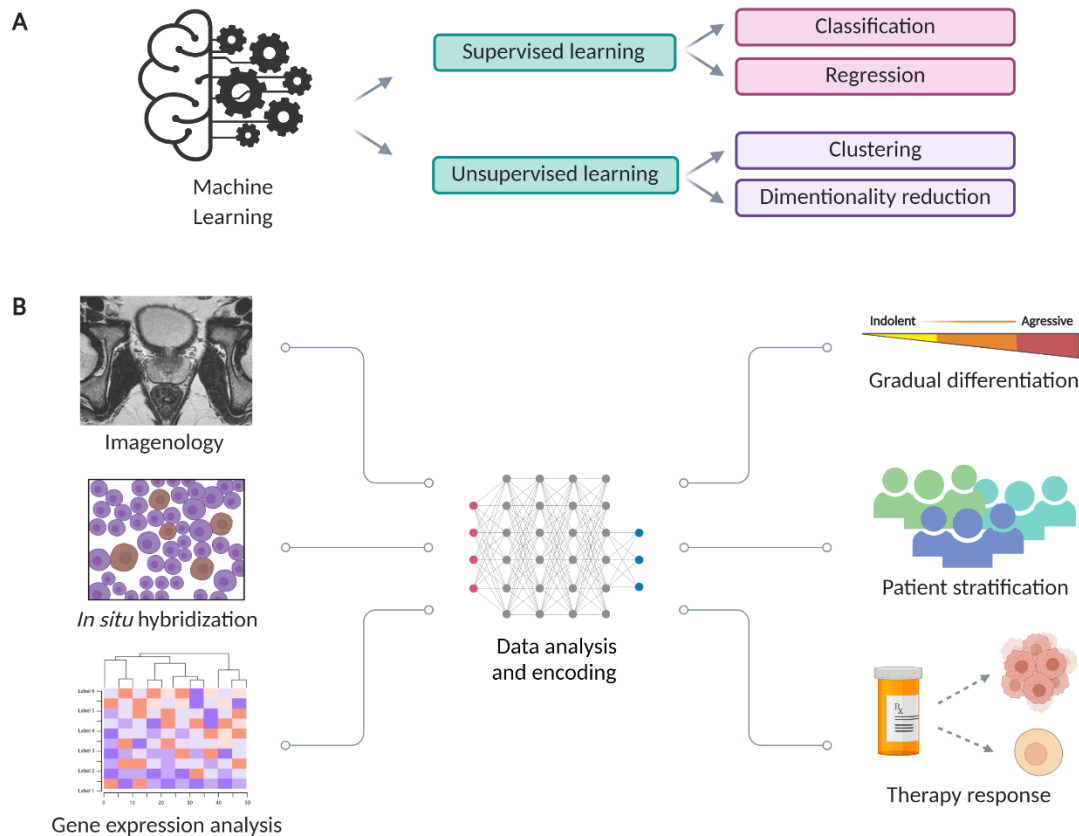


Figure 4. Machine learning and its application in patient stratification in prostate cancer. Machine learning (ML) is an artificial intelligence approach that can predict a possible outcome in prostate cancer research and improve the patient management. **A)** ML techniques. These algorithms are divided into two main types of learning: supervised learning and unsupervised learning. The former uses pre-determined explicit data, it is the most used in radiology and is based on classification and regression (deep learning, convolutional neural network, random forest, support vector machine, decision tree, logistic regression, among others¹⁴⁸). The latter uses the features of the training data and doesn't have a prior division of data in categories, it is based on clustering and dimensional reduction (K-means,

hierarchical clustering, among others). **B)** ML applied in prostate cancer management. A recent application of machine learning is the prediction and analysis of radiomic data. This approach aims to improve the patient stratification and management using imageology, tissue analysis, and molecular data so the clinicians can offer a personalized treatment by differentiating the grade of the disease, stratifying the patients, and determine the therapy response.

Tables

Table 1. Gleason Grading System.

*ISUP group	Gleason score	Histological architecture observed	Risk group and predicted outcome
Group 1 (ISUP1)	GS 6 or less	Only individual discrete well-formed glands.	Low-risk. Excellent prognosis with no potential of lymph node metastasis.
Group 2 (ISUP2)	GS 3+4=7	Predominantly well-formed glands with lesser component of poorly-formed glands.	Intermediate-risk. Very good prognosis with rare metastasis.
Group 3 (ISUP3)	GS 4+3=7	Predominantly poorly-formed glands with lesser component of well-formed glands.	Intermediate-risk. Significantly worse prognosis than group 2.
Group 4 (ISUP4)	GS 4+4=8	Only poorly-formed glands.	High-risk. Significantly better prognosis than group 5.
	GS 3+5=8	Predominantly well-formed glands and lesser component of lacking glands.	
	GS 5+3=8	Predominantly lacking glands and lesser component of well-formed glands.	
Group 5 (ISUP5)	GS 4+5=9	Lacks gland formation (or with necrosis) with or without poorly-formed glands.	High-risk. Poor prognosis with potential of metastasis.
	GS 5+4=9		
	GS 5+5=10		

*ISUP: International Society of Urological Pathology. GS: Gleason score.

Table 2. Molecular biomarkers and their associated drugs in prostate cancer.

Biomarker	Drug	Clinical Trial ID	Reference
TMPRSS2-ERG gene fusion	Abiraterone Acetate, Prednisone, Veliparib	NCT01576172	N/A
	Vorinostat	NCT00330161	N/A
	Romidepsin (histone deacetylase inhibitor)	Preclinical study	58
	Olaparib (PARP1 inhibitor)	NCT01682772	N/A
	MK-4827	NCT00749502	N/A
p53	Enzalutamide	NCT03903835	N/A
	APR-246	NCT00900614	N/A
	Bevacizumab	Preclinical study	63
AR	Enzalutamide	Preclinical study	67
	Arbiraterone	Preclinical study	67
	MDV3100	Preclinical study	149
	Degarelix, Enzalutamide, Trametinib, Dasatinib	Preclinical study	150,151,152,153
	Apalutemide, Androgen deprivation, Docetaxel	Preclinical study	154,155,156,157
	Sodium Fluoride (NaF) Positron Emission Tomography/Computed Tomography (PET/CT) imaging	Preclinical study	158
	pTVG-AR	Preclinical study	159
	Cabazitaxel, Enzalutamida, Abiraterone Acetate, Prednisone	Preclinical study	160,161
	ARN-509, LHRH agonist	Preclinical study	162,163
	Firmagon (degarelix)	Preclinical study	164
BRCA2	Docetaxel	Preclinical study	165
	Arbiraterone	Preclinical study	166
	Enzalutamide	Preclinical study	167
	Olaparib	Preclinical study	26
	Carboplatin	Preclinical study	70
PTEN/P13K/AKT/mTOR	Everolimus	Preclinical study	168,169
	Bicalutamide	Preclinical study	170
	Docetaxel	Preclinical study	171

	Akt inhibitor MK2206, Biclutamide,	Preclinical study	172
	AZD5363, Enzalutamide	Preclinical study	173,174
Circulating tumor cells	Temsirolimus, Diphenhydramine	Preclinical study	175,176,177,178
	huJ591	Preclinical study	179
	Cabazitaxel	Preclinical study	180
Cell-free DNA	Arbiterone	Preclinical study	181,182
	Enzalutamide	Preclinical study	181,183
miRNAs	Docetaxel	Preclinical study	184
	Enzalutamide	Preclinical study	185
	ARV7 y valoración de miRNAs	Preclinical study	186
	Bicalutamide, Leuprolide, Goserelin, Triptorelin, Docetaxel	Preclinical study	187,188
lncRNAs	Enzalutamide	Preclinical study	189

N/A: Not applicable.

