



**Universidad de la República, Facultad de Ciencias**

TESINA PARA OPTAR POR EL GRADO DE LICENCIADO EN CIENCIAS BIOLÓGICAS

# Desarrollo de una línea tumoral prostática *knock-out* para el gen *CD147* utilizando el sistema CRISPR-Cas9

*Rodrigo José Martino Kunsch*

Orientador: Dra. Lucía Pastro

Co-Orientador: Dra. Mercedes Rodríguez-Teja

Laboratorio de Interacción célula-ambiente, Departamento de  
Genética, Facultad de Medicina

Montevideo, Uruguay

2022

**Página de aprobación**

Universidad de la República, Facultad de Ciencias.

**Título:**

Desarrollo de una línea tumoral prostática knock-out para el gen CD147 utilizando el sistema CRISPR-Cas9

**Autor:**

Rodrigo José Martino Kunsch

**Orientador:**

Dra. Lucía Pastro

**Co-Orientador:**

Dra. Mercedes Rodríguez-Teja

**Carrera:**

Licenciatura en Ciencias Biológicas

**Calificación: 11****Tribunal:**

Dra. Cecilia Mathó

Lic. Lucía Brignoni

Dra. Lucía Pastro

Dra. Mercedes Rodríguez-Teja

**Fecha: 15 de junio, 2022**

## Agradecimientos

A Checha y Lucía por todo lo que me enseñaron, por su disposición y paciencia para ayudarme siempre.

A mi esposa, Victoria, que me acompañó y me apoyó en este proceso desde el comienzo.

A mi padre que siempre me apoyó y me impulsó adelante y a cuya memoria dedico este trabajo.

---

## Tabla de contenido

Agradecimientos.....	iii
Tabla de contenido.....	iv
Índice de figuras.....	vi
Índice de tablas.....	viii
Resumen.....	ix
Abreviaturas.....	x
1. Introducción.....	1
1.1. Cáncer de próstata.....	1
1.2. Metástasis y metaloproteinasas.....	2
1.3. CD147.....	6
1.4. Edición génica.....	12
1.5. Historia de CRISPR-Cas9.....	16
2. Hipótesis y objetivos.....	20
2.1. Hipótesis.....	20
2.2. Objetivo general.....	20
2.3. Objetivos específicos.....	20
3. Materiales y métodos.....	21
3.1. Análisis de datos de expresión génica en bases de datos.....	21
3.2. Obtención de secuencias.....	22
3.3. Diseño de cebadores.....	23
3.3.1. Evaluación de cebadores.....	23
3.4. Diseño de ARN guías.....	24
3.5. Electroforesis en gel de agarosa.....	25
3.6. Cuantificación de ácidos nucleicos.....	26
3.7. Generación del vector de edición.....	27

---

3.7.1.	Clonado de ARN guía al vector lentiviral .....	27
3.7.2.	Transformación de bacterias Stbl3 .....	28
3.7.3.	Obtención de ADN plasmídico.....	29
3.8.	Generación de la línea celular <i>knock-out</i> .....	31
3.8.1.	Cultivo de células PC3.....	32
3.8.2.	Transfección de células PC3.....	32
3.8.3.	Selección de clones.....	33
3.9.	Verificación de la línea celular <i>knock-out</i> .....	34
3.9.1.	Obtención de muestras celulares para análisis .....	34
3.9.2.	Análisis de expresión de proteína CD147.....	35
3.9.3.	Secuenciación .....	37
3.9.4.	Análisis de expresión de ARNm .....	39
4.	Resultados y discusión .....	43
4.1.	Análisis de datos de expresión génica en bases de datos .....	43
4.2.	Diseño de ARN guías para la edición del gen <i>CD147</i> .....	50
4.3.	Diseño de cebadores .....	54
4.4.	Clonado de ARNs guías en el vector lentiCRISPR v2.....	58
4.5.	Generación de una línea celular de cáncer de próstata metastásica <i>CD147</i> negativa .....	62
5.	Conclusiones y perspectivas. ....	78
6.	Bibliografía.....	81

## Índice de figuras

Figura 1: Incidencia y mortalidad de cáncer en los principales sitios en hombres. Periodo 2013–2017.....	1
Figura 2: Cascada invasión-metástasis.....	3
Figura 3: Representación esquemática de las isoformas de CD147.....	7
Figura 4: Inactivación de un gen mediante el sistema de edición génica CRISPR-Cas9.....	15
Figura 5: Esquema del vector lentiviral lentiCRISPR v2.....	26
Figura 6: Expresión de <i>CD147</i> en tejido normal, tumores primarios y metástasis. ...	46
Figura 7: Curva de supervivencia Kaplan Meier para expresión de <i>CD147</i> en cáncer de próstata.....	50
Figura 8: Secuencias de referencia para las variantes de ARN mensajero de <i>CD147</i> . .....	52
Figura 9: Sitios de unión de cebadores diseñados para amplificar el blanco genómico de los sgRNA. ....	54
Figura 10: Sitios de unión de cebadores diseñados para amplificar ARNm de <i>CD147</i> . .....	55
Figura 11: Puesta a punto de PCR para la amplificación de los exones 4 y 5 de <i>CD147</i> . .....	57
Figura 12: Puesta a punto de PCR para el par de cebadores TTF2_F y TTF2_R ...	58
Figura 13: Presencia del vector lentiCRISPR v2 en cultivos de bacterias transformadas .....	60
Figura 14: Verificación de la integridad de los plásmidos purificados. ....	62
Figura 15: Detección de proteínas totales transferidas a membrana de nitrocelulosa. ....	64
Figura 16: Expresión proteica de CD147 en líneas celulares generadas.....	65
Figura 17: Integridad del ARN extraído de las líneas celulares generadas.....	67
Figura 18: Curvas de calibración de los cebadores diseñados para qPCR.....	68
Figura 19: Cuantificación relativa de expresión de <i>CD147</i> en líneas celulares.....	70
Figura 20: Verificación de amplificación por PCR de regiones de interés para secuenciación. ....	72
Figura 21: Cromatogramas de la secuencia del exón 4 de <i>CD147</i> de la línea E4-3.	74

Figura 22: Resultados de CRISP-ID para las secuencias del exón 4 de *CD147* en la línea celular editada E4-3..... 75

Figura 23: Secuencias del exón 4 de *CD147* en las líneas E4-3 y vector\_V2..... 77

---

## Índice de tablas

Tabla 1: Clasificación de la familia de las metaloproteinasas (MMP) en humanos. ...	4
Tabla 2: Estudios realizados sobre expresión de CD147 en cáncer de próstata .....	12
Tabla 3: Conteo de muestras con datos de expresión de <i>CD147</i> contenidas en los proyectos GTEx y TCGA.....	44
Tabla 4: Expresión de <i>CD147</i> en muestras de tejido normal no tumoral de los proyectos TCGA y GTEx.....	47
Tabla 5: ARN guías diseñados para la edición genómica del gen <i>CD147</i> utilizando un sistema CRISPR-Cas9.....	52
Tabla 6: Genes potencialmente <i>off-target</i> de los ARN guía diseñados.....	53
Tabla 7: Cebadores diseñados. ....	56
Tabla 8: Concentración y pureza del ADN plasmídico extraído de cultivos bacterianos transformados con el vector lentiCRISPR v2 .....	59
Tabla 9: Concentración y pureza de los vectores lentiCRISPR v2 purificados. ....	61
Tabla 10: Concentración de proteína extraída de las líneas celulares generadas ...	63
Tabla 11: Cuantificación de ARN purificado a partir de extractos celulares. ....	66
Tabla 12: Eficiencia de cebadores diseñados para qPCR .....	69
Tabla 13: Amplificación de <i>CD147</i> y <i>ACTB</i> por qPCR .....	70



## Resumen

El cáncer de próstata es uno de los tipos de cáncer con mayor incidencia y mortalidad a nivel mundial. La muerte en esta enfermedad es causada generalmente por la progresión a la metástasis y la formación de nuevos tumores en pulmón y huesos. La progresión hacia la metástasis está modulada en parte por la acción de proteasas de la familia de las metaloproteinasas de matriz. En el ambiente tumoral la acción de las proteinasas es modulada por la proteína CD147 mediante un mecanismo de acción poco conocido. CD147 se encuentra sobre expresado en la mayoría de los tipos de cáncer y su expresión se asocia a un mal pronóstico de la enfermedad, aunque existe discusión sobre el rol de CD147 en cáncer de próstata metastásico. En este trabajo se utilizaron bases de datos disponibles *online* para analizar la relación entre la expresión de CD147 y el estado de cáncer. Se demostró que esta proteína se encuentra sobreexpresada en cáncer y más aún en cáncer metastásico; y se demostró que CD147 se encuentra sobreexpresado en tumores sólidos de cáncer de próstata. Con el objetivo de estudiar el rol de CD147 en cáncer de próstata metastásico se generó una línea celular de cáncer de próstata metastásico CD147<sup>-/-</sup> mediante la utilización de un sistema de edición génica CRISPR-Cas9. El estudio de esta línea celular permitirá elucidar el rol de CD147 como modulador de MMP y su papel en la progresión de la metástasis en cáncer de próstata.

**Palabras clave:** cáncer; cáncer de próstata; metástasis; CD147; EMMPRIN; MMP; metaloproteinasas de matriz; CRISPR-Cas9;

## Abreviaturas

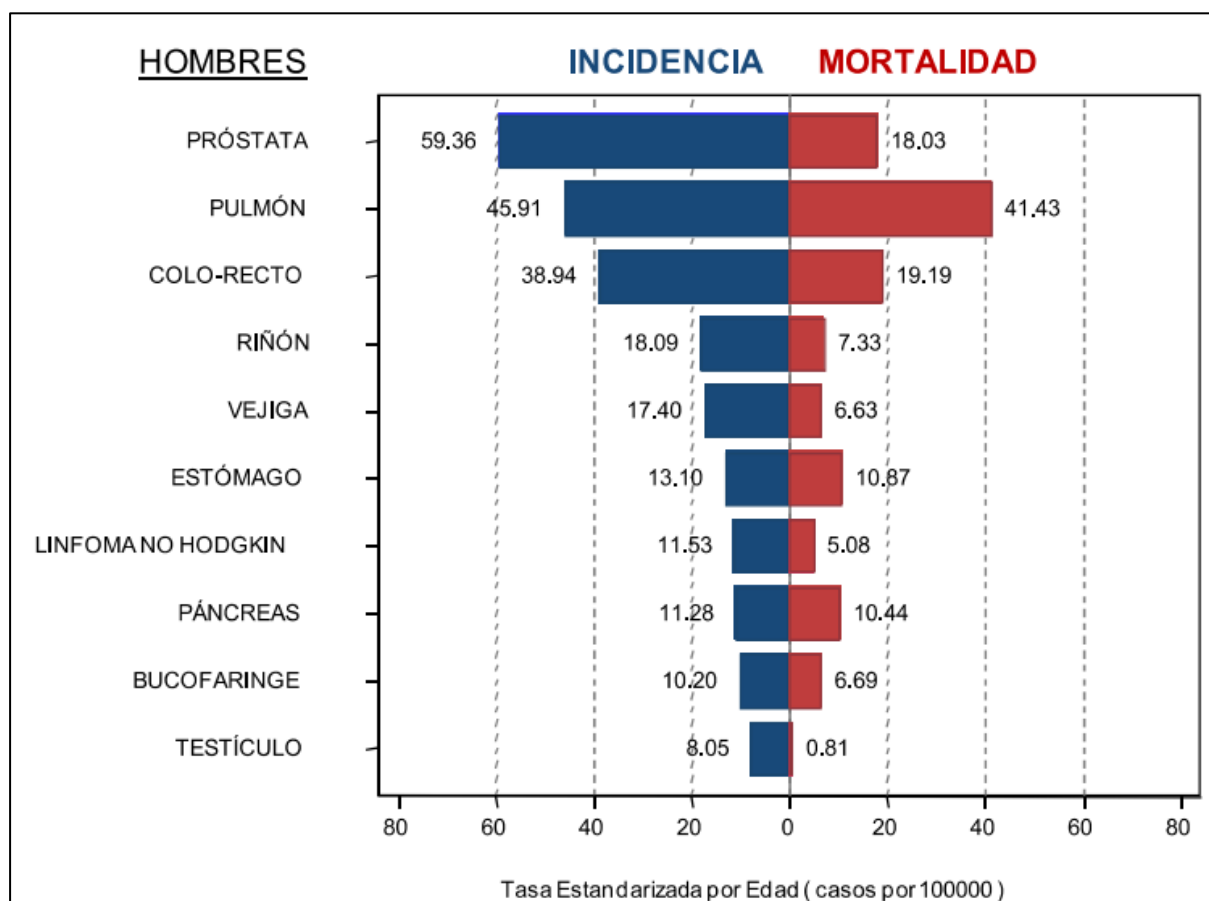
A	
ADN	Ácido desoxirribonucleico
ADNc	Ácido desoxirribonucleico copia
AJCC	<i>American Joint Committee on Cancer</i>
APS	Persulfato amónico
ARN	Ácido ribonucleico
ARNm	ARN mensajero
B	
BSG	Basigin
C	
Cas	<i>CRISPR associated</i>
CD147	<i>Cluster of differentiation 147</i>
CIP	Fosfatasa alcalina de intestino de ternero
Cq	Ciclos de umbral
CRISPR	<i>Clustered regularly interspaced short palindromic repeats</i>
crRNA	ARN de CRISPR
CypA	Ciclofilina A
D	
DMSO	Dimetilsulfóxido
DNasa	Desoxirribonucleasa
DTT	Ditiotreitol
E	
EDTA	Ácido etilendiaminotetraacético
EMMPRIN	<i>Extracellular matrix metalloproteinase inducer</i>
F	
FAK	Quinasas de adhesiones focales
FBS	Suero Fetal Bovino
G	
GTex	<i>Genotype-Tissue Expression</i>
H	
HDR	<i>Homology Directed Repair</i>
I	
IGF	Factor de crecimiento insulínico
IGFBP	Proteína de unión al factor de crecimiento insulínico
IGFR1	Receptor 1 de IGF
L	
LB	Luria–Bertani
M	
MCT	Transportador de monocarboxilatos
MEC	Membrana extracelular
MMP	Metaloproteinasa de matriz
N	
NaAc	Acetato de sodio

NHEJ	<i>Non-homologous End Joining</i>
NMD	<i>Nonsense-mediated decay</i>
P	
PAM	<i>Protospacer adjacent motif</i>
PBS	<i>Buffer</i> fosfato salino
PCR	Reacción en cadena de la polimerasa
PNK	Quinasa de polinucleótidos
PSA	Antígeno prostático específico
Q	
R	
RIPA	<i>Radioimmunoprecipitation assay</i>
RNasa	Ribonucleasa
RPMI	<i>Roswell Park Memorial Institute</i>
RQ	Cuantificación relativa
RT-qPCR	Reacción en cadena de la polimerasa cuantitativa, con transcripción reversa
S	
SDS	Dodecilsulfato sódico
SDS-PAGE	Electroforesis en geles desnaturalizantes de poliacrilamida y SDS
sgRNA	Ácido ribonucleico guía
T	
TAE	Tris-acetato
TALE	<i>Transcription Activator Like Effector</i>
TALEN	Nucleasa de tipo TALE
TARGET	<i>Therapeutically Applicable Research To Generate Effective Treatments</i>
TCGA	<i>The Cancer Genome Atlas</i>
TE	Tris-EDTA
TEM	Transición epitelial-mesenquimal
TEMED	Tetrametiletildiamina
TGF $\beta$	Factor de crecimiento transformante beta
Tm	Temperatura de <i>melting</i>
tracrRNA	ARN de crispr activador en trans
V	
VEGF	Factor de crecimiento endotelial vascular
VEGFR	Receptor de VEGF
Z	
ZFN	<i>Zinc Finger Nucleases</i>

## 1. Introducción

### 1.1. Cáncer de próstata

El cáncer de próstata es una enfermedad que se caracteriza por la división desregulada de células de la glándula prostática, provocando un crecimiento anormal de la misma. Entre los factores de riesgo asociados a esta enfermedad se encuentran la edad, ascendencia africana, historia familiar de cáncer de próstata y predisposición genética <sup>1</sup>. Es el segundo cáncer de mayor incidencia en hombres y la quinta causa más frecuente de muerte <sup>2</sup>. Globalmente, los hombres tienen una probabilidad de 1 en 18 de desarrollar cáncer de próstata en el transcurso de su vida <sup>3</sup>. En Uruguay, el cáncer de próstata es el cáncer de mayor incidencia y la tercera causa más frecuente de muerte por cáncer en hombres <sup>4</sup> (Figura 1).



**Figura 1: Incidencia y mortalidad de cáncer en los principales sitios en hombres. Periodo 2013–2017.**

Tomado del informe 2021 de la Comisión Honoraria de Lucha contra el Cáncer sobre la situación epidemiológica del Uruguay en relación al cáncer <sup>4</sup>.

---

La mayoría de los hombres afectados por cáncer de próstata presentan una forma de la enfermedad de crecimiento lento, localizado, y generalmente no letal. La muerte por cáncer de próstata se produce principalmente por metástasis, cuando las células del cáncer se diseminan a otras zonas del cuerpo, principalmente a los ganglios linfáticos, los huesos y el cerebro <sup>5</sup>.

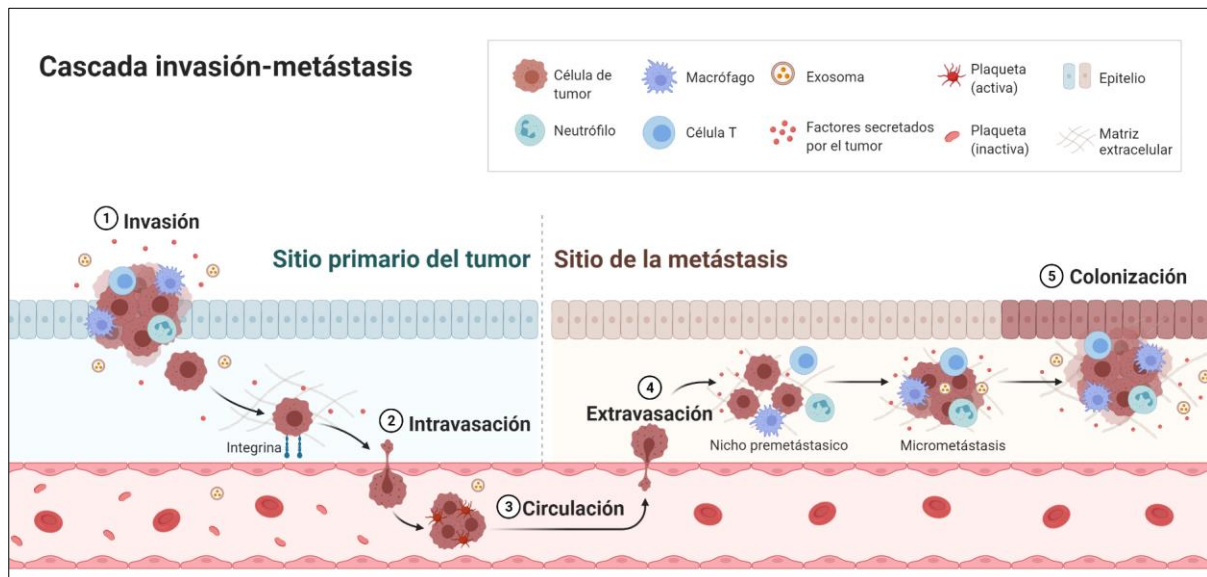
La progresión desde un tumor sólido a una metástasis es gradual. Una de las formas más usadas para clasificar el grado de avance del cáncer de próstata es el sistema de estadios TNM de la *American Joint Committee on Cancer* (AJCC). Este sistema clasifica el cáncer en estadios desde I (menos avanzado) a IV (más avanzado) en base a cinco factores: el nivel de antígeno prostático específico (PSA) al momento de diagnóstico, el puntaje de Gleason, y la clasificación del tumor en las categorías T (tumor), N (nodo) y M (metástasis) <sup>6</sup>.

El PSA es una proteína producida por la próstata que normalmente se encuentra en bajas cantidades en la sangre. Los niveles elevados de PSA en sangre se correlacionan con la presencia y agresividad del cáncer de próstata. El puntaje de Gleason es una metodología que consiste en asignar un puntaje a la medida en que las células cancerosas se parecen a las células prostáticas sanas <sup>7</sup>. Se basa en que existe una correlación entre la agresividad del tumor y la cantidad de cambios observados en el fenotipo celular. Por último, la clasificación TNM clasifica el cáncer en base a el tamaño y extensión del tumor primario (categoría T), la presencia de cáncer en ganglios linfáticos cercanos (categoría N) y la presencia de cáncer en sitios distantes (categoría M).

## 1.2. Metástasis y metaloproteinasas

El proceso de metástasis involucra una compleja serie de pasos biológicos que permiten a la célula cancerosa escapar del tumor primario sólido para formar un nuevo tumor en un sitio distante (Figura 2). Este proceso, denominado la cascada invasión-metástasis, comienza cuando el tumor sólido atraviesa la membrana basal, invadiendo el tejido circundante. Luego de esta invasión local al estroma, las células tumorales invaden en los vasos sanguíneos o linfáticos (intravasación). Luego, las células tumorales pierden sus adhesiones célula-célula y célula-matriz, separándose de la masa tumoral. Las células son transportadas por el sistema circulatorio y, si

sobreviven en la circulación, eventualmente llegan a un nuevo sitio. Por último, las células escapan de los vasos sanguíneos (extravasación) e invaden el tejido circundante. Si las células se adaptan al nuevo ambiente forman nuevas adhesiones que les permite proliferar, formando una metástasis <sup>8</sup>.



**Figura 2: Cascada invasión-metástasis**

En este esquema de la cascada invasión-metástasis se muestran las principales etapas del proceso que llevan a la formación de un nuevo tumor metastásico. El proceso comienza cuando el tumor atraviesa la membrana basal del epitelio e ingresa al estroma (invasión). Células del tumor ingresan a los vasos sanguíneos (intravasación), y son transportados a un segundo sitio (circulación) donde egresan de los vasos (extravasación). Si el ambiente en el nuevo sitio es favorable para la célula, comienza la formación de un nuevo tumor (colonización). Imagen adaptada de "Overview of Metastatic Cascade", por BioRender.com (2021). Obtenido de <https://app.biorender.com/biorender-templates>.

El proceso de metástasis es un proceso ineficiente, heterogéneo y adaptativo, que requiere de la plasticidad celular. Durante este proceso las células tumorales modifican su comportamiento y morfología, perdiendo el fenotipo epitelial y adquiriendo un fenotipo mesenquimal capaz de migrar e invadir otros tejidos. Esta transición fenotípica epitelial-mesenquimal requiere de cambios en los patrones de adhesión célula-célula y célula-matriz extracelular; y de la remodelación de la matriz extracelular (MEC) <sup>8</sup>.

La remodelación de la MEC depende de la acción de proteasas. Una familia de proteasas importantes para esta función son las metaloproteinasas de matriz (MMP);

enzimas proteolíticas capaces de degradar varios de los componentes de la matriz extracelular como el colágeno, elastina, laminina, fibronectina y proteoglicanos. Esta familia está compuesta por 28 enzimas distintas de las cuales se han identificado 23 homólogos en humanos (Tabla 1). Se clasifican generalmente en 6 subfamilias en base a su sustrato y a su estructura <sup>9,10</sup>.

Subfamilia	Nombre	Clasificación numérica
Colagenasas	Colagenasa-1	MMP-1
	Colagenasa-2	MMP-8
	Colagenasa-3	MMP-13
Gelatinasas	Gelatinasa A	MMP-2
	Gelatinasa B	MMP-9
Estromelisinias	Estromelisinina-1	MMP-3
	Estromelisinina-2	MMP-10
	Estromelisinina-3	MMP-11
Matrilisinias	Matrilisinina	MMP-7
	Matrilisinina-2, endometasa	MMP-26
MMP de membrana	MT1-MMP	MMP-14
	MT2-MMP	MMP-15
	MT3-MMP	MMP-16
	MT4-MMP	MMP-17
	MT5-MMP	MMP-24
	MT6-MMP, leucolisina	MMP-25
Otras MMP	Metaloelastasa	MMP-12
	RASI-1	MMP-19
	Enamelisina	MMP-20
	XMMP	MMP-21
	CA-MMP	MMP-23
	CMMP	MMP-27
	Epilisina	MMP-28

**Tabla 1: Clasificación de la familia de las metaloproteinasas (MMP) en humanos.**

Las MMP cumplen diversos roles que contribuyen al inicio y progresión del cáncer y al establecimiento de metástasis <sup>11</sup>. En la primera etapa de esta enfermedad, la MEC y la membrana basal en particular forman una barrera física que contiene al tumor sólido evitando su crecimiento. Cuando el tumor se encuentra confinado de esta manera es considerado benigno. La degradación de la membrana basal por parte de la actividad proteolítica de las MMP permite la invasión del tumor al tejido circundante

---

y marca el inicio de la progresión hacia la malignidad <sup>8</sup>. La degradación de la MEC facilita el crecimiento de los vasos sanguíneos hacia el tumor, proceso importante para la supervivencia del tumor y el inicio de la intravasación.

Las MMP también cumplen funciones en la modulación de señalización celular. Entre los sustratos de estas proteínas se encuentran citocinas, quimiocinas y receptores de superficie celular <sup>12,13</sup>. El procesamiento de estas moléculas por parte de las MMP contribuye a la activación o inactivación de vías de señalización celular relacionadas a la proliferación del tumor, angiogénesis, modulación del sistema inmune y a la migración de las células tumorales <sup>10,14–16</sup>. Un ejemplo de esta función es la regulación de los factores de crecimiento insulínico (IGF). Los IGF son importantes factores de crecimiento con funciones relacionadas a crecimiento, proliferación y supervivencia de las células. Normalmente los IGF tienen su función modulada por su unión a las proteínas de unión a IGF (IGFBP) que bloquea su interacción con el receptor de IGF-1 (IGF1R) <sup>17</sup>. Algunas MMP – como MMP-1, MMP-2, MMP-3 y MMP-11 – pueden degradar las IGFBP, liberando el IGF al medio extracelular. El IGF liberado de esta forma puede unirse al receptor IGF1R estimulando la proliferación de las células circundantes y el crecimiento del tumor <sup>18</sup>.

A su vez, las MMP cumplen un rol en la modulación de la transición epitelial-mesenquimal (TEM) mediante la degradación de proteínas de adhesión en la superficie celular <sup>16</sup>. En particular, la pérdida de E-cadherina por la acción de las MMP puede inducir la TEM <sup>19</sup>. E-cadherina es una importante molécula de adhesión célula-célula en epitelios. Esta molécula puede ser degradada por diversas MMP, principalmente MMP-3, MMP-7, MMP-9 y MT1-MMP (MMP-14) <sup>20</sup>.

La expresión y actividad de las MMP se encuentra elevada en casi todos los tipos de cáncer y esta expresión elevada se asocia generalmente a un mal pronóstico de la enfermedad <sup>16,21,22</sup>. En particular para cáncer de próstata, múltiples estudios han observado que la expresión de diversas MMP se correlaciona con tumores de estado avanzado y metástasis <sup>16,18,23</sup>. Una revisión sistemática de la bibliografía disponible sobre el rol de las MMP en cáncer de próstata encontró que existe evidencia fuerte de una asociación entre la progresión de cáncer de próstata y la expresión elevada de MMP-2, MMP-7, MMP-9 y MT1-MMP. También encuentran evidencia menos fuerte



---

de una asociación con MMP-1, MT3-MMP y MT6-MMP. Para otras MMP, como MMP-11 y MMP-13, los resultados de distintos trabajos son contradictorios y no hay concordancia sobre si contribuyen a la progresión de este cáncer <sup>23</sup>.

No todas las MMP se asocian a progresión del cáncer. Se ha observado que algunas MMP – como MMP-8 y MMP-26 – se asocian generalmente a un mejor pronóstico de la enfermedad, indicando que pueden tener un efecto supresor <sup>24,25</sup>. Las MMP pueden ejercer un efecto protector mediante la modulación de angiogénesis, apoptosis y la respuesta inmune <sup>11</sup>. MMP-8 ejerce su efecto protector mediante la activación de la molécula decorina, una proteína de la MEC. Esta molécula activada se une al factor de crecimiento transformante beta (TGF $\beta$ ), bloqueando su interacción con receptores de membrana y favoreciendo la expresión de genes supresor de tumores <sup>26</sup>.

El rol de las MMP depende en parte del contexto en el que se expresa. Una misma MMP puede tener roles protectores o protumorigénicos dependiendo del tipo de tumor, tipo de célula que lo expresa, o el estadio de progresión del cáncer <sup>11,24</sup>. Estudios en ratones demostraron que MMP-11 tiene funciones antiapoptóticas que favorece el crecimiento y supervivencia del tumor en las primeras etapas de la enfermedad. A su vez, los mismos estudios observaron que la expresión de MMP-11 se correlaciona con un menor número de lesiones metastásicas, indicando que MMP-11 puede ejercer un efecto protector en el cáncer avanzado <sup>27-29</sup>.

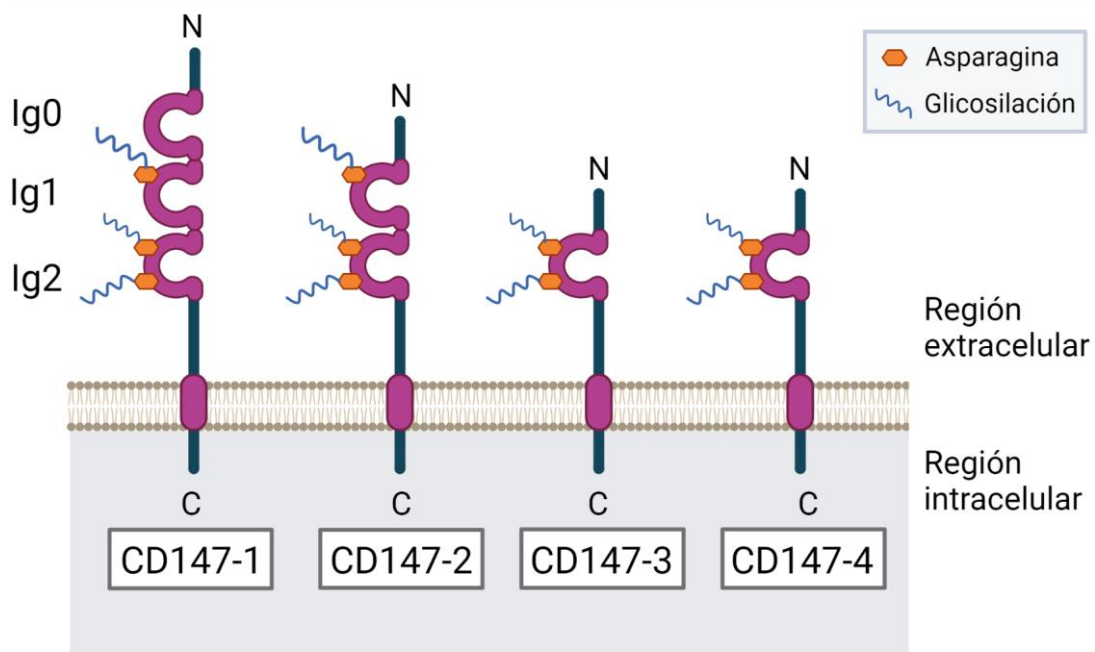
Para comprender mejor el funcionamiento de estas MMP en cáncer es importante entender el contexto en el que se expresan y los factores que modulan su expresión. La mayoría de las MMP no son producidas por células tumorales, sino que se originan principalmente de células del estroma circundante. Las células tumorales reclutan a las células del tumor mediante la secreción de moléculas de señalización o mediante interacciones célula-célula <sup>11</sup>. Mediante estas interacciones, el tumor puede inducir la expresión de MMP en células del estroma. Uno de los factores involucrados en esta función es la proteína CD147, también conocida como *basigin* (BSG) o *extracellular matrix metalloproteinase inducer* (EMMPRIN) <sup>30,31</sup>.

### 1.3. CD147

La proteína CD147 es una glicoproteína transmembrana perteneciente a la superfamilia de las inmunoglobulinas <sup>32</sup>. Es una proteína pleiotrópica con funciones

asociadas a reproducción y desarrollo, funcionamiento del sistema nervioso y sensorial, y a la regulación de la respuesta inmune <sup>33,34</sup>.

El gen *CD147*, ubicado en el cromosoma 19 (p13.3), codifica cuatro isoformas distintas de esta proteína. Estas son producidas por variaciones en el corte y empalme del ARNm y por variaciones en el sitio de inicio de la transcripción (ver Figura 8). La proteína CD147 está formada por un dominio citoplasmático, un dominio transmembrana y, dependiendo de la isoforma, contiene uno (CD147-3 y CD147-4), dos (CD147-2) o tres (CD147-1) dominios extracelulares de tipo inmunoglobulina (Figura 3). La isoforma 2 de CD147 es la más abundante y la mejor caracterizada <sup>35</sup>.



**Figura 3: Representación esquemática de las isoformas de CD147**

La proteína CD147 está compuesta de uno a tres dominios extracelulares de tipo inmunoglobulina (Ig0, Ig1 y Ig2), un dominio transmembrana y un dominio intracelular corto. La isoforma 1 de CD147 se expresa específicamente en la retina y es la única con el dominio Ig0. La isoforma 2, contiene dos dominios Ig y es la isoforma más abundante en los tejidos. Los dominios extracelulares contienen residuos de asparagina conservados que funcionan como sitios de glicosilación. Imagen creada con BioRender.com.

En 1984, CD147 fue inicialmente caracterizada como un factor molecular presente en cáncer de pulmón capaz de inducir la producción de colagenasas en fibroblastos <sup>36</sup>. Desde ese descubrimiento inicial, se ha encontrado que CD147 cumple un rol en la

---

modulación de la expresión y actividad de diversas MMP, incluyendo MMP-1<sup>36-40</sup>, MMP-2<sup>37-42</sup>, MMP-3<sup>37-39</sup>, MMP-9<sup>40,42-45</sup> y las MMP de membrana MT1-MMP<sup>39,41,46</sup> y MT2-MMP<sup>41</sup>. Además de regular la expresión de MMP en fibroblastos<sup>36,41,47</sup>, se ha encontrado evidencia de que CD147 puede regular la expresión de MMP en macrófagos<sup>44,45</sup> y células endoteliales<sup>38</sup>. Incluso se observó que CD147 es capaz de inducir la producción de MMP y CD147 en las mismas células tumorales<sup>38,42,48</sup>.