Table 3. PCa-associated biomarkers approached by clinical trials

Biomarker	Clinical trial phase	Type of cancer	Patients included	Clinical trial ID
TMPRSS2-ERG	Phase II	mCRPC and recurrent PCa	148	NCT01576172
	Phase II	Recurrent PCa, stage IV PCa	29	NCT00330161
	Phase II	Prostatic adenocarcinoma	148	NCT01682772
	Phase I	Advanced or metastatic PCa	113	NCT00749502
	Phase I	High risk PCa	65	NCT02588404
	Phase I	Localized or locally advanced PCa, biochemical recurrent PCa	84	NCT03421015
	Phase II	High risk PCa	208	NCT02573636
p53	Phase III	mCRPC	750	NCT03903835
	Phase I/II	Prostatic neoplasia	36	NCT00900614
	Phase III	Localized PC	7776	NCT00001469
	Phase I	Localized or locally advanced PCa, biochemical recurrent PCa	84	NCT03421015
AR	Phase I	Hormone refractory PCa	140	NCT00510718
	Phase II	PCa	45	NCT01990196
	Phase II	Recurrent PCa	42	NCT03311555
	Phase I	mCRPC	58	NCT01516866
	Phase II	Metastatic PCa, CRPC	60	NCT04090528
	Phase I	PCa	40	NCT02411786
	Phase II	mCRPC	8	NCT02379390
	Phase II	Biochemical recurrent PCa	90	NCT01790126
	Phase II	Advanced hormone dependent PCa	90	NCT01861236
BRCA2	Phase II	High risk PCa	100	NCT02154672
	Phase III	mCRPC	408	NCT03075735

	Phase III	Genetic predisposition to PCa	1700	NCT00261456
	Phase II	mCRPC	40	NCT04038502
	Phase II	mCRPC	70	NCT03012321
	Phase III	mCRPC	387	NCT02987543
PTEN/P13K/AKT/mTOR	Phase II	High risk PCa	208	NCT02573636
	Phase III	CRPC	120	NCT03580239
	Phase II	PCa previously treated	108	NCT01251861
	Phase I	PCa previously treated with enzalutamide	36	NCT03310541
	Phase I	Stage III and IV PCa	62	NCT01480154
	Phase II	mCRPC	9	NCT02091531
Circulating tumor cells	Not applicable	Obese men with PCa	67	NCT02453139
	Phase II	Patients with PSA 4 - 10 ng/mL	500	NCT03488706
	Phase II	Localized prostate cancer	200	NCT01961713
	Phase II	mCRPC	11	NCT00887640
	Phase II	Advanced PCa	24	NCT02552394
	Phase II	mCRPC	140	NCT03050866
	Phase I	PCa	60	NCT02450435
cell-free DNA	Phase III	Metastatic PCa	1038	NCT00134056
	Phase I	PCa	12	NCT04081428
	Phase II	Metastatic PCa	300	NCT02853097
	Phase II	PCa	68	NCT02941029
	Phase II	PCa	30	NCT03284684
Extracellular vesicles	Not applicable	PCa	108	NCT04298398
miRNAs	Phase II	mCRPC	40	NCT02471469
	Phase II	mCRPC	46	NCT04188275
	Phase III	High risk PCa	300	NCT01220427
	Phase I	PCa	240	NCT03911999
	Phase I	PCa	60	NCT02366494
lncRNAs	Not applicable	Prostatic neoplasia	507	NCT01024959

mCRPC: Metastatic castration resistance prostate cancer, CRPC: Castration resistance prostate cancer, PCa: prostate cancer, PSA: Prostate Specific Antigen. Information obtained from <https://clinicaltrials.gov/> (Last accessed: February 25, 2021)

REFERENCES

- 1 Hoffman RM. Clinical practice. Screening for prostate cancer. *N Engl J Med* 2011; **365**: 2013–2019.
- 2 M.D RWR. *Cancer Biology*. Oxford University Press, 2007.
- 3 Cyll K, Ersvær E, Vlatkovic L, Pradhan M, Kildal W, Avranden Kjær M *et al*. Tumour heterogeneity poses a significant challenge to cancer biomarker research. *Br J Cancer* 2017; **117**: 367–375.
- 4 Wang G, Zhao D, Spring DJ, DePinho RA. Genetics and biology of prostate cancer. *Genes Dev* 2018; **32**: 1105–1140.
- 5 Dicker AP MD, Kelly WK, Trabulsi EJ, Zaorsky NG. *Prostate Cancer: A Multidisciplinary Approach to Diagnosis and Management*. Demos Medical Publishing, 2014.
- 6 Park S-Y, Haiman CA, Cheng I, Park SL, Wilkens LR, Kolonel LN *et al*. Racial/ethnic differences in lifestyle-related factors and prostate cancer risk: the Multiethnic Cohort Study. *Cancer Causes Control* 2015; **26**: 1507–1515.
- 7 Rosenberg MT, Froehner M, Albala D, Miner MM. Biology and natural history of prostate cancer and the role of chemoprevention. *International Journal of Clinical Practice* 2010; **64**: 1746–1753.
- 8 Descotes J-L. Diagnosis of prostate cancer. *Asian J Urol* 2019; **6**: 129–136.
- 9 Teo MY, Rathkopf DE, Kantoff P. Treatment of Advanced Prostate Cancer. *Annu Rev Med* 2019; **70**: 479–499.
- 10 Salinas CA, Tsodikov A, Ishak-Howard M, Cooney KA. Prostate cancer in young men: an important clinical entity. *Nat Rev Urol* 2014; **11**: 317–323.
- 11 Rahman N. Realizing the promise of cancer predisposition genes. *Nature* 2014; **505**: 302–308.
- 12 Ni Raghallaigh H, Eeles R. Genetic predisposition to prostate cancer: an update. *Familial Cancer* 2021. doi:10.1007/s10689-021-00227-3.
- 13 Pritzlaff M, Tian Y, Reineke P, Stuenkel AJ, Allen K, Gutierrez S *et al*. Diagnosing hereditary cancer predisposition in men with prostate cancer. *Genet Med* 2020; **22**: 1517–1523.
- 14 Benafif S, Kote-Jarai Z, Eeles RA, PRACTICAL Consortium. A Review of Prostate Cancer Genome-Wide Association Studies (GWAS). *Cancer Epidemiol Biomarkers Prev* 2018; **27**: 845–857.
- 15 Beebe-Dimmer JL, Kapron AL, Fraser AM, Smith KR, Cooney KA. Risk of Prostate Cancer Associated With Familial and Hereditary Cancer Syndromes. *J Clin Oncol* 2020; **38**: 1807–1813.
- 16 D G, J M, A S, F R, M S, Aw W *et al*. Practical considerations for optimising homologous recombination repair mutation testing in patients with metastatic prostate cancer. *J Pathol Clin Res* 2021. doi:10.1002/cjp2.203.