No está claro cuál es el mecanismo de acción por el que CD147 induce la producción de MMP. Múltiples estudios han demostrado que la sobreexpresión de CD147 lleva a un aumento en la expresión de MMP y que la inhibición de CD147 tiene el efecto contrario. Sin embargo, no se ha identificado un mecanismo por el cual CD147 pueda inducir directamente la producción de MMP. Es probable que esta función ocurra de manera indirecta, mediante interacciones con otras proteínas de membrana o mediante la modulación de otros factores involucrados en la regulación de MMP<sup>35,49</sup>.

Los dominios extracelulares de CD147 (Ig1 y Ig2) poseen residuos conservados de asparagina que funcionan como sitios de glicosilación (Figura 3). Se ha observado que CD147 tiene formas con glicosilación baja (LG-CD147) y formas con glicosilación alta (HG-CD147), y que las formas con glicosilación alta son más comunes en tumores<sup>35,50</sup>. El nivel de glicosilación de CD147 es importante para su función como inductor de MMP. Experimentos con formas no glicosiladas de CD147 demostraron que estas no son capaces de inducir la producción de MMP, y que la presencia de CD147 no glicosilado inhibe la inducción de MMP mediada por HG-CD147<sup>35,48,51</sup>. La glicosilación de Ig1 está regulada por interacciones con la proteína caveolin-1 que puede unirse al dominio Ig2 de CD147 e inhibir la conversión de LG-CD147 a HG-CD147<sup>50</sup>.

Otro factor importante en la modulación de MMP es la formación de dímeros CD147-CD147 en la membrana celular. Se ha observado que CD147 puede asociarse consigo mismo mediante la unión de sus dominios Ig1 y que esta dimerización es independiente de la glicosilación. El bloqueo de la dimerización de CD147 mediante anticuerpos o mediante mutaciones en Ig1 resulta en una reducción de su capacidad para inducir MMP en fibroblastos, indicando que la asociación CD147-CD147 es importante para esta función. Las interacciones en *trans* de esta molécula entre las

---

células del tumor y las del estroma permite que esta proteína actúe como su propio receptor celular. De esta manera las células tumorales usan CD147 como un mecanismo para la comunicación célula-célula en el proceso de inducción de MMP en células adyacentes <sup>35,51,52</sup>.

Las células tumorales también pueden usar CD147 para inducir MMP en células más distantes. Se ha visto que CD147 puede ser secretado al espacio extracelular en forma soluble completa o en forma de un fragmento de 22 kDa resultado del procesamiento proteolítico por MT1-MMP <sup>53,54</sup>. También se ha demostrado que CD147 es liberado en microvesículas extracelulares o exosomas y que estas vesículas son capaces de inducir la producción de MMP en fibroblastos <sup>55,56</sup>.

Además de contribuir a la progresión del cáncer por su rol como inductor de MMP, se ha visto que CD147 también contribuye a la progresión del cáncer de manera independiente de MMP. CD147 se puede asociar con diversas moléculas en la membrana y, mediante estas interacciones, CD147 cumple roles en invasión, angiogénesis, proliferación, supervivencia, migración celular, quimio resistencia y en el metabolismo del tumor <sup>35,49,57,58</sup>.

Se ha visto que CD147 forma complejos con integrinas en la membrana celular. Estos complejos activan la vía de señalización mediante quinasas de adhesión focal (FAK) regulando la invasión y migración celular <sup>59</sup>. También se ha visto que CD147 funciona como un receptor celular para la ciclofilina A (CypA). Esta proteína se encuentra sobreexpresada en cáncer y se ha demostrado que su interacción con CD147 induce la proliferación celular y regula la respuesta inmune en cáncer.

Además, se ha demostrado que CD147 puede inducir la expresión del factor de crecimiento endotelial vascular (VEGF) y de su receptor VEGFR-2 tanto en células endoteliales del estroma como en células del tumor, promoviendo la angiogénesis <sup>45,57,60,61</sup>. A su vez, se observó que CD147 interactúa con VEGFR-2 y esta interacción es necesaria para su activación por VEGF <sup>62</sup>.

La modulación del metabolismo celular en cáncer es una de las principales funciones protumorales de CD147. Esta proteína cumple un rol esencial en el funcionamiento de los transportadores de monocarboxilatos (MCT) que catalizan el transporte de ácido láctico a través de la membrana celular <sup>63,64</sup>. Las células tumorales dependen

---

de la glicolisis aeróbica para sus necesidades energéticas y deben eliminar el lactato producido por este proceso para mantener el pH intracelular <sup>65</sup>. La acidificación del microambiente tumoral por el lactato exportado de la célula promueve invasión y metástasis. El lactato también estimula la expresión de ácido hialurónico y de su receptor CD44. Se ha demostrado que la interacción CD44-CD147-ácido hialurónico promueve invasión y quimio resistencia mediante la estimulación de la activación de CD44 <sup>49</sup>.

CD147 está sobre expresado en la mayoría de los tipos de cáncer y su expresión se asocia a un mal pronóstico de la enfermedad <sup>66-71</sup>. Se ha visto que CD147 está sobreexpresado en cáncer oral, colorrectal, de esófago, gástrico, en glioblastoma, cáncer de hueso, en leucemia mieloide aguda, en cáncer de mama, melanoma, de ovario, de próstata, de pulmón, renal y de vejiga <sup>57</sup>. CD147 se expresa principalmente en los bordes de tumores sólidos, en la interfase tumor-estroma, donde contribuye a la progresión del cáncer por su rol como inductor de MMP en fibroblastos <sup>70</sup>. A su vez, se ha observado que la expresión de CD147 se encuentra elevada también en células metastásicas obtenidas de la médula ósea en pacientes con cáncer de mama, próstata, y pulmón <sup>72,73</sup>.

La mayoría de los estudios realizados sobre CD147 en próstata encontraron que la expresión de esta proteína se encuentra elevada en cáncer en comparación con próstata normal <sup>74-78</sup> (Tabla 2). La presencia de esta proteína en el tumor se asocia a la expresión de MMP-1, MMP-2 y MMP-9 <sup>74</sup> y a proliferación, invasión y metástasis. A su vez, la expresión de CD147 se correlaciona con el puntaje de Gleason, nivel de PSA en sangre, el estadio TNM, resistencia a quimioterapia y recurrencia bioquímica <sup>74-80</sup>. Sin embargo, algunos trabajos han encontrado que la expresión de CD147 se encuentra disminuida en cáncer de próstata y que los niveles bajos de esta proteína en el tumor se asocian a puntajes de Gleason altos, estadios de tumor avanzados, niveles de PSA altos y recurrencia bioquímica <sup>81-83</sup>.

Autor	Número de pacientes	Principales resultados
Zhong <i>et al.</i> , 2008 <sup>74</sup>	62 PCa 15 BPH	Expresión elevada de CD147, MMP-1, MMP-2 y MMP-9 en PCa en comparación con BPH
		Correlación entre expresión de CD147 y puntaje de Gleason alto ( $\geq 7$ ), estadio TNM avanzado (IIb y III), y reducción de sobrevida
Han <i>et al.</i> , 2009 <sup>75</sup>	62 PCa 30 BPH	Expresión elevada de CD147 en PCa en comparación con BPH
		Correlación entre expresión de CD147 y puntaje de Gleason alto ( $\geq 7$ ), estadio TNM avanzado (IIb y III), invasión, y reducción de sobrevida
Han <i>et al.</i> , 2010 <sup>76</sup>	101 PCa 68 Normal sano	No encuentran expresión positiva (>5% de células marcadas) de CD147 en tejidos normal
		Correlación positiva entre expresión de CD147 y estadio TNM, reducción de sobrevida
Bi <i>et al.</i> , 2011 <sup>77</sup>	300	Expresión de ARNm de CD147 elevada en cáncer en comparación con tejido normal adyacente. No encuentran relación entre expresión de ARNm de CD147 y características clinicopatológicas
		Expresión de proteína CD147 elevada en PCa sobre normal
		Correlación entre expresión de proteína CD147 y nivel de PSA preoperatorio, estadio TNM y puntaje de Gleason
Madigan <i>et al.</i> , 2008 <sup>78</sup>	120 PCa 20 Metástasis 40 Normal sano	Expresión de CD147 elevada en PCa sobre normal y mayor en PCa de grado alto sobre PCa de grado bajo
		Correlación entre expresión de proteína CD147 y nivel de PSA preoperatorio, estadio TNM y puntaje de Gleason
Pértega-Gomes <i>et al.</i> , 2011 <sup>83</sup>	171 PCa 135 Normal adyacente 40 PIN 14 Normal sano	Expresión de CD147 disminuida en PCa sobre normal sano y en PIN sobre normal sano.
		No hay diferencias en la expresión de CD147 entre PCa y normal adyacente, entre PIN y normal adyacente o entre PIN y PCa
		Correlación entre expresión de proteína CD147 y nivel de PSA preoperatorio, estadio TNM avanzado (pT3), puntaje de Gleason, recurrencia bioquímica e invasión
Zhong <i>et al.</i> , 2012 <sup>80</sup>	240	Correlación entre expresión de proteína CD147 y puntaje de Gleason, margen quirúrgico positivo, recurrencia bioquímica, metástasis y reducción de sobrevida
Bauman <i>et al.</i> , 2015 <sup>81</sup>	846	Expresión de CD147 y E-cadherina disminuida en PCa local, PCa agresivo, y metástasis en comparación con tejido normal adyacente
		No hay diferencias en la expresión de CD147 o E-cadherina en BPH, PIN en comparación con tejido normal adyacente
		Correlación inversa entre expresión de proteína CD147 y puntaje de Gleason
Grupp <i>et al.</i> , 2013 <sup>82</sup>	11152	Expresión de CD147 disminuida en PCa en comparación con tejido normal
		Correlación inversa entre expresión de proteína CD147 y valores de PSA preoperatorios, puntaje de Gleason y estadio TNM

**Tabla 2: Estudios realizados sobre expresión de CD147 en cáncer de próstata**

PCa: muestras de tumor primario de cáncer de próstata; BPH: muestras de hiperplasia prostática benigna; Normal sano: muestras de tejido de próstata de individuos sin cáncer diagnosticado; Normal adyacente: muestra de tejido prostático normal adyacente al tumor primario; PIN: muestra de neoplasia intraepitelial prostática.

Las diferencias observadas entre los distintos trabajos se pueden deber a diferencias metodológicas. En todos los trabajos mencionados la expresión de CD147 se midió mediante inmunohistoquímica (IHC), una técnica basada en el uso de anticuerpos para detectar una proteína en una muestra de tejido. Este método mide la expresión de una proteína de manera cualitativa o semicuantitativa en base al conteo manual del porcentaje de células marcadas y una evaluación subjetiva de la intensidad del marcado. Los resultados de este análisis dependen del operador y de la sensibilidad de los anticuerpos usados, dificultando realizar comparaciones entre distintos trabajos. Estas diferencias podrían explicar las diferencias en los resultados obtenidos por los diferentes trabajos. Es importante destacar que los trabajos que muestran una correlación positiva entre CD147 y un mal pronóstico de la enfermedad analizan un bajo número de pacientes, menor a 300 pacientes en todos los casos. Por otro lado, el trabajo de Grupp *et al.* y el de Bauman *et al.*, que encuentran resultados opuestos, analizaron un total de 11152 y 846 muestras de pacientes respectivamente <sup>81,82</sup>.

#### 1.4. Edición génica

Una de las herramientas más ampliamente utilizadas para estudiar la función de una proteína en una célula es anular su función mediante la delección o silenciamiento del gen que la expresa. Uno de los métodos más comunes para este propósito es el uso de nucleasas para inducir mutaciones de pérdida de función en el gen estudiado <sup>84</sup>.

Las nucleasas son enzimas capaces de cortar la cadena de nucleótidos de los ácidos nucleicos. Mediante el uso de nucleasas *in vivo* se puede inducir daños al ADN de manera dirigida a un gen de interés. Normalmente el ADN es reparado por la célula, pero cuando el daño resulta en un corte de ambas hebras la reparación puede ocurrir con errores. Los cortes de doble hebra activan una de dos vías de reparación. La vía de reparación por recombinación homóloga (HDR del inglés *Homology Directed Repair*) utiliza una molécula de ADN homólogo como molde para restaurar la molécula

---

dañada. Este homólogo puede ser una cromátida hermana, un cromosoma homólogo o una molécula de ADN sintética diseñada para este propósito. La vía de HDR resulta generalmente en la reparación sin errores, aunque cualquier diferencia que exista en la secuencia de la molécula homóloga va a ser copiada a la molécula dañada durante la reparación.

La alternativa a la vía HDR es la reparación por unión de extremos no homólogos (NHEJ del inglés *Non-homologous End Joining*). La vía NHEJ repara el ADN uniendo directamente ambos extremos de la cadena cortada. Al no usar una secuencia homóloga como molde esta forma de reparación frecuentemente resulta en la introducción de errores en la forma de inserciones o deleciones de nucleótidos (*indels*). Cuando estos *indels* ocurren en la secuencia codificante de un gen causan que se altere el marco de lectura del mismo. Esto causa que se altere la secuencia de codones codificada por el gen y frecuentemente resulta en que en el nuevo marco de lectura aparezcan codones de terminación prematuros.

Las mutaciones que introducen codones de terminación prematuros se denominan mutaciones sin sentido o mutaciones *nonsense* y resultan en la producción de una proteína incompleta y generalmente no funcional. Para prevenir la expresión de este tipo de proteínas, los eucariotas cuentan con un mecanismo de control que degrada los ARN mensajeros de estos genes. Este mecanismo de degradación de ARN mediada por mutaciones sin sentido (NMD del inglés *Nonsense-mediated decay*) se activa durante la traducción cuando el ribosoma encuentra un codón de terminación a una distancia mayor a 50 bases previo al último exón <sup>85</sup>.

Mediante la generación de cortes doble hebra en el ADN y la subsecuente reparación con errores por la vía NHEJ se pueden inducir mutaciones de pérdida de función en un gen de interés. Existen distintos tipos de nucleasas capaces de generar cortes de doble hebra. Las más usadas son aquellas que tienen especificidad por una secuencia determinada y cuya especificidad se puede alterar para dirigir las a un sitio de corte deseado <sup>86</sup>. Entre estas se encuentran las nucleasas de tipo dedo de zinc (*Zinc Finger Nucleases*; ZFN), las nucleasas de tipo *Transcription Activator Like Effector* (TALEN) y el sistema CRISPR-Cas9 <sup>84</sup>.

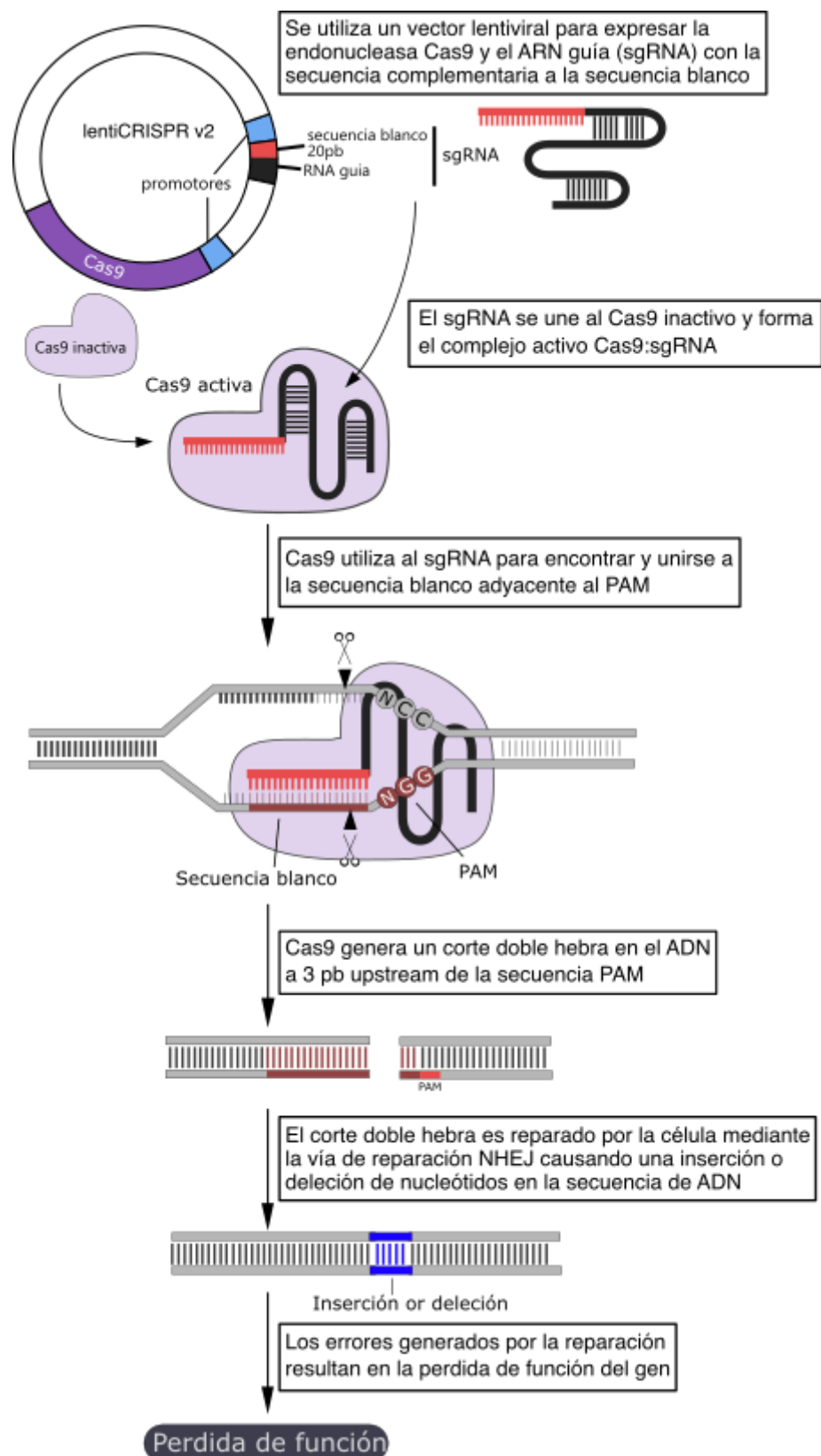


---

Tanto ZFN como TALEN son nucleasas artificiales construidas por la unión de un dominio de reconocimiento de ADN y un dominio de corte inespecífico. El primer dominio reconoce y se une a una secuencia específica de ADN dirigiendo la actividad del dominio de corte al sitio deseado. En ambos casos la especificidad del dominio de reconocimiento puede ser modificada y dirigida a un gen de interés. Esta alteración se logra mediante cambios en su estructura lo que requiere de complejos procesos de ingeniería proteica <sup>87</sup>.

Por otro lado, el sistema CRISPR-Cas9 proviene de un sistema natural de defensa en procariontes que fue adaptado para su uso como herramienta de ingeniería genética <sup>88</sup>. El sistema consiste de la proteína Cas9 con actividad endonucleasa y una molécula de ARN que guía la nucleasa a un sitio específico <sup>89</sup>. La especificidad de secuencia de este sistema está determinada por una secuencia de 20 bases de largo contenida en el ARN guía (Figura 4). Esta molécula reconoce su blanco en el ADN mediante apareamiento de bases Watson-Crick. La secuencia del blanco, también denominada protoespaciador, puede ser cualquier secuencia de 20 nucleótidos que se encuentre adyacente a un pequeño motivo con secuencia 5'-NGG. En el genoma humano este motivo, denominado PAM (del inglés *protospacer adjacent motif*), se encuentra en promedio una vez cada 8-12 pares de bases <sup>87</sup>.

Una de las ventajas principales de este sistema frente al ZFN y TALEN, es que la especificidad de secuencia del sistema CRISPR-Cas9 no depende de la estructura de la proteína, sino que es determinada por los apareamientos de bases Watson-Crick entre el ARN guía y la molécula a editar. La modificación de su especificidad es sencilla, requiriendo solamente de la síntesis y clonado de una nueva molécula de ARN de 20 bases de largo.



**Figura 4: Inactivación de un gen mediante el sistema de edición genética CRISPR-Cas9**

Esquema del mecanismo de acción del sistema de edición genética CRISPR-Cas9 usado para generar una mutación de pérdida de función de un gen. Se muestra únicamente el sistema utilizado en la realización de esta tesis.

---

### 1.5. Historia de CRISPR-Cas9

El desarrollo del sistema CRISPR-Cas9 como una herramienta de edición génica comienza en 1987 cuando Ishino *et al.* descubren en bacterias *Escherichia coli* un grupo de secuencias repetidas que se encontraban cercanas al gen *iap* que estaban estudiando <sup>90</sup>. Posteriormente, el mismo tipo de secuencias repetidas fueron encontradas en múltiples especies de procariontes y se establece que estas secuencias pertenecen a una misma familia. La familia se define como un grupo de repeticiones palindrómicas cortas y regularmente separadas por secuencias espaciadoras no repetitivas. La presencia de estos espaciadores caracteriza a esta familia y la distingue de otras secuencias repetidas <sup>91</sup>.

En 2002 Jansen *et al.* proponen usar la nomenclatura CRISPR (del inglés *clustered regularly interspaced short palindromic repeats*) para referirse a esta familia de repetidos. Identifican cuatro genes asociados a CRISPR a los que le dan el nombre de *cas* (del inglés *CRISPR-associated*). Muestran que los genes *cas* se encuentran invariablemente adyacentes al loci de CRISPR, sugiriendo una relación funcional entre ellos. Mediante análisis de secuencia encuentran que los genes *cas* contienen motivos helicasas y exonucleasas y proponen que pueden cumplir funciones en reparación de ADN <sup>92</sup>.

En 2005 se publican tres trabajos que muestran que las secuencias espaciadoras en los loci CRISPR derivan de elementos genéticos foráneos como ser plásmidos y bacteriófagos <sup>93-95</sup>. En particular, Mojica *et al.* demuestran que las secuencias de los espaciadores tienen homología con secuencias virales y descubren que las células que portan espaciadores originados de bacteriófagos son resistentes a la infección por los mismos. Este descubrimiento sugiere una relación entre las secuencias CRISPR y la inmunidad contra ADN invasivo por parte de las bacterias <sup>93</sup>.

En 2006, Makarova *et al.* son los primeros en hipotetizar que los genes *cas* actúan en conjunto con las secuencias CRISPR asociadas y pueden constituir un sistema de defensa contra bacteriófagos y plásmidos. Ellos proponen que este sistema CRISPR-*cas* produce pequeñas moléculas de ARN que funcionan de manera análoga a los ARN pequeños de interferencia en eucariotas <sup>96</sup>. Poco tiempo después Barrangou *et*

---

*al.* demuestran *in vivo* que existe una asociación funcional entre los genes *cas*, los espaciadores de CRISPR y la resistencia a bacteriófagos <sup>97</sup>.

En 2008, Devau *et al.* investigan el rol de CRISPR en la resistencia a bacteriófagos en bacterias de *Streptococcus thermophilus*. Proponen el nombre de protoespaciadores para referirse a la región del genoma bacteriófago homóloga al espaciador adquirido del mismo. Demuestran que la resistencia a bacteriófagos mediada por CRISPR requiere que los espaciadores tengan secuencia idéntica a la del protoespaciador, y que la adquisición de espaciadores no ocurre al azar, sino que provienen solo de protoespaciadores que se encuentran adyacentes una secuencia 5'-NNAGAAW <sup>98</sup>. Posteriormente, Mojica *et al.* demuestran que el requisito de un motivo adyacente al protoespaciador (PAM del inglés *protospacer adjacent motifs*) es una característica común a los sistemas CRISPR-Cas y que la secuencia de este motivo depende de la variante específica de CRISPR-Cas <sup>99</sup>.

En 2010, Garneau *et al.* presentan la primera evidencia del mecanismo de acción de CRISPR-Cas. Demuestran en *S. thermophilus* que el sistema CRISPR-Cas genera cortes en ADN de doble hebra y que este corte ocurre dentro de la secuencia del protoespaciador. Observan que el corte ocurre siempre en la misma posición del protoespaciador, a 3 pb de distancia del extremo adyacente al PAM <sup>100</sup>.

En 2011, Deltcheva *et al.* estudian la maduración de los ARN de CRISPR (crRNA) en *Streptococcus pyogenes*. Los crRNA se transcriben como un precursor de ARN (pre-crRNA) a partir de una unidad espaciador-repetido de CRISPR y deben ser procesados para generar el crRNA maduro funcional. Demuestran que en esta bacteria la maduración del crRNA está dirigida por otro transcripto de ARN al que llaman ARN CRISPR activador en trans (tracrRNA). También demuestran que esta maduración depende de la RNasa III y de la proteína Cas9 (llamada Csn1 en este trabajo) <sup>101</sup>.

En 2012, el mismo grupo, dirigido por Emmanuelle Charpentier y Jennifer A. Doudna, publican un trabajo demostrando el mecanismo de acción del sistema CRISPR-Cas9 <sup>88</sup>. Demuestran que además de su rol en la maduración de crRNA, Cas9 participa directamente en la destrucción del ADN foráneo. Encuentran que Cas9 genera un corte de doble hebra en el ADN dirigido por complementariedad de secuencia entre

---

crRNA y el protoespaciador. Observan que, al igual que lo observado por Garneau *et al.* <sup>100</sup>, en el sistema CRISPR-Cas9 el corte doble hebra ocurre siempre en la misma posición del protoespaciador, a 3 pb de distancia del PAM. En este trabajo también demuestran que el crRNA por su cuenta no es suficiente para activar a Cas9 y que la acción endonucleasa de esta proteína es guiada por una estructura de ARN formada por el apareamiento de bases de tracrRNA y la sección de repetido de crRNA. Comprueban que la dupla tracrRNA:crRNA se puede fusionar en una única molécula y que este ARN quimérico también es capaz de dirigir la actividad de Cas9. Demuestran que una única molécula de ARN puede ser usada para programar la endonucleasa Cas9 dirigiéndola a un sitio de interés y proponen que el sistema CRISPR-Cas9 puede ser usado como una herramienta de edición génica que es eficiente, versátil y fácilmente programable. En 2020, el Premio Nobel de Química fue otorgado conjuntamente a Emmanuelle Charpentier y Jennifer A. Doudna “por el desarrollo de un método para la edición del genoma” <sup>102</sup>.

En 2013, dos trabajos demuestran la aplicabilidad del sistema CRISPR-Cas9 como una herramienta de ingeniería genética en eucariotas. Mali *et al.* logran expresar el sistema CRISPR-Cas9 en células humanas y demostraron que el sistema es capaz de generar cortes doble hebra guiadas por ARN en estas células <sup>103</sup>. Por otro lado, Cong *et al.* desarrollan una herramienta de edición génica eucariota basada en el sistema CRISPR-Cas9 (ver Figura 4) y demuestran que su sistema de expresión puede reconstituir el sistema CRISPR de *S. pyogenes* en células de humano y ratón. También demuestran que usando múltiples ARN guías este sistema puede editar múltiples sitios simultáneamente <sup>104</sup>.

En la última década se desarrollaron múltiples tecnologías basadas en sistemas CRISPR-Cas que varían tanto en la especificidad de secuencia como en sus funciones. Estas nuevas tecnologías expanden las capacidades del sistema CRISPR-Cas9 y permiten nuevas formas de manipular el genoma. Entre las aplicaciones de estos sistemas se encuentran la edición de ADN, edición de ARN y la regulación de expresión génica <sup>105,106</sup>.

En este trabajo se utilizó el sistema CRISPR-Cas9 desarrollado por Cong *et al.* <sup>104</sup> para inducir mutaciones de pérdida de función en el gen CD147 y generar una línea

celular de cáncer de próstata metastásico CD147<sup>-/-</sup>. Mediante el estudio de esta línea celular se espera poder elucidar el rol de CD147 como modulador de MMP y su papel en la progresión de la metástasis en cáncer de próstata.

## 2. Hipótesis y objetivos

### 2.1. Hipótesis

La transfección de células provenientes de una metástasis ósea de un tumor prostático con un vector de expresión para el sistema de edición genómica CRISPR-Cas9 dirigido hacia el gen *CD147*, produce mutaciones de pérdida de función del gen.

### 2.2. Objetivo general

Generar una línea celular metastásica prostática (PC3) estable *knock-out* para el gen *CD147*<sup>-/-</sup> mediante el sistema de edición genómica CRISPR-Cas9.

### 2.3. Objetivos específicos

- a) Realizar un meta-análisis de expresión génica del gen *CD147* de bases de datos disponibles
- b) Diseñar y clonar los ARN guía para la generación del vector CRISPR-Cas9-*CD147*
- c) Transfectar, seleccionar y verificar la línea celular de cáncer de próstata *knock-out* (*CD147*<sup>-/-</sup>)

---

### 3. Materiales y métodos

#### 3.1. Análisis de datos de expresión génica en bases de datos

Todos los análisis de expresión génica en bases de datos se realizaron utilizando el portal *UCSC Xena* <sup>107</sup>. Para los análisis se seleccionaron los estudios “TCGA Pan-Cancer (PANCAN)”, “TCGA Prostate Cancer (PRAD)” y “TCGA TARGET GTEx” <sup>108</sup>. El estudio “TCGA Pan-Cancer” contiene datos de expresión provenientes del proyecto *The Cancer Genome Atlas* (TCGA), mientras que el estudio “TCGA TARGET GTEx” contiene datos combinados provenientes de los proyectos TCGA, *Therapeutically Applicable Research To Generate Effective Treatments* (TARGET), y *Genotype-Tissue Expression* (GTEx). El estudio “TCGA Prostate Cancer (PRAD)” es un subconjunto del estudio “TCGA Pan-Cancer (PANCAN)” que contiene solo muestras de pacientes con cáncer de próstata.

A partir de estos estudios se crearon subconjuntos de datos seleccionando muestras que contengan datos de expresión de *CD147* en tejido normal no tumoral (sample\_type = “Solid Tissue Normal” o sample\_type = “Normal Tissue”), en tumor primario (sample\_type = “Primary Tumor”) o en metástasis (sample\_type = “Metastatic”). El resto de las muestras se descartaron para los análisis.

El estudio “TCGA Pan-Cancer (PANCAN)” se utilizó para analizar la expresión de *CD147* en muestras pareadas. En este estudio el nombre de cada muestra está compuesto de el nombre del proyecto del cual proviene la muestra (TCGA en este caso), un identificador de paciente, y un código numérico de dos dígitos que identifica el tipo de muestra (01 para tejido tumoral, 11 para tejido normal y 06 para metástasis). Utilizando la función de filtros y subgrupos del portal *UCSC Xena* <sup>107</sup>, se usó el nombre de las muestras para identificar las muestras que provienen de un mismo paciente (es decir, que tienen el mismo identificador de paciente) y se crearon dos subconjuntos de datos; uno conteniendo muestras de tejido normal no tumoral y muestras de tejido tumoral, y el otro conteniendo muestras de tejido tumoral y tejido metastásico. Se repitió esta metodología para el análisis de muestras pareadas en el estudio “TCGA Prostate Cancer (PRAD)”.

Para el análisis de muestras no pareadas se utilizó el estudio “TCGA TARGET GTEx”. Se emplearon los datos de expresión de *CD147* en tejido normal no tumoral



---

provenientes del proyecto GTEx en conjunto con los datos de expresión en tejido tumoral y metástasis provenientes del proyecto TCGA. Las muestras provenientes del proyecto TARGET no fueron incluidas en este análisis.

Para comparar la expresión de *CD147* entre los tejidos normales no tumorales provenientes del proyecto TCGA y los tejidos normales no tumorales provenientes del proyecto GTEx se utilizó la función de filtros y subgrupos del portal *UCSC Xena*<sup>107</sup> para seleccionar las muestras que provienen del mismo tipo de tejido (*primary\_site*) en ambos proyectos. Luego se comparó la expresión de *CD147* en cada subgrupo.

Para estudiar la sobrevida se utilizó el estudio TCGA Prostate Cancer (PRAD) y se generaron curvas Kaplan Meier. Se compararon las muestras en el cuartil superior de expresión de *CD147* (muestras con valores de expresión mayor que 75% de las muestras) con muestras en el cuartil inferior de expresión (muestras con valores de expresión menor que 75% de las muestras).

En todos los análisis se utilizaron las herramientas estadísticas incluidas en el portal *UCSC Xena*<sup>107</sup> para evaluar la significación estadística de los resultados, considerándose como significativos valores *p* menores a 0,05. Se utilizó la prueba *t* de Welch en los análisis de expresión y la prueba *logRank* en el análisis de sobrevida

### 3.2. Obtención de secuencias

Las secuencias para el diseño de cebadores y para el diseño de ARN guías se obtuvieron utilizando la base de datos *RefSeq*<sup>109</sup> (<https://www.ncbi.nlm.nih.gov/refseq/>). Se obtuvo la secuencia de referencia del gen *CD147* (NG\_007468.1) y se identificaron y obtuvieron las secuencias de todas las variantes de transcritos de este gen. Además, se obtuvieron las secuencias de referencia de los posibles genes *off-target* del sistema CRISPR-Cas9 (genes *TTF2*, *AP3D1*, *CELSR2*, *PPP1R3F*, *MRGPRX1*, *AGAP1*, *FHAD1*, *RIPK4* y *MAPKBP1*).

La secuencia del vector lentiviral *lentiCRISPR v2* se obtuvo de la web de Addgene (<https://www.addgene.org/52961/sequences/>).

### 3.3. Diseño de cebadores

Para el diseño de todos los cebadores se utilizó la herramienta *primer-BLAST* (<https://www.ncbi.nlm.nih.gov/tools/primer-blast>)<sup>110</sup> y se utilizaron como molde las secuencias de referencia obtenidas de la base de datos *RefSeq*.

Para el diseño de los cebadores que amplifican los blancos genómicos del sistema CRISPR-Cas9 se limitó el tamaño del amplicón entre 400 y 600 pb y para el resto de los parámetros se utilizaron los valores que la herramienta ofrece por defecto. Entre los posibles cebadores dados por la herramienta se seleccionaron preferentemente cebadores que tengan como última base un C o G, que tengan pocas bases repetidas, y que tengan similar temperatura de *melting* ( $T_m$ ) entre pares. Para el diseño de los cebadores utilizados en RT-qPCR se limitó el tamaño del amplicón entre 60 y 90 pb de largo y se limitó la búsqueda a cebadores que aparezcan en la unión exón-exón del ARNm de *CD147*. Para el resto de los parámetros, se utilizaron los valores que la herramienta ofrece por defecto.