- 17 Merseburger AS, Waldron N, Ribal MJ, Heidenreich A, Perner S, Fizazi K *et al.* Genomic Testing in Patients with Metastatic Castration-resistant Prostate Cancer: A Pragmatic Guide for Clinicians. *European Urology* 2021; **0**. doi:10.1016/j.eururo.2020.12.039.
- 18 AlDubayan SH. Considerations of multigene test findings among men with prostate cancer - knowns and unknowns. *Can J Urol* 2019; **26**: 14–16.
- 19 Heidegger I, Tsaour I, Borgmann H, Surcel C, Kretschmer A, Mathieu R *et al.* Hereditary prostate cancer - Primetime for genetic testing? *Cancer Treat Rev* 2019; **81**: 101927.
- 20 Giri VN, Hegarty SE, Hyatt C, O’Leary E, Garcia J, Knudsen KE *et al.* Germline genetic testing for inherited prostate cancer in practice: Implications for genetic testing, precision therapy, and cascade testing. *Prostate* 2019; **79**: 333–339.
- 21 Huang J, Wang JK, Sun Y. Molecular pathology of prostate cancer revealed by next-generation sequencing: opportunities for genome-based personalized therapy. *Current Opinion in Urology* 2013; **23**: 189–193.
- 22 Simon R, Roychowdhury S. Implementing personalized cancer genomics in clinical trials. *Nat Rev Drug Discov* 2013; **12**: 358–369.
- 23 Professionals S-O. EAU Guidelines: Prostate Cancer. Uroweb. <https://uroweb.org/guideline/prostate-cancer/> (accessed 25 Jan2021).
- 24 Mateo J, Boysen G, Barbieri CE, Bryant HE, Castro E, Nelson PS *et al.* DNA Repair in Prostate Cancer: Biology and Clinical Implications. *European Urology* 2017; **71**: 417–425.
- 25 Boyle JL, Hahn AW, Kapron AL, Kohlmann W, Greenberg SE, Parnell TJ *et al.* Pathogenic Germline DNA Repair Gene and HOXB13 Mutations in Men With Metastatic Prostate Cancer. *JCO Precision Oncology* 2020; : 139–151.
- 26 Bono JD, Mateo J, Fizazi K, Saad F, Shore N, Sandhu S *et al.* Olaparib for metastatic castration-resistant prostate cancer. *NEW ENGL JMED* 2020; **382**: 2091–2102.
- 27 Scheinberg T, Goodwin A, Ip E, Linton A, Mak B, Smith DP *et al.* Evaluation of a Mainstream Model of Genetic Testing for Men With Prostate Cancer. *JCO Oncol Pract* 2021; **17**: e204–e216.
- 28 Saini S. PSA and beyond: alternative prostate cancer biomarkers. *Cell Oncol (Dordr)* 2016; **39**: 97–106.
- 29 Kretschmer A, Tilki D. Biomarkers in prostate cancer - Current clinical utility and future perspectives. *Crit Rev Oncol Hematol* 2017; **120**: 180–193.
- 30 Paner GP, Stadler WM, Hansel DE, Montironi R, Lin DW, Amin MB. Updates in the Eighth Edition of the Tumor-Node-Metastasis Staging Classification for Urologic Cancers. *Eur Urol* 2018; **73**: 560–569.
- 31 Amin MB, Edge S, Greene F, Byrd DR, Brookland RK, Washington MK *et al.* (eds.). *AJCC Cancer Staging Manual*. 8th ed. Springer International Publishing, 2017 <https://www.springer.com/gp/book/9783319406176> (accessed 8 Feb2021).
- 32 Epstein JI, Egevad L, Amin MB, Delahunt B, Srigley JR, Humphrey PA *et al.* The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on

Gleason Grading of Prostatic Carcinoma: Definition of Grading Patterns and Proposal for a New Grading System. *Am J Surg Pathol* 2016; **40**: 244–252.

33 De Nunzio C, Pastore AL, Lombardo R, Simone G, Leonardo C, Mastroianni R *et al*. The new Epstein gleason score classification significantly reduces upgrading in prostate cancer patients. *Eur J Surg Oncol* 2018; **44**: 835–839.

34 Welti J, Rodrigues DN, Sharp A, Sun S, Lorente D, Riisnaes R *et al*. Analytical Validation and Clinical Qualification of a New Immunohistochemical Assay for Androgen Receptor Splice Variant-7 Protein Expression in Metastatic Castration-resistant Prostate Cancer. *European Urology* 2016; **70**: 599–608.

35 Van der Kwast TH. Prognostic prostate tissue biomarkers of potential clinical use. *Virchows Arch* 2014; **464**: 293–300.

36 Bazzichetto C, Conciatori F, Pallocca M, Falcone I, Fanciulli M, Cognetti F *et al*. PTEN as a Prognostic/Predictive Biomarker in Cancer: An Unfulfilled Promise? *Cancers (Basel)* 2019; **11**. doi:10.3390/cancers11040435.

37 Lotan TL, Tomlins SA, Bismar TA, Van der Kwast TH, Grignon D, Egevad L *et al*. Report From the International Society of Urological Pathology (ISUP) Consultation Conference on Molecular Pathology of Urogenital Cancers. I. Molecular Biomarkers in Prostate Cancer. *Am J Surg Pathol* 2020; **44**: e15–e29.

38 Cullen J, Rosner IL, Brand TC, Zhang N, Tsiatis AC, Moncur J *et al*. A Biopsy-based 17-gene Genomic Prostate Score Predicts Recurrence After Radical Prostatectomy and Adverse Surgical Pathology in a Racially Diverse Population of Men with Clinically Low- and Intermediate-risk Prostate Cancer. *Eur Urol* 2015; **68**: 123–131.

39 Health Quality Ontario. Prolaris Cell Cycle Progression Test for Localized Prostate Cancer: A Health Technology Assessment. *Ont Health Technol Assess Ser* 2017; **17**: 1–75.

40 Blume-Jensen P, Berman DM, Rimm DL, Shipitsin M, Putzi M, Nifong TP *et al*. Development and clinical validation of an in situ biopsy-based multimarker assay for risk stratification in prostate cancer. *Clin Cancer Res* 2015; **21**: 2591–2600.

41 Spratt DE, Yousefi K, Deheshi S, Ross AE, Den RB, Schaeffer EM *et al*. Individual Patient-Level Meta-Analysis of the Performance of the Decipher Genomic Classifier in High-Risk Men After Prostatectomy to Predict Development of Metastatic Disease. *J Clin Oncol* 2017; **35**: 1991–1998.

42 Shtivelman E, Beer TM, Evans CP. Molecular pathways and targets in prostate cancer. *Oncotarget* 2014; **5**: 7217–7259.

43 Wallis CJD, Nam RK. Prostate Cancer Genetics: A Review. *EJIFCC* 2015; **26**: 79–91.

44 Pomerantz MM, Freedman ML. The Genetics of Cancer Risk. *Cancer J* 2011; **17**: 416–422.

45 Berger MF, Lawrence MS, Demichelis F, Drier Y, Cibulskis K, Sivachenko AY *et al*. The genomic complexity of primary human prostate cancer. *Nature* 2011; **470**: 214–220.

46 Phin S, Moore M, Cotter PD. Genomic Rearrangements of PTEN in Prostate Cancer. *Front Oncol* 2013; **3**. doi:10.3389/fonc.2013.00240.

- 47 Castro E, Goh C, Olmos D, Saunders E, Leongamornlert D, Tymrakiewicz M *et al.* Germline BRCA Mutations Are Associated With Higher Risk of Nodal Involvement, Distant Metastasis, and Poor Survival Outcomes in Prostate Cancer. *JCO* 2013; **31**: 1748–1757.
- 48 Ecke TH, Schlechte HH, Schiemenz K, Sachs MD, Lenk SV, Rudolph BD *et al.* TP53 Gene Mutations in Prostate Cancer Progression. *Anticancer Research* 2010; **30**: 1579–1586.
- 49 Schrecengost R, Knudsen KE. Molecular Pathogenesis and Progression of Prostate Cancer. *Seminars in Oncology* 2013; **40**: 244–258.
- 50 Goossens N, Nakagawa S, Sun X, Hoshida Y. Cancer biomarker discovery and validation. *Translational Cancer Research* 2015; **4**. doi:10.21037/4536.
- 51 FDA-NIH Biomarker Working Group. *BEST (Biomarkers, EndpointS, and other Tools) Resource*. Food and Drug Administration (US): Silver Spring (MD), 2016 <http://www.ncbi.nlm.nih.gov/books/NBK326791/> (accessed 28 Jun2021).
- 52 Rubin EH, Allen JD, Nowak JA, Bates SE. Developing Precision Medicine in a Global World. *Clin Cancer Res* 2014; **20**: 1419–1427.
- 53 Burdova A, Bouchal J, Tavandzis S, Kolar Z. TMPRSS2-ERG gene fusion in prostate cancer. *Biomedical Papers* 2014; **158**: 502–510.
- 54 Barbieri CE, Baca SC, Lawrence MS, Demichelis F, Blattner M, Theurillat J-P *et al.* Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nature Genetics* 2012; **44**: 685–689.
- 55 Tu JJ, Rohan S, Kao J, Kitabayashi N, Mathew S, Chen Y-T. Gene fusions between TMPRSS2 and ETS family genes in prostate cancer: frequency and transcript variant analysis by RT-PCR and FISH on paraffin-embedded tissues. *Modern Pathology* 2007; **20**: 921–928.
- 56 Feng FY, Brenner JC, Hussain M, Chinnaiyan AM. Molecular Pathways: Targeting ETS Gene Fusions in Cancer. *Clin Cancer Res* 2014; **20**: 4442–4448.
- 57 Bradley D, Rathkopf D, Dunn R, Stadler WM, Liu G, Smith DC *et al.* Vorinostat in advanced prostate cancer patients progressing on prior chemotherapy (National Cancer Institute Trial 6862). *Cancer* 2009; **115**: 5541–5549.
- 58 Molife LR, Attard G, Fong PC, Karavasilis V, Reid AHM, Patterson S *et al.* Phase II, two-stage, single-arm trial of the histone deacetylase inhibitor (HDACi) romidepsin in metastatic castration-resistant prostate cancer (CRPC). *Annals of Oncology* 2010; **21**: 109–113.
- 59 Rivlin N, Brosh R, Oren M, Rotter V. Mutations in the p53 Tumor Suppressor Gene: Important Milestones at the Various Steps of Tumorigenesis. *Genes & Cancer* 2011; **2**: 466–474.
- 60 Baca SC, Prandi D, Lawrence MS, Mosquera JM, Romanel A, Drier Y *et al.* Punctuated Evolution of Prostate Cancer Genomes. *Cell* 2013; **153**: 666–677.
- 61 Robinson D, Van Allen EM, Wu Y-M, Schultz N, Lonigro RJ, Mosquera J-M *et al.* Integrative Clinical Genomics of Advanced Prostate Cancer. *Cell* 2015; **161**: 1215–1228.

- 62 Vinall RL, Chen JQ, Hubbard NE, Sulaimon SS, Shen MM, White RWD *et al.* Initiation of prostate cancer in mice by Tp53R270H: evidence for an alternative molecular progression. *Disease Models & Mechanisms* 2012; **5**: 914–920.
- 63 Said R, Hong DS, Warneke CL, Lee JJ, Wheler JJ, Janku F *et al.* P53 Mutations in Advanced Cancers: Clinical Characteristics, Outcomes, and Correlation between Progression-Free Survival and Bevacizumab-Containing Therapy. *Oncotarget* 2013; **4**: 705–714.
- 64 Wang Q, Li W, Zhang Y, Yuan X, Xu K, Yu J *et al.* Androgen Receptor Regulates a Distinct Transcription Program in Androgen-Independent Prostate Cancer. *Cell* 2009; **138**: 245–256.
- 65 Tan ME, Li J, Xu HE, Melcher K, Yong E. Androgen receptor: structure, role in prostate cancer and drug discovery. *Acta Pharmacologica Sinica* 2015; **36**: 3–23.
- 66 Joseph JD, Lu N, Qian J, Sensintaffar J, Shao G, Brigham D *et al.* A Clinically Relevant Androgen Receptor Mutation Confers Resistance to Second-Generation Antiandrogens Enzalutamide and ARN-509. *Cancer Discov* 2013; **3**: 1020–1029.
- 67 Antonarakis ES, Lu C, Wang H, Luber B, Nakazawa M, Roeser JC *et al.* AR-V7 and Resistance to Enzalutamide and Abiraterone in Prostate Cancer. <http://dx.doi.org/10.1056/NEJMoa1315815>. 2014. doi:10.1056/NEJMoa1315815.
- 68 Roy R, Chun J, Powell SN. BRCA1 and BRCA2: different roles in a common pathway of genome protection. *Nature Reviews Cancer* 2012; **12**: 68–78.
- 69 Kaufman B, Shapira-Frommer R, Schmutzler RK, Audeh MW, Friedlander M, Balmaña J *et al.* Olaparib Monotherapy in Patients With Advanced Cancer and a Germline BRCA1/2 Mutation. *JCO* 2014; **33**: 244–250.
- 70 Fujiwara M, Akamatsu S, Sumiyoshi T, Segawa T, Mizuno K, Yoshino T *et al.* Efficacy and Safety of Carboplatin Plus Paclitaxel as the First-, Second-, and Third-line Chemotherapy in Men With Castration-resistant Prostate Cancer. *Clin Genitourin Cancer* 2019; **17**: e923–e929.
- 71 Taneja Samir S. Re: DNA-Repair Defects and Olaparib in Metastatic Prostate Cancer. *Journal of Urology* 2016; **195**: 925–928.
- 72 Chalhoub N, Baker SJ. PTEN and the PI3-Kinase Pathway in Cancer. *Annu Rev Pathol* 2009; **4**: 127–150.
- 73 Porta C, Paglino C, Mosca A. Targeting PI3K/Akt/mTOR Signaling in Cancer. *Front Oncol* 2014; **4**. doi:10.3389/fonc.2014.00064.
- 74 Jamaspishvili T, Berman DM, Ross AE, Scher HI, De Marzo AM, Squire JA *et al.* Clinical implications of PTEN loss in prostate cancer. *Nat Rev Urol* 2018; **15**: 222–234.
- 75 Huret J-L, Ahmad M, Arsaban M, Bernheim A, Cigna J, Desangles F *et al.* Atlas of Genetics and Cytogenetics in Oncology and Haematology in 2013. *Nucleic Acids Research* 2013; **41**: D920–D924.
- 76 Templeton AJ, Dutoit V, Cathomas R, Rothermundt C, Bärtschi D, Dröge C *et al.* Phase 2 Trial of Single-agent Everolimus in Chemotherapy-naïve Patients with Castration-resistant Prostate Cancer (SAKK 08/08). *European Urology* 2013; **64**: 150–158.