Se diseñaron cebadores para amplificar por PCR los blancos genómicos del sistema CRISPR-Cas9. Se diseñaron cebadores que amplifican las secuencias blanco de los ARN guía (exón 4 de *CD147* y exón 5 de *CD147*) así como para las secuencias *off-target* de los mismos (genes TTF2, AP3D1, CELSR2, PPP1R3F, MRGPRX1, AGAP1, FHAD1, RIPK4 y MAPKBP1). También se diseñaron cebadores para amplificar el ADN copia (ADNc) del ARNm de *CD147*. Los cebadores diseñados fueron sintetizados por MacroGen (Seúl, Corea del Sur).

Adicionalmente, el laboratorio cuenta con un par de cebadores que amplifican el gen de actina *ACTB* ( $\beta$ 1actina 1 [CACCCAGCACAAATGAAGATCA] y  $\beta$ 1actina 2 [CTCGTCATACTCCTGCTTGC]), y un par de cebadores que amplifican el gen cas9 (cas9\_F [GCCGAGAATATCATCCACCTG] y cas9\_R [CTCTTTGGTGCTGGTGACCT]) que también fueron utilizados en este trabajo (Tabla 7).

#### 3.3.1. Evaluación de cebadores

Para cada par de cebadores se realizaron PCR de prueba utilizando distintas temperaturas de hibridación en el termociclador. Para cada par de cebadores se hicieron entre 2 y 5 reacciones de prueba con temperaturas de entre 0 y 5 °C por debajo de la temperatura de fusión de los cebadores. Como ADN molde se utilizó

---

ADN genómico humano (24,1 ng/μL) o el vector lentiCRISPR v2 (3,8 μg/μL). Luego, cada producto de PCR se resolvió por electroforesis en gel de agarosa 1% y se visualizó mediante fluorescencia en UV por la adición de bromuro de etidio. Para los pares de cebadores que no producen una única banda en las condiciones elegidas (PPP1R3F y MRGPRX1; ver Tabla 7), se repitió la prueba con la adición de 5% de dimetilsulfóxido (DMSO) en la mezcla de reacción.

Las PCR se realizaron en 20 μL de volumen final conteniendo 0,5 μM de cada cebador, 1X de mezcla de reacción MangoMix™ (Meridian Bioscience ® cat. BIO-25033) y 1 μL de ADN molde (ADN genómico humano o el vector lentiCRISPR v2). Para cada par de cebadores evaluados también se hizo una reacción de control sin ADN molde, completando el volumen con H<sub>2</sub>O. Pevio calentamiento de la reacción a 95 °C durante 5 minutos, se realizaron 35 ciclos de amplificación, en los cuales la temperatura alternó en tres etapas sucesivas de desnaturalización del ADN, hibridación de los cebadores y polimerización del ADN. Para la desnaturalización del ADN se utilizó una temperatura de 95 °C durante 15 segundos. En la etapa de hibridación se evaluaron en simultáneo distintas temperaturas de hibridación para cada par de cebadores, manteniéndose la reacción a la temperatura de prueba durante 30 segundos. La etapa de polimerización se realizó a una temperatura de 72 °C durante 30 segundos. Luego de los ciclos de amplificación se dejó la reacción a 72 °C durante 7 minutos para asegurar la polimerización completa de todos los fragmentos.

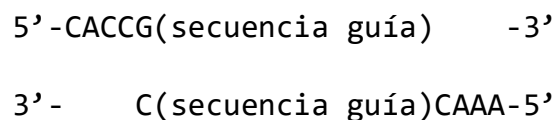
Los productos de PCR se analizaron por electroforesis en gel de agarosa 1% como se describe en la sección 3.5.

#### 3.4. Diseño de ARN guías

A partir de las secuencias de referencia obtenidas de la base de datos *RefSeq*, se identificaron los exones de ARNm de *CD147* que son comunes a todas las variantes de transcripto de este gen. Luego se utilizó la secuencia de referencia de la variante 2 del transcripto (NCBI NM\_198589.3) para diseñar ARN guía que tengan como blanco uno de los exones identificados. Se seleccionaron los exones 4 y 5 de este ARNm, ya que son los primeros exones que son comunes a todas las variantes de *CD147*.

Para diseñar los ARN guía se utilizó la herramienta web *CRISPR design tool*<sup>111</sup> (<http://crispr.mit.edu>) ingresando como parámetro las secuencias del exón 4 o exón 5 de *CD147*. La herramienta toma como entrada una secuencia genómica a la cual se quiere editar y busca todos los posibles ARN guía que tengan esta secuencia como blanco de Cas9. Luego, le asigna a cada guía un puntaje (*hit score*) de 0 a 100% que es aproximadamente igual a la probabilidad que tiene el ARN guía para unirse a su secuencia genómica objetivo y no a secuencias *off-target*. Finalmente, se presenta una lista de los ARN guía ordenada en base a este puntaje, y se presenta una lista de los potenciales *off-target* de cada guía. Para la selección de los ARN guía se siguieron las recomendaciones de los creadores de la herramienta y se consideró como blanco adecuado aquel que presente un *hit score* mayor a 50 y que no tenga genes *off-targets* con un hit score alto (> 5). A partir de los resultados producidos por la herramienta se seleccionaron las 3 secuencias blanco con puntaje más alto, uno con blanco en el exón 4 de *CD147*, y 2 con blancos en el exón 5 de *CD147*.

Luego de determinar las secuencias que guiarán al sistema CRISPR-Cas9, se diseñaron los ARN guías a ser sintetizados. Para cada secuencia blanco se diseñaron y sintetizaron un par de oligonucleótidos de simple hebra con la siguiente estructura:



Los oligonucleótidos diseñados fueron sintetizados por MacroGen (Seúl, Corea del Sur).

### 3.5. Electroforesis en gel de agarosa

Se prepararon geles de agarosa 0,7 %, 1 % o 1,5 % m/v dependiendo del rango de tamaño de los fragmentos analizados. Los geles fueron preparados en *buffer* Tris-acetato (TAE: Tris-acetato 40 mM; EDTA 1 mM pH 8), y este mismo *buffer* fue utilizado como *buffer* de corrida. En la preparación de los geles se añadió bromuro de etidio en concentración final de 0,5 µg/mL. Se utilizó 5 µL de KAPA™ Express Ladder (Kapa biosystems cat. KK6305) como marcador de peso molecular y 1X de KAPA Loading Dye (Kapa biosystems cat. KK6305) como *buffer* de carga. Las separaciones se realizaron a 5 V/cm y se visualizaron en transiluminador UV (Image Quant LAS500, General Electric).



### 3.7. Generación del vector de edición

Para generar la línea celular knock-out se utilizó el sistema CRISPR-Cas9<sup>87</sup> para inducir mutaciones en el gen *CD147* resultando en una pérdida de función (Figura 4). Para introducir el sistema CRISPR-Cas9 en la línea celular se utilizó el vector *lentiCRISPR v2* (Addgene plasmid cat. 52961)<sup>113</sup>. Este vector cuenta con sistemas de expresión para la proteína Cas9 y para el ARN guía quimérico (sgRNA) que dirige la endonucleasa al sitio de corte específico (Figura 5).

#### 3.7.1. Clonado de ARN guía al vector lentiviral

Los ARN guía diseñados anteriormente se clonaron en el vector siguiendo el protocolo de clonación establecido para este plásmido<sup>113</sup>. En primer lugar, se digirió el plásmido con la enzima de restricción BsmBI (Thermo scientific cat. ER0451). La reacción de digestión se realizó en un volumen final de 200 µL conteniendo 10 µg del plásmido, 3 unidades de enzima BsmBI, 1X de *buffer* Tango (Thermo scientific cat. BY5) y 20 mM de ditioneitol (DTT). La mezcla de reacción se incubó 16 horas en baño de agua a 37 °C, seguido de 20 minutos a 65 °C para inactivar la enzima.

Siguiendo el protocolo de clonación, se desfosforiló el plásmido digerido. La reacción de desfosforilación se realizó en 20 µL de volumen final con 1 pmol de plásmido digerido, 1X de CutSmart® Buffer (New England Biolabs cat. B7204S) y 1 unidad de Fosfatasa alcalina de intestino de ternero (CIP) (New England Biolabs cat. M0290S). La mezcla de reacción se incubó a 37 °C durante 1 hora.

La digestión con BsmBI genera un plásmido lineal de aproximadamente 12 kb y un fragmento de aproximadamente 2 kb que es descartado. Los productos de la digestión se separaron mediante electroforesis en gel de agarosa al 0,7 % siguiendo el protocolo descrito en la sección 3.5. A partir de la visualización en el transiluminador se identificó y se cortó la banda de mayor tamaño, correspondiente al plásmido digerido. A partir del fragmento de gel aislado, se purificó el plásmido utilizando el kit de extracción Zymoclean™ Gel DNA Recovery Kit (Zymo Research cat. D4001) siguiendo las indicaciones del fabricante. La concentración y pureza del plásmido se cuantificó siguiendo el protocolo descrito en la sección 3.6.

Cada par de oligonucleótidos fue hibridado con su complementario y se fosforiló usando la quinasa T4 PNK (Thermo Scientific cat. EK0031). Se preparó una mezcla

de reacción con volumen final 10  $\mu\text{L}$  conteniendo 1  $\mu\text{L}$  de cada uno del par de oligonucleótidos a una concentración de 100  $\mu\text{M}$ , 0,5 unidades de T4 PNK y 1X de T4 DNA Ligation Buffer (Thermo Scientific cat. EL0014). Las reacciones de fosforilación e hibridación se realizaron en un termociclador utilizando una temperatura inicial de 37 °C durante 30 minutos, seguido de 5 minutos a 95 °C. Luego se redujo gradualmente la temperatura 5 °C/min hasta alcanzar una temperatura final de 25 °C.

Finalmente, los oligonucleótidos hibridados se clonaron al vector digerido y purificado. Para esta reacción de ligación se preparó una dilución 1:200 de los oligonucleótidos hibridados y se prepararon tres mezclas de reacción, una para cada par de oligonucleótidos. Las mezclas de reacción se prepararon con 50 ng del plásmido digerido purificado, 1  $\mu\text{L}$  de la dilución de oligonucleótidos, 1x de T4 DNA Ligation Buffer, 1  $\mu\text{L}$  de T4 DNA Ligase (Thermo Scientific cat. EL0014) y H<sub>2</sub>O suficiente para un volumen final de 11  $\mu\text{L}$ . Las mezclas de reacción se incubaron a temperatura ambiente durante 10 minutos.

### 3.7.2. Transformación de bacterias Stbl3

Los vectores lentivirales se expandieron en bacterias *Escherichia coli* de la cepa Stbl3 (New England BioLabs cat. C3040; genotipo F' *proA+B+ lacI<sup>q</sup> Δ(lacZ)M15 zzf::Tn10 (Tet<sup>R</sup>)/ Δ(ara-leu) 7697 araD139 fhuA ΔlacX74 galK16 galE15 e14- Φ80dlacZΔM15 recA1 relA1 endA1 nupG rpsL (Str<sup>R</sup>) rph spoT1 Δ(mrr-hsdRMS-mcrBC)*). Estas bacterias poseen la mutación *recA1*, favoreciendo la integridad del vector lentiviral durante su expansión. Se realizaron cuatro transformaciones, una con cada uno de los vectores generados y una con el vector lentiCRISPR v2 sin un ARN guía.

Para la transformación, las bacterias competentes Stbl3 se descongelaron colocándolas en hielo durante 20 a 30 minutos. Se mezclaron suavemente 5  $\mu\text{L}$  del vector (aproximadamente 50 ng) con 50  $\mu\text{L}$  de células y se colocó la mezcla en hielo durante 30 minutos. Para aplicar un shock térmico a las células, se colocaron en un baño a 42 °C durante 60 segundos seguido de 2 minutos en hielo. Se agregaron 500  $\mu\text{L}$  de medio LB a las células y se incubaron 45 minutos a 37 °C con agitación. Luego se centrifugaron las células a 2100 g en centrífuga de mesa (Centrifuge 5418, Eppendorf) durante 30 segundos y se quitaron 300  $\mu\text{L}$  de sobrenadante. Se

---

resuspendió el *pellet* de células en el medio restante y se colocaron 150  $\mu$ L de las bacterias transformadas en placas de cultivo con medio sólido agar-LB con ampicilina. Como control de viabilidad de las células también se sembraron bacterias transformadas y no transformadas en medio sólido agar-LB sin ampicilina.

#### 3.7.2.1. Medios de cultivo

El medio LB se preparó con 10 g de triptona, 5 g de extracto de levadura, 10 g de NaCl y H<sub>2</sub>O suficiente para completar 950 ml. Se ajustó el pH con NaOH hasta alcanzar un pH 7 y se agregó H<sub>2</sub>O suficiente para alcanzar un volumen final de 1 L. La mezcla resultante se esterilizó mediante autoclave por 25 minutos a 121 °C.

Para el medio sólido agar-LB se utilizó un volumen final de 1 L con 10 g de triptona, 5 g de extracto de levadura, 5 g de NaCl, 1 mL NaOH 1 N y 15 g de agar. La mezcla resultante se autoclavó 25 minutos a 121 °C y se colocó en placas de cultivo. Para el medio sólido con ampicilina se adicionó 50  $\mu$ g/mL de ampicilina al medio luego de la esterilización por autoclave.

#### 3.7.3. Obtención de ADN plasmídico

La extracción de ADN plasmídico en pequeñas cantidades se realizó por el método de lisis alcalina <sup>114</sup>, a partir de 3 mL de cultivos de bacterias Stb13 transformadas con los plásmidos generados. Las bacterias inicialmente se cultivaron en medio sólido agar-LB a 37 °C durante toda la noche y a partir de este cultivo se seleccionaron colonias de bacterias transformadas con los plásmidos (portadores de un gen de resistencia a ampicilina). Usando una punta de pipeta esterilizada se transfirieron las colonias seleccionadas a tubos de cultivo conteniendo 3 mL de medio LB líquido con ampicilina. Los tubos de cultivo se incubaron toda la noche a 37 °C con agitación.

A partir del cultivo en medio líquido se transfirieron 150  $\mu$ L del cultivo a un segundo tubo y se centrifugó a 2100 g en centrífuga de mesa (Centrifuge 5418, Eppendorf) durante 5 minutos. Se descartó el sobrenadante y se resuspendió el *pellet* de células en *buffer* TENS (10 mM Tris-HCl pH 8,0; EDTA 1 M; NaOH 0,1 N; SDS 0,5 % v/v) mezclando 5 veces por inversión. Este paso rompe la membrana celular, desnaturaliza el ADN e inhibe la función de DNAsas mediante la quelación de los cationes metálicos necesarios para su funcionamiento. Luego se agregó 150  $\mu$ L de NaAc 3 M, pH 5,5. La adición de NaAc causa la precipitación del ADN cromosomal y



---

de complejos proteína-SDS mientras que el ADN plasmídico permanece en suspensión, permitiendo separarlo mediante centrifugación. Finalmente, la mezcla se centrifugó a 12000 g en centrífuga de mesa durante 5 minutos y se transfirieron 400  $\mu$ L del sobrenadante a un tubo nuevo. Para mejorar la pureza del plásmido extraído se siguió a la extracción con un paso de purificación mediante fenol-cloroformo. Se agregó un volumen (400  $\mu$ L) de una mezcla fenol-cloroformo en proporción 1:1 al extracto de ADN plasmídico obtenido en el paso anterior. Se mezcló por *vortex* 10 segundos y se centrifugó 15 segundos a 12000 g en centrífuga de mesa para alcanzar la separación de fases. Se transfirió a un tubo nuevo la fase acuosa, que contiene el ADN, y se agregó 0,1 volúmenes (40  $\mu$ L) de NaAc 3 M. Luego, para precipitar el ADN, se agregó 2 volúmenes (800  $\mu$ L) de EtOH puro, frío, y se dejó en hielo 5 minutos. Se centrifugó 5 minutos a 12000 g en centrífuga de mesa (Centrifuge 5418, Eppendorf) y se descartó el sobrenadante. El *pellet* restante se lavó dos veces con EtOH al 70%, centrifugando y descartando el sobrenadante entre cada lavado. Finalmente, el *pellet* conteniendo el ADN plasmídico purificado se dejó secar a aire durante 10 minutos y se resuspendió en 20  $\mu$ L de *buffer* TE (10 mM Tris-HCl, pH 8,0; 1mM EDTA) con RNasa. La concentración y pureza del plásmido se cuantificó siguiendo el protocolo descrito en la sección 3.6.

La identidad del plásmido extraído se verificó mediante PCR. Se prepararon mezclas de reacción de PCR para cada extracto utilizando los cebadores hU6-F y Cas9-R (GAGGGCCTATTTCCCATGATT y CTCTTTGGTGCTGGTGTACCT) que hibridan en el promotor U6 y en el gen Cas9 presentes en el vector lentiCRISPR v2 (Figura 5). Las mezclas de PCR se realizaron en 20  $\mu$ L de volumen final conteniendo 0,5  $\mu$ M de cada cebador, 1X de mezcla de reacción MangoMix™ (Meridian Bioscience ® cat. BIO-25033) y 10 ng de cada plásmido purificado. Previo calentamiento de la reacción a 95 °C durante 5 minutos, se realizaron 35 ciclos de amplificación en los cuales la temperatura alternó en tres etapas sucesivas: 30 segundos a 95 °C, seguido de 30 segundos a 63 °C y terminando el ciclo con 30 segundos a 72 °C. Finalmente, se dejó la reacción a 72 °C durante 7 minutos para asegurar la polimerización completa de todos los fragmentos. Los productos de PCR se analizaron por electroforesis en gel de agarosa 1 % siguiendo el protocolo descrito en la sección 3.5.

---

Para la extracción de ADN plasmídico en grandes cantidades se prepararon cultivos de las bacterias transformadas seleccionadas previamente. Se cultivaron bacterias en tubos de cultivo con 100 mL de medio LB con ampicilina (50 µg/mL) y se incubaron toda la noche a 37 °C con agitación. La extracción del plásmido se realizó utilizando el kit Plasmid Maxi Kit (QIAGEN cat. 12163) siguiendo las instrucciones del fabricante del kit. Se cuantificó la concentración y pureza del plásmido siguiendo el protocolo descrito en la sección 3.6. La integridad del plásmido extraído se verificó mediante electroforesis en gel de agarosa 1 % siguiendo el protocolo descrito en la sección 3.5, utilizando 500 ng de cada extracción.

Se realizó PCR para verificar la presencia en los plásmidos de los ARN guías clonados en los vectores. Para esto se utilizó como cebador reverso a los oligonucleótidos que se diseñaron para funcionar como ARN guía. Se prepararon mezclas de reacción de PCR para cada extracto utilizando el cebador hU6-F como cebador directo (GAGGGCCTATTTCCCATGATT). Las mezclas de PCR se realizaron en 20 µL de volumen final conteniendo 0,5 µM del cebador hU6-F, 1X de mezcla de reacción MangoMix™ (Meridian Bioscience ® cat. BIO-25033) y 10 ng de cada plásmido purificado. Adicionalmente en cada mezcla se incluyó como cebador reverso 0,5 µM del oligonucleótido correspondiente al ARN guía clonado en el plásmido analizado. Previo calentamiento de la reacción a 95 °C durante 5 minutos, se realizaron 35 ciclos de amplificación en los cuales la temperatura alternó en tres etapas sucesivas: 30 segundos a 95 °C, seguido de 30 segundos a 65 °C y terminando el ciclo con 30 segundos a 72 °C. Finalmente, se dejó la reacción a 72 °C durante 7 minutos para asegurar la polimerización completa de todos los fragmentos. Los productos de PCR se analizaron por electroforesis en gel de agarosa 1 % siguiendo el protocolo descrito en la sección 3.5.

### 3.8. Generación de la línea celular *knock-out*

La línea celular *knock-out* CD147 <sup>-/-</sup> se generó a partir de la línea celular prostática PC3 (ATCC® CRL-1435™). Esta línea celular proviene de un hombre caucásico de 62 años y es derivada de una metástasis en hueso de adenocarcinoma prostático grado IV <sup>115</sup>. Las células fueron transfectadas con los vectores lentiCRISPR v2 generados previamente y se usó la resistencia a puromicina otorgada por el vector como marcador de selección.

### 3.8.1. Cultivo de células PC3

En todo momento se trabajó en condiciones de esterilidad en cámara de flujo laminar. Las células PC3 usadas en este trabajo se cultivaron en medio RPMI 1640 (Gibco cat. 11-100-1G) suplementado con 10 % v/v suero fetal bovino (Capricorn Scientific. cat. FBS-11A), 100 U/mL de penicilina y 100 µg/mL de estreptomicina (Gibco cat. 15140122). Las células se mantuvieron a 37 °C en una atmósfera con 5 % CO<sub>2</sub> y el medio se cambió cada 48 horas. Las células se observaron diariamente bajo microscopio invertido para monitorear el crecimiento y asegurar que el cultivo esté libre de contaminación.

Para el pasaje de células, se descartó el medio de cultivo y se lavó el remanente de medio con 1-3 mL de *buffer* fosfato salino (PBS). Se removió el PBS y se agregó 0,5-1 mL de TrypLE™ Express Enzyme (Gibco cat. 12604021) para desprender las células de la placa de cultivo. Se incubó a 37 °C para facilitar el desprendimiento. Se incubó 5 minutos o hasta observar desprendimiento de 90 % de las células. Se agregó dos volúmenes de medio RPMI para inactivar el TrypLE por dilución y se transfirió la suspensión de células a placa de cultivo nueva con medio RPMI suficiente para el tamaño de la placa.

Para preservar los cultivos celulares a largo plazo, estos se congelaron en nitrógeno líquido. Para esto, se preparó un medio de criopreservación con medio RPMI suplementado con 10 % v/v FBS y 10 % v/v de dimetilsulfóxido (DMSO). Se separaron las células de la placa de cultivo usando 0,5-1 mL de TrypLE al igual que para el pasaje de células y se agregó dos volúmenes de medio RPMI. Se centrifugó la suspensión de células a 5000 g (Centrífuga Sorvall ST16R, Thermo Scientific) durante 5 minutos y se descartó el sobrenadante. El *pellet* de células se resuspendió en el medio de criopreservación y se transfirió a un criovial (Greiner Bio-one International cat. 126263). El criovial se envolvió en múltiples capas de papel absorbente y se dejó a 2-8 °C durante 10 minutos, seguido de 24 horas a -80 °C. Finalmente se quitó el envoltorio de papel absorbente y se transfirió el criovial a un contenedor con nitrógeno líquido para su preservación a largo plazo.

### 3.8.2. Transfección de células PC3

Para la transfección se prepararon 5 cultivos de células PC3. Inicialmente se cultivaron células en una placa de cultivo de 100 mm de diámetro hasta alcanzar 90 % de confluencia observada en microscopio. Veinticuatro horas antes de realizar la transfección, se realizaron pasajes de las células a una placa de cultivo de 6 pocillos. Se descartó el medio de cultivo de la placa de 100 mm, se lavó con 1 mL de PBS, y se agregó 1 mL de TrypLE para desprender las células de la placa. Se incubó 5 minutos o hasta observar desprendimiento de 90 % de las células y se transfirió la suspensión de células a tubo de 15 mL con 2 volúmenes (2 mL) de medio RPMI. Se contaron las células usando una cámara de Neubauer y se sembraron  $8 \times 10^5$  células en 2,5 mL de medio RPMI en cada pocillo de la placa de cultivo de 6 pocillos. Una hora previo a la transfección se descartó el medio de cultivo de cada pocillo, se lavó con PBS y se reemplazó el medio por medio Opti-MEM™ (Gibco cat. 31985062).

Los cultivos se transfectaron mediante lipofección usando el reactivo Lipofectamine™ 3000 (ThermoFisher Scientific cat. L3000015) siguiendo las instrucciones del fabricante. Se realizaron 5 reacciones de transfección, una en cada pocillo de los cultivos preparados previamente. Se realizó una reacción de transfección con cada uno de los plásmidos preparados previamente, una reacción con el vector lentiCRISPR v2 sin un ARN guía clonado y una reacción sin la adición de un plásmido. Doce horas después de la transfección se cambió el medio Opti-MEM por medio RPMI suplementado con 10 % v/v FBS.

### 3.8.3. Selección de clones

Para seleccionar las células que incorporaron el vector en la transfección, se suplementó el medio RPMI con 1 µg/mL de puromicina. El vector lentiCRISPR v2 le confiere resistencia a este antibiótico a las células que lo incorporaron, funcionando como marcador de selección. Las células se cultivaron bajo selección con puromicina hasta observar crecimiento de colonias de células resistentes. Bajo observación en microscopio, se marcó la posición de cada colonia en el pocillo y usando un *cloning ring* (Merck cat. C2059) se aislaron las colonias individuales. Los *cloning rings* son anillos de vidrio que se colocan sobre una colonia de células en una placa de cultivo. El anillo sella un área de la placa permitiendo manipular la colonia de células sin afectar al resto del cultivo. Usando una pipeta se descartó el medio de cultivo

---

contenido en el anillo. Se lavó el interior del anillo con PBS y se agregó TrypLE, se incubó a 37 °C hasta observar desprendimiento de las células y se transfirió la suspensión de células a un pocillo de una placa nueva conteniendo 3 mL de medio RPMI. Se repitió el procedimiento para todas las colonias observables y aislables de los cultivos transfectados con los vectores diseñados. Para el cultivo transfectado con el vector lentiCRISPR v2 sin ARN guía no se realizó la selección de clones y se trabajó con el conjunto de células policlonales seleccionadas con puromicina.

Los cultivos de células transfectadas monoclonales se cultivaron hasta observar confluencia. Una vez alcanzada 90 % de confluencia se hizo el pasaje a 2 placas de cultivo de 100 mm con 10 mL de medio RPMI 10 % v/v FBS sin puromicina y se cultivaron hasta confluencia.

### 3.9. Verificación de la línea celular *knock-out*

La pérdida de función de *CD147* en las líneas celulares generadas se verificó mediante *western blot*, secuenciación Sanger y RT-qPCR.

#### 3.9.1. Obtención de muestras celulares para análisis

Se prepararon *pellets* de células a partir de los cultivos de las líneas celulares generadas y a partir de cultivos de células PC3 sin modificar. Para cada cultivo se descartó el medio de cultivo y se lavó la placa con 1 mL de PBS. Se descartó el PBS del primer lavado, se agregó 1 mL de PBS a la placa y se usó un raspador de células para desprender las células de la placa. Luego se transfirió el PBS con las células a un tubo de 1,5 mL, se centrifugó a 2100 g y se descartó el sobrenadante. El procedimiento se repitió con una segunda placa de cultivo, usando 2 mL de PBS en el raspado de las células y dividiendo la suspensión de células en 2 tubos de 1,5 mL. Las células del primer raspado se conservaron a -20 °C y se usaron en el análisis por *western blot*. Uno de los tubos del segundo raspado se conservó a -20 °C y se usó para extraer ADN para secuenciar el gen editado, mientras que el tubo restante se conservó a -80 °C y se usó para extraer ARN para analizar la expresión del gen por RT-qPCR.

### 3.9.2. Análisis de expresión de proteína CD147

#### 3.9.2.1. Preparación de extractos proteicos

Para evaluar la expresión de la proteína CD147, se realizó un extracto de proteínas solubles a partir de las líneas celulares generadas. Para cada línea celular, se desprendieron las células de las placas de cultivo como se describió en 3.9.1., se resuspendió el *pellet* de células (aproximadamente  $7 \times 10^6$  células) en 50  $\mu$ L de *buffer* de lisis RIPA (50 mM Tris-HCl pH 7,4; 1 % v/v NP-40; 0,5 % v/v desoxicolato de sodio; 0,1 % v/v SDS; 150 mM NaCl; 2 mM EDTA pH 8; 5 mM NaF; 1 mM  $\text{Na}_3\text{VO}_4$ ; cOmplete™ Mini EDTA-free Protease Inhibitor Cocktail [Roche cat. 11836170001]) y se dejó actuar en hielo durante 15 minutos. Se centrifugó la suspensión de células 2100 g durante 5 minutos a 4 °C y se conservó el sobrenadante. La proteína total extraída se cuantificó en el Qubit 4.0 (Invitrogen) empleando el kit Qubit™ Protein Assay Kit (Invitrogen cat. Q33211).

#### 3.9.2.2. SDS PAGE

Las proteínas purificadas se separaron en base a su tamaño mediante electroforesis en geles desnaturizantes de poliacrilamida y SDS (SDS-PAGE) usando el método Laemmli <sup>116</sup>. El detergente SDS desnaturiza las proteínas confiriéndoles una carga negativa relativa a su largo y de esta manera permite usar electroforesis para separar las proteínas en base a su tamaño, independiente de su carga eléctrica original. El método Laemmli de SDS-PAGE consiste en usar geles discontinuos compuestos por un gel concentrador y un gel separador que difieren en su pH y en el porcentaje de acrilamida/bis-acrilamida. El gel concentrador tiene un bajo porcentaje de acrilamida permitiendo que las proteínas migren rápidamente y se concentren en una banda apretada antes de ingresar al gel separador.

El gel concentrador se preparó con 125 mM Tris-HCl pH 6,8; 0,1 % v/v SDS; 4 % v/v de solución 37,5:1 acrilamida/bisacrilamida (Bio Rad cat. 1610148); 0,05 % v/v persulfato amónico (APS) y 0,1 % v/v tetrametiletilendiamina (TEMED). El gel separador se preparó con 375 mM Tris-HCl pH 8,8; 10 % v/v de acrilamida/bisacrilamida; 0,1 % v/v SDS; 0,05 % v/v APS y 0,05 % v/v TEMED.

Las muestras de proteínas purificadas en la sección 3.9.2.1 se prepararon con un volumen igual de *buffer* de carga para SDS PAGE 2x (375 mM Tris-HCl pH 6,8; 48 %

---

v/v glicerol; 6 % v/v SDS; 0,03 % v/v azul de bromofenol; 5 % v/v  $\beta$ -mercaptoetanol) y se desnaturalizaron a 95°C durante 4 min antes de ser cargadas en el gel. La electroforesis se realizó en *buffer* de corrida para SDS-PAGE (25 mM Tris, 192 mM glicina, 0,1% v/v SDS) utilizando una cuba Mini-PROTEAN® Tetra System (Bio Rad) a un voltaje de 120 V. Se cargó 60  $\mu$ g de proteína total en cada pocillo del gel y se usó PageRuler Prestained Protein Ladder (Thermo Scientific cat. 26616) como marcador de peso molecular.

### 3.9.2.3. Western blot

Se utilizó *western blot*<sup>117</sup> para detectar la expresión de la proteína CD147 en las líneas celulares generadas. Las proteínas purificadas en la sección 3.9.2.1. se separaron en base a su tamaño mediante SDS-PAGE como se describe en la sección 3.9.2.2. y luego se transfirieron a una membrana de nitrocelulosa (Odyssey Nitrocellulose membrane, LI-COR cat. 926-31092) para su posterior detección por inmunofluorescencia. La proteína se transfirió mediante electrotransferencia<sup>118</sup> usando la cuba Mini-PROTEAN® Tetra System (Bio Rad). Se dejó transferir durante 1 h a 100 V en *buffer* de transferencia Tris-Glicina-Etanol (25 mM Tris; 192 mM Glicina; 20 % v/v Etanol; pH 8,3).

La transferencia de proteínas se verificó mediante tinción con Ponceau S<sup>119</sup>. La membrana de nitrocelulosa se sumergió en tinción Ponceau (0,1 % m/v Ponceau S; 5 % v/v ácido acético) y se incubó a temperatura ambiente 5 minutos con agitación. Se lavó con agua destilada para quitar la tinción de fondo y se fotografió. La tinción se revirtió lavando con una solución de PBS con 0,1 % v/v Tween-20 (PBS-Tween).

Previo a la detección de proteínas, es necesario bloquear la membrana de nitrocelulosa ya que esta es capaz de unir inespecíficamente con todo tipo de proteínas lo que significa que une también a los anticuerpos usados en la detección de la proteína de interés. Una forma de evitar esta unión inespecífica de los anticuerpos es primero saturar todos los sitios de unión de la membrana con una solución de proteínas inespecíficas. De esta manera, cuando se agreguen los anticuerpos solo van a poder unirse a las proteínas de interés. Para bloquear la membrana de nitrocelulosa se incubó 1 h a temperatura ambiente en una solución PBS-Tween con 5 % leche descremada.

---

La membrana se incubó con los anticuerpos primarios en PBS durante 1 h a temperatura ambiente con agitación. Luego, la membrana se incubó con anticuerpo monoclonal de ratón anti-EMMPRIN [CD147] (Sigma-Aldrich cat. MAB2623) en dilución 1 en 1000 en PBS-Tween con 1 % m/v BSA; y con anticuerpo monoclonal de conejo anti-GAPDH (Cell Signaling cat. 2118) en dilución 1 en 5000 en PBS-Tween con 1 % m/v BSA. Luego de la incubación se lavó la membrana con PBS-Tween durante 5 minutos a temperatura ambiente con agitación y se repitió el lavado 3 veces más. La membrana se incubó con los anticuerpos secundarios de la misma manera que para los anticuerpos primarios. Como anticuerpo secundario se usó el anticuerpo anti-ratón IRDye® 800CW Goat (polyclonal) Anti-Mouse IgG (LI-COR cat. 926-68071) y el anticuerpo anti-conejo IRDye® 680RD Goat (polyclonal) Anti-Rabbit IgG (LI-COR cat. 926-32210) ambos en dilución 1 en 5000 en PBS-Tween con 1 % v/v BSA. Estos anticuerpos llevan unidos un fluoróforo que emite fluorescencia al ser excitado por luz en el infrarrojo cercano. La fluorescencia emitida por los anticuerpos secundarios se detectó y fotografió mediante sistema Odyssey® Infrared Imaging System (LI-COR).

La expresión de CD147 se evaluó de manera semicuantitativa a partir de las imágenes de fluorescencia obtenidas. Utilizando el programa ImageJ <sup>120</sup> se midió la intensidad de las señales fluorescentes como la densidad de píxeles de cada banda en las fotografías. Para cada muestra se obtuvo una medida correspondiente a la señal fluorescente para CD147 y una medida correspondiente a la señal fluorescente para GAPDH. Las medidas de fluorescencia de GAPDH se utilizaron para normalizar las medidas de fluorescencia de CD147. Finalmente, se calculó la relación en la expresión normalizada de CD147 entre cada muestra y la línea control vector\_V2.