- 77 Wang C, Li Y, Fang X. Research progress on liquid biopsy in oncology and its clinical applications. *Yi Chuan* 2017; **39**: 220–231.
- 78 Sharova E, Grassi A, Marcer A, Ruggero K, Pinto F, Bassi P *et al.* A circulating miRNA assay as a first-line test for prostate cancer screening. *Br J Cancer* 2016; **114**: 1362–1366.
- 79 O'Brien J, Hayder H, Zayed Y, Peng C. Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. *Front Endocrinol (Lausanne)* 2018; **9**: 402.
- 80 Weber JA, Baxter DH, Zhang S, Huang DY, How Huang K, Jen Lee M *et al.* The MicroRNA Spectrum in 12 Body Fluids. *Clinical Chemistry* 2010; **56**: 1733–1741.
- 81 Lin S, Gregory RI. MicroRNA biogenesis pathways in cancer. *Nat Rev Cancer* 2015; **15**: 321–333.
- 82 Di Leva G, Croce CM. miRNA profiling of cancer. *Current Opinion in Genetics & Development* 2013; **23**: 3–11.
- 83 Arroyo JD, Chevillet JR, Kroh EM, Ruf IK, Pritchard CC, Gibson DF *et al.* Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. *PNAS* 2011; **108**: 5003–5008.
- 84 Endzeliņš E, Melne V, Kalniņa Z, Lietuvietis V, Riekstiņa U, Llorente A *et al.* Diagnostic, prognostic and predictive value of cell-free miRNAs in prostate cancer: a systematic review. *Molecular Cancer* 2016; **15**: 41.
- 85 Turchinovich A, Weiz L, Langheinz A, Burwinkel B. Characterization of extracellular circulating microRNA. *Nucleic Acids Research* 2011; **39**: 7223–7233.
- 86 Australian Prostate Cancer BioResource, Matin F, Jeet V, Moya L, Selth LA, Chambers S *et al.* A Plasma Biomarker Panel of Four MicroRNAs for the Diagnosis of Prostate Cancer. *Sci Rep* 2018; **8**: 6653.
- 87 Sita-Lumsden A, Dart DA, Waxman J, Bevan CL. Circulating microRNAs as potential new biomarkers for prostate cancer. *Br J Cancer* 2013; **108**: 1925–1930.
- 88 Bautista-Sánchez D, Arriaga-Canon C, Pedroza-Torres A, De La Rosa-Velázquez IA, González-Barrios R, Contreras-Espinosa L *et al.* The Promising Role of miR-21 as a Cancer Biomarker and Its Importance in RNA-Based Therapeutics. *Mol Ther Nucleic Acids* 2020; **20**: 409–420.
- 89 Yaman Agaoglu F, Kovancilar M, Dizdar Y, Darendeliler E, Holdenrieder S, Dalay N *et al.* Investigation of miR-21, miR-141, and miR-221 in blood circulation of patients with prostate cancer. *Tumor Biol* 2011; **32**: 583–588.
- 90 Endzeliņš E, Berger A, Melne V, Bajo-Santos C, Soboļevska K, Ābols A *et al.* Detection of circulating miRNAs: comparative analysis of extracellular vesicle-incorporated miRNAs and cell-free miRNAs in whole plasma of prostate cancer patients. *BMC Cancer* 2017; **17**: 730.
- 91 Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL *et al.* Circulating microRNAs as stable blood-based markers for cancer detection. *PNAS* 2008; **105**: 10513–10518.

- 92 Lin H-M, Castillo L, Mahon KL, Chiam K, Lee BY, Nguyen Q *et al.* Circulating microRNAs are associated with docetaxel chemotherapy outcome in castration-resistant prostate cancer. *British Journal of Cancer* 2014; **110**: 2462–2471.
- 93 Zhang W, Meng Y, Liu N, Wen X-F, Yang T. Insights into Chemoresistance of Prostate Cancer. *Int J Biol Sci* 2015; **11**: 1160–1170.
- 94 Kosaka N, Iguchi H, Ochiya T. Circulating microRNA in body fluid: a new potential biomarker for cancer diagnosis and prognosis. *Cancer Science* 2010; **101**: 2087–2092.
- 95 Rosenfeld N, Aharonov R, Meiri E, Rosenwald S, Spector Y, Zepeniuk M *et al.* MicroRNAs accurately identify cancer tissue origin. *Nat Biotechnol* 2008; **26**: 462–469.
- 96 Ignatiadis M, Sledge GW, Jeffrey SS. Liquid biopsy enters the clinic - implementation issues and future challenges. *Nat Rev Clin Oncol* 2021; **18**: 297–312.
- 97 Campos-Fernández E, Barcelos LS, de Souza AG, Goulart LR, Alonso-Goulart V. Research landscape of liquid biopsies in prostate cancer. *Am J Cancer Res* 2019; **9**: 1309–1328.
- 98 Kim C-J, Dong L, Amend SR, Cho Y-K, Pienta KJ. The role of liquid biopsies in prostate cancer management. *Lab Chip* 2021; **21**: 3263–3288.
- 99 Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A *et al.* Landscape of transcription in human cells. *Nature* 2012; **489**: 101–108.
- 100 Iyer MK, Niknafs YS, Malik R, Singhal U, Sahu A, Hosono Y *et al.* The landscape of long noncoding RNAs in the human transcriptome. *Nat Genet* 2015; **47**: 199–208.
- 101 Geisler S, Collier J. RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nat Rev Mol Cell Biol* 2013; **14**: 699–712.
- 102 Rinn JL, Chang HY. Genome Regulation by Long Noncoding RNAs. *Annual Review of Biochemistry* 2012; **81**: 145–166.
- 103 Ulitsky I, Bartel DP. lincRNAs: Genomics, Evolution, and Mechanisms. *Cell* 2013; **154**: 26–46.
- 104 Sun M, Kraus WL. From discovery to function: the expanding roles of long noncoding RNAs in physiology and disease. *Endocr Rev* 2015; **36**: 25–64.
- 105 Camacho CV, Choudhari R, Gadad SS. Long noncoding RNAs and cancer, an overview. *Steroids* 2018; **133**: 93–95.
- 106 Qiu M-T, Hu J-W, Yin R, Xu L. Long noncoding RNA: an emerging paradigm of cancer research. *Tumour Biol* 2013; **34**: 613–620.
- 107 Pennisi E. Cell biology. Lengthy RNAs earn respect as cellular players. *Science* 2014; **344**: 1072.
- 108 Ponting CP, Oliver PL, Reik W. Evolution and functions of long noncoding RNAs. *Cell* 2009; **136**: 629–641.
- 109 Arriaga-Canon C, De La Rosa-Velázquez IA, González-Barríos R, Montiel-Manríquez R, Oliva-Rico D, Jiménez-Trejo F *et al.* The use of long non-coding RNAs as prognostic biomarkers and therapeutic targets in prostate cancer. *Oncotarget* 2018; **9**: 20872–20890.

- 110 Bussemakers MJG, Bokhoven A van, Verhaegh GW, Smit FP, Karthaus HFM, Schalken JA *et al.* DD3::A New Prostate-specific Gene, Highly Overexpressed in Prostate Cancer. *Cancer Res* 1999; **59**: 5975–5979.
- 111 Durand X, Moutereau S, Xylinas E, de la Taille A. Progensa™ PCA3 test for prostate cancer. *Expert Rev Mol Diagn* 2011; **11**: 137–144.
- 112 Sartori DA, Chan DW. Biomarkers in prostate cancer: what's new? *Curr Opin Oncol* 2014; **26**: 259–264.
- 113 Haese A, de la Taille A, van Poppel H, Marberger M, Stenzl A, Mulders PFA *et al.* Clinical utility of the PCA3 urine assay in European men scheduled for repeat biopsy. *Eur Urol* 2008; **54**: 1081–1088.
- 114 Hessels D, Klein Gunnewiek JMT, van Oort I, Karthaus HFM, van Leenders GJL, van Balken B *et al.* DD3(PCA3)-based molecular urine analysis for the diagnosis of prostate cancer. *Eur Urol* 2003; **44**: 8–15; discussion 15-16.
- 115 Xue W-J, Ying X-L, Jiang J-H, Xu Y-H. Prostate cancer antigen 3 as a biomarker in the urine for prostate cancer diagnosis: a meta-analysis. *J Cancer Res Ther* 2014; **10 Suppl**: C218-221.
- 116 Prensner JR, Iyer MK, Sahu A, Asangani IA, Cao Q, Patel L *et al.* The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. *Nat Genet* 2013; **45**: 1392–1398.
- 117 Mehra R, Shi Y, Udager AM, Prensner JR, Sahu A, Iyer MK *et al.* A novel RNA in situ hybridization assay for the long noncoding RNA SChLAP1 predicts poor clinical outcome after radical prostatectomy in clinically localized prostate cancer. *Neoplasia* 2014; **16**: 1121–1127.
- 118 Prensner JR, Zhao S, Erho N, Schipper M, Iyer MK, Dhanasekaran SM *et al.* Nomination and validation of the long noncoding RNA SChLAP1 as a risk factor for metastatic prostate cancer progression: a multi-institutional high-throughput analysis. *Lancet Oncol* 2014; **15**: 1469–1480.
- 119 Luo J, Wang K, Yeh S, Sun Y, Liang L, Xiao Y *et al.* LncRNA-p21 alters the antiandrogen enzalutamide-induced prostate cancer neuroendocrine differentiation via modulating the EZH2/STAT3 signaling. *Nat Commun* 2019; **10**: 2571.
- 120 Prensner JR, Iyer MK, Balbin OA, Dhanasekaran SM, Cao Q, Brenner JC *et al.* Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. *Nat Biotechnol* 2011; **29**: 742–749.
- 121 Xiong T, Li J, Chen F, Zhang F. PCAT-1: A Novel Oncogenic Long Non-Coding RNA in Human Cancers. *Int J Biol Sci* 2019; **15**: 847–856.
- 122 Prensner JR, Chen W, Iyer MK, Cao Q, Ma T, Han S *et al.* PCAT-1, a long noncoding RNA, regulates BRCA2 and controls homologous recombination in cancer. *Cancer Res* 2014; **74**: 1651–1660.
- 123 Morganti S, Tarantino P, Ferraro E, D'Amico P, Duso BA, Curigliano G. Next Generation Sequencing (NGS): A Revolutionary Technology in Pharmacogenomics and Personalized Medicine in Cancer. *Adv Exp Med Biol* 2019; **1168**: 9–30.

- 124 Pine AC, Fioretti FF, Brooke GN, Bevan CL. Advances in genetics: widening our understanding of prostate cancer. *F1000Res* 2016; **5**. doi:10.12688/f1000research.8019.1.
- 125 Beltran H, Eng K, Mosquera JM, Sigaras A, Romanel A, Rennert H *et al*. Whole-Exome Sequencing of Metastatic Cancer and Biomarkers of Treatment Response. *JAMA Oncol* 2015; **1**: 466–474.
- 126 Xuan J, Yu Y, Qing T, Guo L, Shi L. Next-generation sequencing in the clinic: promises and challenges. *Cancer Lett* 2013; **340**: 284–295.
- 127 Byron SA, Van Keuren-Jensen KR, Engelthaler DM, Carpten JD, Craig DW. Translating RNA sequencing into clinical diagnostics: opportunities and challenges. *Nature Reviews Genetics* 2016; **17**: 257–271.
- 128 Berglund E, Maaskola J, Schultz N, Friedrich S, Marklund M, Bergenstråhle J *et al*. Spatial maps of prostate cancer transcriptomes reveal an unexplored landscape of heterogeneity. *Nat Commun* 2018; **9**: 2419.
- 129 Robinson D, Van Allen EM, Wu Y-M, Schultz N, Lonigro RJ, Mosquera J-M *et al*. Integrative clinical genomics of advanced prostate cancer. *Cell* 2015; **161**: 1215–1228.
- 130 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**: 645–651.
- 131 Abeshouse A, Ahn J, Akbani R, Ally A, Amin S, Andry CD *et al*. The Molecular Taxonomy of Primary Prostate Cancer. *Cell* 2015; **163**: 1011–1025.
- 132 Deo RC. Machine Learning in Medicine. *Circulation* 2015; **132**: 1920–1930.
- 133 Bishop CM. Pattern recognition and machine learning. CERN Document Server. 2006. <https://cds.cern.ch/record/998831> (accessed 13 Aug2020).
- 134 Goldenberg SL, Nir G, Salcudean SE. A new era: artificial intelligence and machine learning in prostate cancer. *Nature Reviews Urology* 2019; **16**: 391–403.
- 135 Shilo S, Rossman H, Segal E. Axes of a revolution: challenges and promises of big data in healthcare. *Nature Medicine* 2020; **26**: 29–38.
- 136 Goecks J, Jalili V, Heiser LM, Gray JW. How Machine Learning Will Transform Biomedicine. *Cell* 2020; **181**: 92–101.
- 137 Litjens G, Toth R, van de Ven W, Hoeks C, Kerkstra S, van Ginneken B *et al*. Evaluation of prostate segmentation algorithms for MRI: the PROMISE12 challenge. *Med Image Anal* 2014; **18**: 359–373.
- 138 Libbrecht MW, Noble WS. Machine learning applications in genetics and genomics. *Nat Rev Genet* 2015; **16**: 321–332.
- 139 Karnes RJ, Bergstralh EJ, Davicioni E, Ghadessi M, Buerki C, Mitra AP *et al*. Validation of a genomic classifier that predicts metastasis following radical prostatectomy in an at risk patient population. *J Urol* 2013; **190**: 2047–2053.
- 140 Bera K, Schalper KA, Rimm DL, Velcheti V, Madabhushi A. Artificial intelligence in digital pathology - new tools for diagnosis and precision oncology. *Nat Rev Clin Oncol* 2019; **16**: 703–715.