### 3.9.3. Secuenciación

A partir de las líneas celulares generadas se purificó ADN genómico para secuenciar las regiones de interés. Se desprendieron las células de las placas de cultivo como se describió en la sección 3.9.1 y, a partir de 1 mL de suspensión de células (aproximadamente  $3,5 \times 10^6$  células), se purificó ADN genómico mediante el uso del kit PureLink™ Genomic DNA Mini Kit (Invitrogen cat. K182001) siguiendo las instrucciones del fabricante.



---

La concentración y pureza del ADN purificado se cuantificó mediante espectrofotometría como se describe en la sección 3.6. y se amplificó por PCR las regiones de interés. Se amplificaron el exón 4 del gen *CD147*, el exón 5 de *CD147*, el gen *off-target* del ARN guía usado, y el gen *cas9*. Para amplificar *cas9* se prepararon mezclas de reacción de PCR usando el par de cebadores *cas9\_F* y *cas9\_R* (Tabla 7). Las mezclas de PCR se realizaron en 20  $\mu$ L de volumen final conteniendo 0,5  $\mu$ M de cada cebador, 1X de mezcla de reacción MangoMix™ (Meridian Bioscience® cat. BIO-25033) y 50 ng del ADN purificado.

Para mejorar la especificidad y sensibilidad de la reacción de PCR, los exones de *CD147* y el gen *off-target* se amplificaron usando Platinum™ Taq DNA Polymerase (Invitrogen 10966018). Las mezclas de reacción se prepararon en volumen final de 50  $\mu$ L conteniendo 1X de *buffer* de PCR sin  $MgCl_2$  (5  $\mu$ L), 1,5 mM de  $MgCl_2$  (1,5  $\mu$ L), 0,2 mM de mezcla de dNTP (1  $\mu$ L de mezcla 10 mM), 0,2  $\mu$ M de cada cebador (1  $\mu$ L de solución 10  $\mu$ M), 50 ng del ADN purificado y 2 unidades de polimerasa Platinum™ Taq (0,2  $\mu$ L). Se usó el par de cebadores *e4\_F* y *e4\_R* para amplificar el exón 4 de *CD147*; el par de cebadores *e5\_F* y *e5\_R* para amplificar el exón 5 de *CD147*; y el par de cebadores *TTF2\_F* y *TTF2\_R* para amplificar el gen *off-target* del ARN guía *sg\_E4* (Tabla 7).

Para todas las reacciones de PCR se usó el mismo programa en el termociclador, difiriendo solo en la temperatura usada para la hibridación de cebadores. Previo calentamiento de la reacción a 95 °C durante 5 minutos, se realizaron 35 ciclos de amplificación en los cuales la temperatura alternó en tres etapas sucesivas: 30 segundos a 95 °C, seguido de 30 segundos a la temperatura de hibridación específica de cada par de cebadores (Tabla 7) y terminando el ciclo con 30 segundos a 72 °C. Finalmente, se dejó la reacción a 72 °C durante 5 minutos para asegurar la polimerización completa de todos los fragmentos.

Los productos de PCR se analizaron mediante electroforesis en gel de agarosa 1 % como se describe en la sección 3.5. Se identificaron y recortaron las bandas conteniendo los fragmentos amplificados. Los fragmentos de gel conteniendo los productos de PCR se enviaron a secuenciar a Macrogen (Seúl, Corea del Sur), con la excepción del fragmento amplificado a partir del gen *cas9*.

---

Las secuencias obtenidas se analizaron mediante el *software on-line* CRISP-ID (<http://crispid.gbiomed.kuleuven.be/>)<sup>121</sup> usando los parámetros por defecto. Las secuencias obtenidas se alinearon con secuencias de referencia de los genes obtenidas de la base de datos *RefSeq* (<https://www.ncbi.nlm.nih.gov/refseq/>)<sup>109</sup>. Las secuencias se alinearon en el *software* Mega X<sup>122</sup> usando el algoritmo Clustal W<sup>123</sup>.

### 3.9.4. Análisis de expresión de ARNm

#### 3.9.4.1. Obtención de ADN copia

A partir de las líneas celulares generadas se purificó ARN para analizar la expresión de *CD147* mediante RT-PCR cuantitativo (RT-qPCR). Se desprendieron las células de las placas de cultivos como se describió en la sección 3.9.1 y, a partir de 1 mL de suspensión de células (aproximadamente  $3,5 \times 10^6$  células), se purificó ARN mediante el uso del kit Direct-zol™ RNA Miniprep Plus (Zymo Research cat. R2070) siguiendo las instrucciones del fabricante. La concentración y pureza del ARN extraído se cuantificó mediante espectrofotometría como se describe en la sección 3.6. La integridad del ARN purificado se verificó mediante electroforesis en gel de agarosa 1,5 % siguiendo el protocolo descrito en la sección 3.5 y usando 2 µg del ARN purificado.

A partir de cada extracto de ARN se generó ADN copia (ADNc) mediante retrotranscripción usando la transcriptasa reversa SuperScript™ II (Invitrogen cat. 18064-022). La retrotranscripción se realizó con 2 µg de ARN siguiendo las instrucciones del fabricante.

#### 3.9.4.2. PCR cuantitativo

Las reacciones de PCR cuantitativo (qPCR) se realizaron usando el kit SYBR Green PCR Kit (QuantiNova cat. 208052). La mezcla de reacción se preparó en un volumen final de 9 µL conteniendo 1X de SYBR Green PCR Master Mix (5 µL), 0,9 µM de cada uno del par cebadores usado (0,9 µL de solución 10 µM), y 1 µL del colorante ROX. La mezcla de reacción se colocó en un pocillo de placa para qPCR y se agregó 1 µL de ADNc o 1 µL de H<sub>2</sub>O.

Las reacciones de qPCR se realizaron en el sistema Eco™ Real-Time PCR System (Illumina cat. EC-900-1001). El sistema se programó con un paso inicial de activación de la enzima a 95 °C por 2 minutos, seguido de 40 ciclos de dos pasos: un primer

paso de desnaturalización de 5 segundos a 95 °C seguido de un segundo paso de hibridación y extensión de 10 segundos a 60 °C.

### 3.9.4.3. Evaluación de la eficiencia de amplificación de cebadores

Para poder comparar la expresión de genes distintos mediante qPCR es importante determinar la eficiencia de amplificación de los cebadores usados. La eficiencia de amplificación se define como la fracción de las moléculas de ADN que son replicadas en cada ciclo de PCR <sup>124</sup>. Si existe una diferencia en las eficiencias entre los cebadores usados para amplificar el gen de interés y los cebadores usado para amplificar el gen normalizador esa diferencia se debe tomar en cuenta en el cálculo de la expresión relativa del gen.

El ciclo de umbral ( $C_q$ ) es el ciclo de la reacción de PCR en que la fluorescencia emitida supera a un umbral de detección predeterminado. En cada ciclo de amplificación, la fluorescencia emitida por la reacción de qPCR es proporcional a la cantidad de ADN en la reacción. La cantidad de ADN amplificado en cada ciclo ( $C_n$ ) depende tanto de la cantidad de ADN inicial ( $C_0$ ) como de la eficiencia de amplificación de los cebadores ( $E$ ) y del número de ciclos de amplificación transcurridos ( $n$ ), como se muestra en la ecuación 1. Dado que, para un umbral dado el valor de  $C_n$  permanece constante, se puede reordenar la ecuación 1 para mostrar que el número de ciclos requeridos para superar este umbral ( $C_q$ ) depende linealmente del logaritmo de  $C_0$  y que la pendiente de esta línea es igual a  $-\frac{1}{\log_{10} E}$  (ecuación 2). Finalmente, a partir de esta pendiente se puede determinar la eficiencia de amplificación de los cebadores mediante la ecuación 3.

$$C_n = C_0 \times E^n \quad (\text{Ecuación 1})$$

$$C_q = \left( \frac{1}{\log_{10} E} \times \log_{10} C_n \right) - \frac{1}{\log_{10} E} \times \log_{10} C_0 \quad (\text{Ecuación 2})$$

$$E = 10^{\left( \frac{1}{\text{pendiente}} \right)} - 1 \quad (\text{Ecuación 3})$$

---

Para determinar la eficiencia de amplificación de los cebadores se generó una curva de calibración a partir de reacciones de qPCR realizadas sobre diluciones seriadas de ADNc. Mediante el uso de distintas  $C_0$ , se puede graficar los  $C_q$  medidos correspondientes a las distintas  $C_0$  y de esta manera generar una curva de calibración cuya pendiente corresponde a  $-\frac{1}{\log_{10} E}$ . Para generar la curva, se prepararon 4 diluciones de ADNc obtenido a partir de células PC3 sin edición génica. El ADNc se obtuvo como se describe en la sección 3.9.4.1 y se prepararon diluciones 1:10, 1:20, 1:40, y 1:80 del mismo. Se prepararon mezclas de reacción de qPCR como se describe en la sección 3.9.4.2, utilizando los cebadores diseñados para qPCR (qPCR1\_F, qPCR1\_R, qPCR2\_F, qPCR2\_R, qPCR3\_F, y qPCR3\_R,) y el par de cebadores para *ACTB* ( $\beta$ 1actina 1 y  $\beta$ 1actina 2) (ver Tabla 7). Para cada par de cebadores se prepararon reacciones usando como ADN molde una de las diluciones de ADNc preparadas previamente o usando agua en lugar del ADN molde. Para las reacciones realizadas con los cebadores que amplifican el ADNc del ARNm de *CD147*, se realizaron 3 réplicas técnicas de la reacción de qPCR, mientras que para el par de cebadores de *ACTB* se realizaron 2 réplicas técnicas. La reacción de qPCR se llevó a cabo como se describe en la sección 3.9.4.2. y los resultados de la reacción se visualizaron mediante el *software* EcoStudy <sup>125</sup>.

A partir de los resultados de esta reacción, se graficaron los ciclos de umbral ( $C_q$ ) medidos para cada dilución contra el logaritmo de la dilución ( $\log_{10} \text{dilución}$ ) para obtener una curva de calibración. Luego, se utilizó el *software* GraphPad Prism 8.4.2 (GraphPad Software, San Diego, California USA, [www.graphpad.com](http://www.graphpad.com)) para realizar una regresión lineal simple. Para cada curva se calculó el cuadrado del coeficiente de correlación de Pearson ( $R^2$ ) para verificar la existencia de una relación lineal entre  $C_q$  y  $\log_{10} \text{dilución}$  para cada par de cebadores. A partir de este modelo lineal se determinó la pendiente de la curva de calibración y se calculó la eficiencia (E) de cada par de cebadores utilizando la ecuación 3.

#### 3.9.4.4. Expresión relativa de *CD147*

Se realizaron reacciones de qPCR como se describe en la sección 3.9.4.2 usando el par de cebadores qPCR1\_F y qPCR1\_R o el par de cebadores  $\beta$ 1actina 1 y  $\beta$ 1actina 2. Para cada par de cebadores se prepararon 4 mezclas de reacción; 1 usando como

ADN molde el ADNc obtenido de la línea celular editada con el sg\_E4, 1 usando el ADNc obtenido de la línea celular transfectada con el vector lentiCRISPR v2 sin ARN guía, 1 usando el ADNc obtenido de la línea celular PC3 sin modificar, y 1 con la adición de agua en lugar de ADN molde. Se realizaron 3 réplicas técnicas de cada reacción de qPCR para cuantificar la incertidumbre en las medidas obtenidas.

Se utilizó el *software* EcoStudy <sup>125</sup> para visualizar los resultados y calcular la expresión de *CD147* en las líneas celulares editadas. La expresión se calculó mediante la opción de expresión relativa (RQ). Esta opción utiliza el método Pfaffl <sup>126</sup> para determinar la diferencia relativa en el nivel de expresión del gen de interés con respecto a un gen control. Este método determina cuántas veces mayor o menor es la expresión del gen en la muestra evaluada con respecto a una muestra control. Utiliza la expresión de un gen control (*ACTB* en este caso) para normalizar las medidas de expresión del gen de interés (*CD147*), tomando en cuenta la eficiencia de amplificación de cada par de cebadores. El resultado se expresa como el cambio en la expresión del gen de interés en relación al cambio en la expresión del gen control. El cálculo utilizado se muestra en la ecuación 4:

$$RQ_{muestra} = \frac{E_{qPCR1}^{(Cq_{wt} - Cq_{muestra})_{qPCR1}}}{E_{actina}^{(Cq_{wt} - Cq_{muestra})_{actina}}} \quad (\text{Ecuación 4})$$

Donde  $E_{qPCR1}$  y  $E_{actina}$  son las eficiencias calculadas para los pares de cebadores qPCR1 y  $\beta$ 1actina,  $Cq_{wt}$  y  $Cq_{muestra}$  son los promedios de los ciclos de umbral para las células PC3 sin modificar y la muestra evaluada respectivamente.

## 4. Resultados y discusión

### 4.1. Análisis de datos de expresión génica en bases de datos

Para analizar la expresión de *CD147* en cáncer se utilizó la herramienta *UCSC Xena*<sup>127</sup> y se analizaron datos de expresión provenientes de los proyectos *The Cancer Genome Atlas* (TCGA) y *Genotype-Tissue Expression* (GTEx)<sup>108</sup>. Los datos de expresión génica de esta base de datos fueron generados originalmente a partir de lecturas de secuenciación de ARN mensajero (*RNAseq*) cuantificadas por la metodología RSEM (RNA-Seq by Expectation Maximization)<sup>128</sup> y normalizadas mediante la herramienta DESeq2<sup>129</sup>. El RNA-seq es una técnica basada en tecnologías de secuenciación masiva que permite detectar y cuantificar la totalidad de ARN presente en una muestra biológica en un momento dado<sup>130</sup>. Mediante esta tecnología se puede determinar y comparar la expresión de distintos genes entre distintas muestras.

A partir de este conjunto de datos se utilizaron solo las muestras con datos de expresión de *CD147* en tejido normal no tumoral, en tumor primario o en metástasis. En la Tabla 3 se detalla la distribución de las muestras por sitio primario.

Sitio Primario	Tejido Normal (GTEx)	Tejido Normal (TCGA)	Tumor Primario (TCGA)	Metástasis (TCGA)
Bazo	100	0	0	0
Cabeza y cuello	0	44	518	2
Cerebro	1148	5	662	0
Colon	307	41	287	1
Corazón	376	0	0	0
Cuello uterino	10	3	303	2
Ducto biliar	0	9	36	0
Endometrio	0	23	180	0
Esófago	652	13	181	1
Estómago	174	36	413	0
Glándula Adrenal	126	0	77	0
Glándula salival	55	0	0	0
Glándula tiroides	279	59	504	8
Hígado	110	50	369	0
Intestino delgado	92	0	0	0
Mama	178	113	1092	7
Músculo	395	0	0	0
Nervios	278	0	0	0
Ojo	0	0	79	0

Ovario	88	0	419	0
Páncreas	167	4	178	1
Paraganglia	0	3	177	2
Piel	555	1	102	366
Pituitaria	107	0	0	0
Próstata	100	51	494	1
Pulmón	288	109	1011	0
Recto	0	10	92	0
Revestimiento de cavidades	0	0	87	0
Riñón	28	140	884	1
Sangre	337	0	0	0
Tejido Adiposo	512	0	0	0
Tejido blando, hueso	0	2	258	1
Tejido linfático	0	0	47	0
Testículo	165	0	148	0
Timo	0	2	119	0
Trompa uterina	5	0	0	0
Útero	78	0	57	0
Vagina	85	0	0	0
Vaso sanguíneo	604	0	0	0
Vejiga	9	19	407	0

**Tabla 3: Conteo de muestras con datos de expresión de *CD147* contenidas en los proyectos GTEx y TCGA.**

Se muestra para cada sitio el número de muestras provenientes de tejido normal no tumoral (Tejido Normal (GTEx) y Tejido Normal (TCGA)), tumor primario y metástasis que contienen datos de expresión génica de *CD147*. Las muestras de tejido normal del proyecto GTEx son muestras de tejido no tumoral provenientes de individuos sin un diagnóstico de cáncer, mientras que las del proyecto TCGA fueron tomadas del tejido sólido no tumoral adyacentes al tumor sólido principal.

Inicialmente se utilizó el conjunto de datos TCGA Pan-Cancer Atlas (PANCAN) que contiene muestras provenientes de los 33 tipos de tumores más prevalentes<sup>131</sup>. Este conjunto de datos contiene, para algunos pacientes, la muestra del tumor correspondiente y una muestra del tejido normal no tumoral adyacente al tumor sólido. Esto permite utilizar muestras pareadas (provenientes de un mismo paciente) para analizar diferencias de expresión entre el fenotipo tumoral y el fenotipo normal.