- 141 Tolkach Y, Dohmgörge T, Toma M, Kristiansen G. High-accuracy prostate cancer pathology using deep learning. *Nature Machine Intelligence* 2020; **2**: 411–418.
- 142 Fujita K, Nonomura N. Urinary biomarkers of prostate cancer. *International Journal of Urology* 2018; **25**: 770–779.
- 143 Dong L, Huang C-Y, Johnson EJ, Yang L, Zieren RC, Horie K *et al.* High-Throughput Simultaneous mRNA Profiling Using nCounter Technology Demonstrates That Extracellular Vesicles Contain Different mRNA Transcripts Than Their Parental Prostate Cancer Cells. *Anal Chem* 2021; **93**: 3717–3725.
- 144 McKiernan J, Donovan MJ, O'Neill V, Bentink S, Noerholm M, Belzer S *et al.* A Novel Urine Exosome Gene Expression Assay to Predict High-grade Prostate Cancer at Initial Biopsy. *JAMA Oncol* 2016; **2**: 882–889.
- 145 Stark R, Grzelak M, Hadfield J. RNA sequencing: the teenage years. *Nature Reviews Genetics* 2019; **20**: 631–656.
- 146 Suzuki Y. Advent of a new sequencing era: long-read and on-site sequencing. *J Hum Genet* 2020; **65**: 1–1.
- 147 Handa S, Hans B, Goel S, Bashorun HO, Dovey Z, Tewari A. Immunotherapy in prostate cancer: current state and future perspectives. *Ther Adv Urol* 2020; **12**: 1756287220951404.
- 148 Pineda-Jaramillo JD. A review of Machine Learning (ML) algorithms used for modeling travel mode choice. http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0012-73532019000400032 (accessed 18 Aug2021).
- 149 Scher HI, Beer TM, Higano CS, Anand A, Taplin M-E, Efstathiou E *et al.* Antitumour activity of MDV3100 in castration-resistant prostate cancer: a phase 1-2 study. *Lancet* 2010; **375**: 1437–1446.
- 150 Tosco L, Laenen A, Gevaert T, Salmon I, Decaestecker C, Davicioni E *et al.* Neoadjuvant degarelix with or without apalutamide followed by radical prostatectomy for intermediate and high-risk prostate cancer: ARNEO, a randomized, double blind, placebo-controlled trial. *BMC Cancer* 2018; **18**: 354.
- 151 Sternberg CN. Enzalutamide, an oral androgen receptor inhibitor for treatment of castration-resistant prostate cancer. *Future Oncol* 2019; **15**: 1437–1457.
- 152 Nickols NG, Nazarian R, Zhao SG, Tan V, Uzunangelov V, Xia Z *et al.* MEK-ERK signaling is a therapeutic target in metastatic castration resistant prostate cancer. *Prostate Cancer Prostatic Dis* 2019; **22**: 531–538.
- 153 Liu Y, Karaca M, Zhang Z, Gioeli D, Earp HS, Whang YE. Dasatinib inhibits site-specific tyrosine phosphorylation of androgen receptor by Ack1 and Src kinases. *Oncogene* 2010; **29**: 3208–3216.
- 154 Rathkopf DE, Smith MR, Ryan CJ, Berry WR, Shore ND, Liu G *et al.* Androgen receptor mutations in patients with castration-resistant prostate cancer treated with apalutamide. *Ann Oncol* 2017; **28**: 2264–2271.

- 155 Karantanos T, Evans CP, Tombal B, Thompson TC, Montironi R, Isaacs WB. Understanding the mechanisms of androgen deprivation resistance in prostate cancer at the molecular level. *Eur Urol* 2015; **67**: 470–479.
- 156 Harris WP, Mostaghel EA, Nelson PS, Montgomery B. Androgen deprivation therapy: progress in understanding mechanisms of resistance and optimizing androgen depletion. *Nat Clin Pract Urol* 2009; **6**: 76–85.
- 157 Conteduca V, Jayaram A, Romero-Laorden N, Wetterskog D, Salvi S, Gurioli G *et al.* Plasma Androgen Receptor and Docetaxel for Metastatic Castration-resistant Prostate Cancer. *Eur Urol* 2019; **75**: 368–373.
- 158 Harmon SA, Perk T, Lin C, Eickhoff J, Choyke PL, Dahut WL *et al.* Quantitative Assessment of Early [18F]Sodium Fluoride Positron Emission Tomography/Computed Tomography Response to Treatment in Men With Metastatic Prostate Cancer to Bone. *J Clin Oncol* 2017; **35**: 2829–2837.
- 159 Olson BM, Johnson LE, McNeel DG. The androgen receptor: a biologically relevant vaccine target for the treatment of prostate cancer. *Cancer Immunol Immunother* 2013; **62**: 585–596.
- 160 de Wit R, de Bono J, Sternberg CN, Fizazi K, Tombal B, Wülfing C *et al.* Cabazitaxel versus Abiraterone or Enzalutamide in Metastatic Prostate Cancer. *N Engl J Med* 2019; **381**: 2506–2518.
- 161 Fizazi K, Tran N, Fein L, Matsubara N, Rodriguez-Antolin A, Alekseev BY *et al.* Abiraterone plus Prednisone in Metastatic, Castration-Sensitive Prostate Cancer. *N Engl J Med* 2017; **377**: 352–360.
- 162 Nj C, J W, Jd J, C T, S O, A D *et al.* ARN-509: a novel antiandrogen for prostate cancer treatment. *Cancer research* 2012; **72**. doi:10.1158/0008-5472.CAN-11-3948.
- 163 Kittai AS, Blank J, Graff JN. Gonadotropin-Releasing Hormone Antagonists in Prostate Cancer. *Oncology (Williston Park)* 2018; **32**: 599–602, 604–606.
- 164 Cucchiara V, Yang JC, Liu C, Adomat HH, Tomlinson Guns ES, Gleave ME *et al.* GnRH Antagonists Have Direct Inhibitory Effects On Castration-Resistant Prostate Cancer Via Intracrine Androgen and AR-V7 Expression. *Mol Cancer Ther* 2019; **18**: 1811–1821.
- 165 Pomerantz MM, Spisák S, Jia L, Cronin AM, Csabai I, Ledet E *et al.* The association between germline BRCA2 variants and sensitivity to platinum-based chemotherapy among men with metastatic prostate cancer. *Cancer* 2017; **123**: 3532–3539.
- 166 Annala M, Vandekerckhove G, Khalaf D, Taavitsainen S, Beja K, Warner EW *et al.* Circulating Tumor DNA Genomics Correlate with Resistance to Abiraterone and Enzalutamide in Prostate Cancer. *Cancer Discov* 2018; **8**: 444–457.
- 167 Antonarakis ES, Lu C, Luber B, Liang C, Wang H, Chen Y *et al.* Germline DNA-repair Gene Mutations and Outcomes in Men with Metastatic Castration-resistant Prostate Cancer Receiving First-line Abiraterone and Enzalutamide. *Eur Urol* 2018; **74**: 218–225.
- 168 Courtney KD, Manola JB, Elfiky AA, Ross R, Oh WK, Yap JT *et al.* A phase I study of everolimus and docetaxel in patients with castration-resistant prostate cancer. *Clin Genitourin Cancer* 2015; **13**: 113–123.

- 169 George DJ, Halabi S, Healy P, Jonasch D, Anand M, Rasmussen J *et al.* Phase 2 clinical trial of TORC1 inhibition with everolimus in men with metastatic castration-resistant prostate cancer. *Urol Oncol* 2020; **38**: 79.e15-79.e22.
- 170 Ham WS, Cho NH, Kim WT, Ju HJ, Lee JS, Choi YD. Pathological effects of prostate cancer correlate with neuroendocrine differentiation and PTEN expression after bicalutamide monotherapy. *J Urol* 2009; **182**: 1378–1384.
- 171 Rescigno P, Lorente D, Dolling D, Ferraldeschi R, Rodrigues DN, Riisnaes R *et al.* Docetaxel Treatment in PTEN- and ERG-aberrant Metastatic Prostate Cancers. *Eur Urol Oncol* 2018; **1**: 71–77.
- 172 Tee SS, Suster I, Truong S, Jeong S, Eskandari R, DiGialleonardo V *et al.* Targeted AKT Inhibition in Prostate Cancer Cells and Spheroids Reduces Aerobic Glycolysis and Generation of Hyperpolarized [1-13C] Lactate. *Mol Cancer Res* 2018; **16**: 453–460.
- 173 De Velasco MA, Kura Y, Yoshikawa K, Nishio K, Davies BR, Uemura H. Efficacy of targeted AKT inhibition in genetically engineered mouse models of PTEN-deficient prostate cancer. *Oncotarget* 2016; **7**: 15959–15976.
- 174 Lamoureux F, Thomas C, Crafter C, Kumano M, Zhang F, Davies BR *et al.* Blocked autophagy using lysosomotropic agents sensitizes resistant prostate tumor cells to the novel Akt inhibitor AZD5363. *Clin Cancer Res* 2013; **19**: 833–844.
- 175 Emmenegger U, Booth CM, Berry S, Sridhar SS, Winquist E, Bandali N *et al.* Temsirolimus Maintenance Therapy After Docetaxel Induction in Castration-Resistant Prostate Cancer. *Oncologist* 2015; **20**: 1351–1352.
- 176 Barata PC, Cooney M, Mendiratta P, Gupta R, Dreicer R, Garcia JA. Phase I/II study evaluating the safety and clinical efficacy of temsirolimus and bevacizumab in patients with chemotherapy refractory metastatic castration-resistant prostate cancer. *Invest New Drugs* 2019; **37**: 331–337.
- 177 Armstrong AJ, Shen T, Halabi S, Kemeny G, Bitting RL, Kartcheske P *et al.* A phase II trial of temsirolimus in men with castration-resistant metastatic prostate cancer. *Clin Genitourin Cancer* 2013; **11**: 397–406.
- 178 Pantel K, Hille C, Scher HI. Circulating Tumor Cells in Prostate Cancer: From Discovery to Clinical Utility. *Clin Chem* 2019; **65**: 87–99.
- 179 Pandit-Taskar N, O'Donoghue JA, Beylertgil V, Lyashchenko S, Ruan S, Solomon SB *et al.* ⁸⁹Zr-huJ591 immuno-PET imaging in patients with advanced metastatic prostate cancer. *Eur J Nucl Med Mol Imaging* 2014; **41**: 2093–2105.
- 180 Sieuwerts AM, Onstenk W, Kraan J, Beaufort CM, Van M, De Laere B *et al.* AR splice variants in circulating tumor cells of patients with castration-resistant prostate cancer: relation with outcome to cabazitaxel. *Mol Oncol* 2019; **13**: 1795–1807.
- 181 Torquato S, Pallavajjala A, Goldstein A, Toro PV, Silberstein JL, Lee J *et al.* Genetic Alterations Detected in Cell-Free DNA Are Associated With Enzalutamide and Abiraterone Resistance in Castration-Resistant Prostate Cancer. *JCO Precis Oncol* 2019; **3**. doi:10.1200/PO.18.00227.
- 182 Gordevičius J, Kriščiūnas A, Groot DE, Yip SM, Susic M, Kwan A *et al.* Cell-Free DNA Modification Dynamics in Abiraterone Acetate-Treated Prostate Cancer Patients. *Clin Cancer Res* 2018; **24**: 3317–3324.

- 183 Wyatt AW, Azad AA, Volik SV, Annala M, Beja K, McConeghy B *et al.* Genomic Alterations in Cell-Free DNA and Enzalutamide Resistance in Castration-Resistant Prostate Cancer. *JAMA Oncol* 2016; **2**: 1598–1606.
- 184 Wang Y, Lieberman R, Pan J, Zhang Q, Du M, Zhang P *et al.* miR-375 induces docetaxel resistance in prostate cancer by targeting SEC23A and YAP1. *Mol Cancer* 2016; **15**: 70.
- 185 Dang Q, Li L, Xie H, He D, Chen J, Song W *et al.* Anti-androgen enzalutamide enhances prostate cancer neuroendocrine (NE) differentiation via altering the infiltrated mast cells → androgen receptor (AR) → miRNA32 signals. *Mol Oncol* 2015; **9**: 1241–1251.
- 186 Kumar B, Khaleghzadegan S, Mears B, Hatano K, Kudrolli TA, Chowdhury WH *et al.* Identification of miR-30b-3p and miR-30d-5p as direct regulators of androgen receptor signaling in prostate cancer by complementary functional microRNA library screening. *Oncotarget* 2016; **7**: 72593–72607.
- 187 Guan Y, Guan X, An H, Baihetiya A, Wang W, Shao W *et al.* Epigenetic silencing of miR-137 induces resistance to bicalutamide by targeting TRIM24 in prostate cancer cells. *Am J Transl Res* 2019; **11**: 3226–3237.
- 188 Lehmusvaara S, Erkkilä T, Urbanucci A, Jalava S, Seppälä J, Kaipia A *et al.* Goserelin and bicalutamide treatments alter the expression of microRNAs in the prostate. *Prostate* 2013; **73**: 101–112.
- 189 Luo J, Wang K, Yeh S, Sun Y, Liang L, Xiao Y *et al.* LncRNA-p21 alters the antiandrogen enzalutamide-induced prostate cancer neuroendocrine differentiation via modulating the EZH2/STAT3 signaling. *Nat Commun* 2019; **10**: 2571.