Se comparó la expresión media de *CD147* en este conjunto para muestras pareadas entre tumor y tejido normal ( $n = 711$ ) o entre tumor y metástasis ( $n = 28$ ) (Figura 6A) para luego realizarse una prueba *t* de Welch para evaluar si las medias de las poblaciones muestreadas son significativamente distintas. La expresión media de *CD147* en muestras pareadas es significativamente mayor ( $t = -6,804$ ,  $p < 0.001$ ) en tumor primario ( $13,6 \log_2[\text{valor normalizado} + 1]$ ) que en tejido normal ( $13,3 \log_2[\text{valor}$

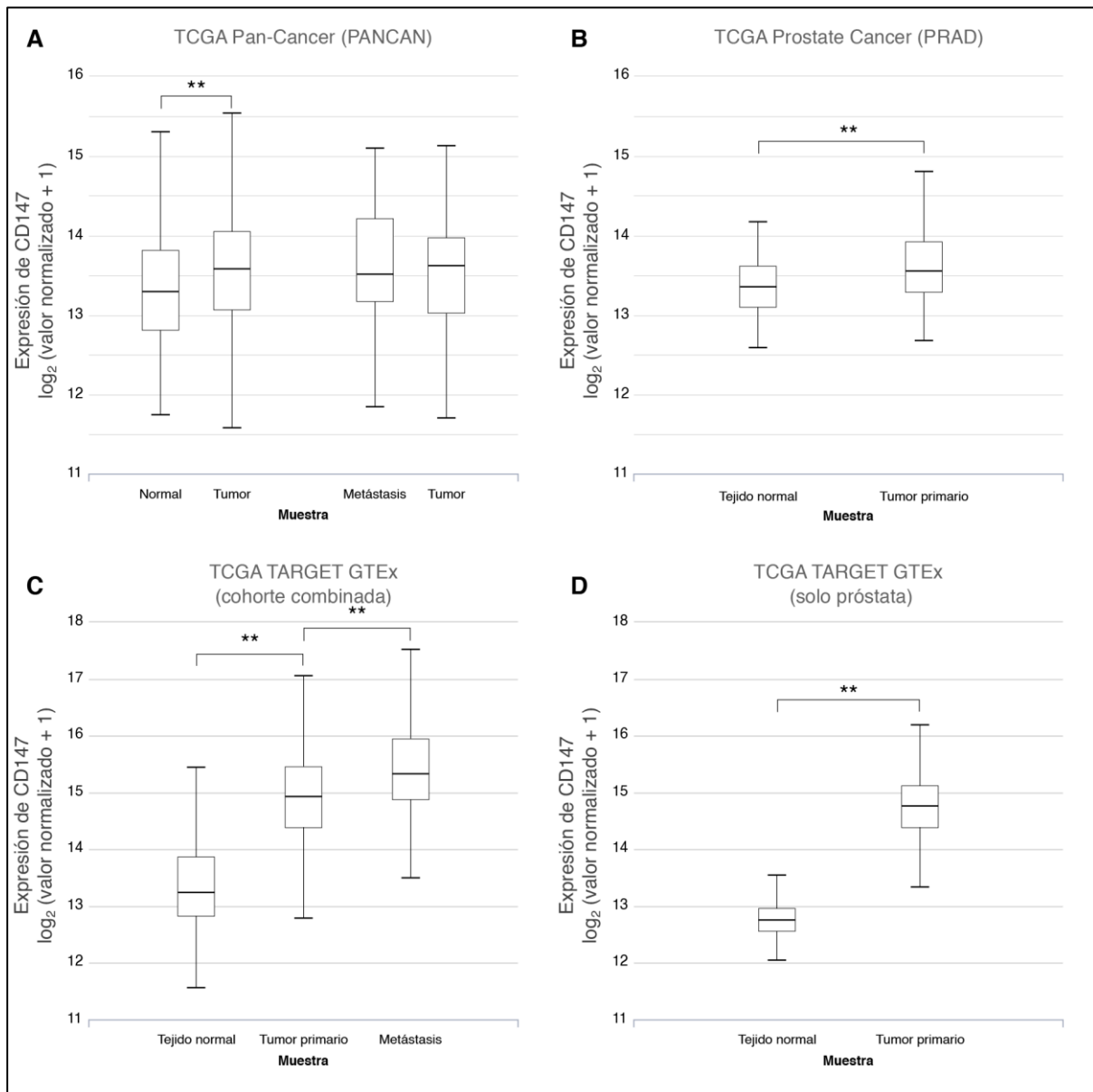
---

normalizado + 1]). En este set de datos se observa una diferencia no significativa entre la expresión de *CD147* en tejido metastásico ( $13,5 \log_2[\text{valor normalizado} + 1]$ ) y la expresión en tejido tumoral primario ( $13,6 \log_2[\text{valor normalizado} + 1]$ ). En términos generales, estos resultados sugieren que la expresión de *CD147* aumenta con el desarrollo inicial del tumor y apoya los trabajos que proponen un papel oncogénico de *CD147*<sup>74–80</sup>.

Para evaluar la expresión de *CD147* en cáncer de próstata se repitió el análisis utilizando el conjunto de datos pareados provenientes de pacientes con cáncer de próstata (TCGA Prostate Cancer,  $n = 52$ ). Debido a que solo hay una muestra de metástasis de próstata en esta base de datos, no se pudo realizar un análisis de la expresión en metástasis, por lo que se descartó esta muestra y se prosiguió a analizar únicamente tejido normal y tumor primario (Figura 6B). Los resultados de este análisis son congruentes con lo observado en el análisis anterior y muestran que en próstata la expresión media de *CD147* es significativamente mayor en tumores primarios ( $13,6 \log_2[\text{valor normalizado} + 1]$ ) con respecto a la expresión en el tejido prostático normal del mismo paciente ( $13,4 \log_2[\text{valor normalizado} + 1]$ , prueba t de Welch  $t = -2,894$ ,  $p = 0,005$ ).

El uso de muestras pareadas permite comparar muestras de tumores con muestras de tejido normal no tumoral provenientes de un mismo individuo y obtenidas usando la misma metodología. Esto permite asociar diferencias en la expresión génica con diferencias entre los fenotipos tumoral y normal y no a diferencias entre patrones de expresión de base que existen entre individuos. Sin embargo, existen dudas sobre si estas muestras normales son realmente “normales” o si contienen cambios en sus patrones de expresión causados por la presencia de un tumor cercano<sup>132,133</sup>. El proyecto *Genotype-Tissue Expression* (GTEx) contiene datos de expresión para distintos tejidos normales no tumorales provenientes de individuos sin un diagnóstico de cáncer. Cuando se comparan las muestras de tejido “normal” provenientes del proyecto TCGA con las muestras de tejido normal provenientes de GTEx, se observa que para todos los sitios estudiados la expresión media de *CD147* en los tejidos normales de pacientes con cáncer es mayor que la expresión media de *CD147* en los tejidos de individuos sanos (Tabla 4).





**Figura 6: Expresión de *CD147* en tejido normal, tumores primarios y metástasis.**

Datos de expresión obtenidos del *UCSC Toil RNAseq Recompute Compendium*<sup>108</sup> en el portal de *Xena Browser*<sup>127</sup>. La expresión de *CD147* se muestra en escala logarítmica convertida mediante el cálculo  $\log_2$  (valor normalizado + 1). Valor p evaluado mediante prueba t de Welch. **A.** Expresión en muestras pareadas de tejido normal, tumor primario y metástasis de la cohorte TCGA Pan-Cancer. Izquierda: muestras pareadas de tejido normal y tumor ( $n = 711$ ,  $t = -6,804$  y  $p < 0.001$ ). Derecha: muestras pareadas de tumor y metástasis ( $n = 28$ ,  $t = 0,4710$ ,  $p = 0,6395$ ). **B.** Expresión en muestras pareadas de tejido normal y tumor primario de próstata de la cohorte TCGA Prostate Cancer ( $n = 52$ ,  $t = -2,894$  y  $p = 0,005$ ). **C.** Expresión en muestras de tumor primario ( $n = 9181$ ), metástasis ( $n = 393$ ) y tejido normal ( $n = 7408$ ) en la cohorte combinada TCGA TARGET GTEx ( $n_{\text{total}} = 16892$ ). Prueba t de Welch: tejido normal versus tumor primario  $t = -94,71$  y  $p < 0.001$  y tumor primario versus metástasis  $t = -10,08$  y  $p < 0.001$ . **D.** Expresión en muestras de tumor primario ( $n = 494$ ) y tejido normal ( $n = 100$ ) en próstata de la cohorte combinada TCGA TARGET GTEx ( $n_{\text{total}} = 594$ ). Prueba t de Welch  $t = -43.26$  y  $p < 0.001$ . \*\*  $p < 0,01$ .

Sitio	Tejido Normal (GTEX)	n	Tejido Normal (TCGA)	n	Diferencia
Cerebro	13,8	1148	14,4	5	0,6 **
Colon	13,8	307	16,2	41	2,4 **
Esófago	12,9	652	15,0	13	2,1 **
Estómago	13,8	174	15,1	36	1,3 **
Hígado	12,5	110	13,2	50	0,7 **
Mama	12,9	178	13,6	113	0,7 **
Páncreas	12,7	167	14,9	4	2,2 *
Próstata	12,8	100	14,6	51	1,8 **
Pulmón	12,9	288	14,8	109	1,9 **
Riñón	14,0	28	14,8	140	0,8 *
Vejiga	13,3	9	15,1	19	1,8 **

**Tabla 4: Expresión de *CD147* en muestras de tejido normal no tumoral de los proyectos TCGA y GTEX.**

Se muestra la mediana de la expresión de *CD147* ( $\log_2[\text{valor normalizado} + 1]$ ) en muestras de tejido normal no tumoral para todos los tejidos para los cuales se cuenta con muestras, tanto en el proyecto GTEX como TCGA. Para todos los tejidos la expresión de *CD147* es significativamente mayor en las muestras provenientes del proyecto TCGA. La significación estadística fue evaluada mediante prueba t de Welch. \*  $p < 0,05$  \*\*  $p < 0,01$ .

A pesar de las diferencias encontradas, los datos de expresión de individuos sanos provenientes del proyecto GTEX son adecuados para usar como referencia para comparar la expresión génica en el tumor con respecto a la expresión en el tejido no tumoral, como lo muestran estudios realizados previamente<sup>134</sup>. Por esta razón, y para aumentar el poder estadístico de estos análisis, se amplió el número de datos analizados utilizando *UCSC Xena* con el conjunto de datos combinados de los proyectos TCGA y GTEX. A partir de este conjunto de datos se seleccionaron solo las muestras que contengan datos de expresión de *CD147* en tejido normal no tumoral ( $n = 7408$ ), en tumor primario ( $n = 9181$ ) o en metástasis ( $n = 393$ ). Se comparó la expresión media de *CD147* entre tejido normal no tumoral y tumor primario ( $n_{\text{total}} = 16589$ ), así como entre tumor primario y metástasis ( $n_{\text{total}} = 9574$ ).

Los resultados se muestran en la Figura 6C. Se observa un aumento significativo en la expresión de *CD147* en muestras de tejido tumoral primario (expresión media de *CD147* =  $14,9 \log_2[\text{valor normalizado} + 1]$ ) cuando se compara con su expresión en tejido no tumoral ( $13,2 \log_2[\text{valor normalizado} + 1]$ ), prueba t de Welch =  $-94,71$  y  $p < 0.001$ ). Más aún, se observa un aumento mayor de la expresión del gen *CD147* en

---

muestras metastásicas ( $15,3 \log_2[\text{valor normalizado} + 1]$ ) y este cambio es significativo con respecto al valor obtenido con muestras de tumores primarios (prueba t de Welch = -10,08,  $p < 0.001$ ). A diferencia de los resultados anteriores, estos resultados sugieren que no solo la expresión de *CD147* aumenta con el desarrollo inicial del tumor, sino que también existe un aumento desde el tejido tumoral al nicho metastásico, apoyando el papel pro-metastásico de *CD147*<sup>74-80</sup>.

Para evaluar la expresión de *CD147* en cáncer de próstata se repitió el análisis utilizando el mismo conjunto de datos seleccionando solo las muestras procedentes de tejido prostático ( $n_{\text{total}} = 594$ ). Al igual que en el primer análisis, debido a que solo hay una muestra de metástasis de próstata en esta base de datos, se descartó esta muestra y se prosiguió a analizar únicamente tejido normal ( $n = 100$ ) y tumor primario ( $n = 494$ ) (Figura 6D). Los resultados muestran que la expresión media de *CD147* expresada en  $\log_2(\text{valor normalizado} + 1)$  es igual a 12,8 en tejido normal y 14,8 en tumor primario, observándose una diferencia significativa entre las medias de estas poblaciones ( $t = -43.26$  y  $p < 0.001$ ). Estos resultados apoyan lo observado anteriormente con las muestras pareadas (Figura 6B) y muestran un aumento significativo de la expresión de *CD147* en muestras de tumor primario prostático cuando se compara con muestras de tejido prostático normal.

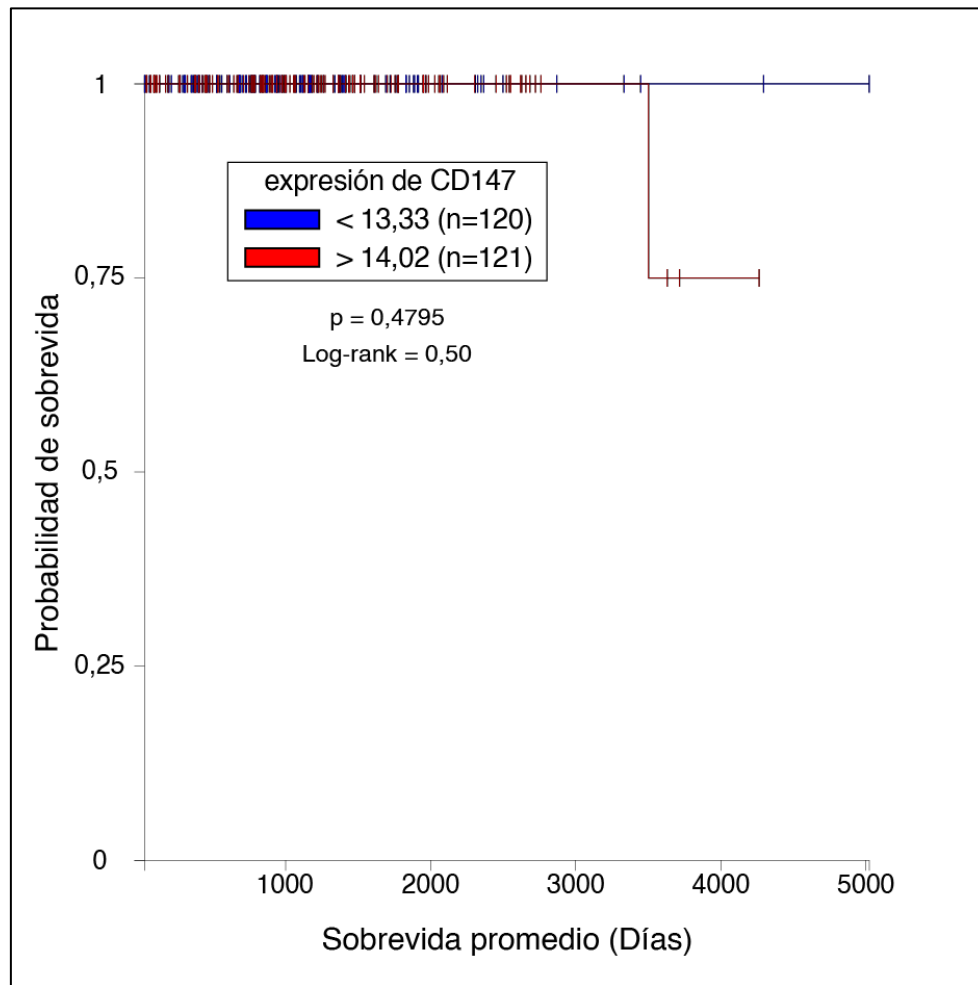
Sin embargo, es importante mencionar que las muestras empleadas en estos últimos análisis no son pareadas. Los datos de expresión de tejido normal provienen del proyecto GTEx mientras que los datos de expresión de tumores provienen del proyecto TCGA. Las metodologías usadas para obtener y procesar los datos son distintas entre los proyectos lo que puede resultar en *batch effects* (variabilidad generada por diferencias en la metodología experimental). Si bien, los datos de expresión usados en este análisis provienen del *UCSC Toil RNAseq Recompute Compendium*<sup>108</sup> y fueron reprocesados a partir de los datos de expresión de TCGA y GTEx para eliminar los *batch effects* generados por el procesamiento computacional, no es posible eliminar los *batch effects* producidos por la obtención de las lecturas de RNAseq.

Para estudiar el efecto del nivel de expresión de *CD147* en tumores primarios en la supervivencia de pacientes con cáncer de próstata se generaron curvas Kaplan Meier

---

utilizando el portal de *UCSC Xena* (Figura 7). Se comparó la sobrevida de pacientes con tumores prostáticos con niveles altos de expresión de *CD147* (> 75 %, en el cuartil superior) con la sobrevida de pacientes con niveles bajos de expresión de *CD147* (< 25 %, en el cuartil inferior). Se evaluó la significación estadística mediante la prueba logrank. Si bien la curva de sobrevida de los pacientes que expresan altos niveles de *CD147* en el tumor prostático se encuentra por debajo de la curva de los pacientes que presentan bajos niveles, esta diferencia no es significativa ( $p = 0,480$  y Log-rank = 0,500).

En su conjunto, los resultados presentados muestran que la expresión de *CD147* está elevada en cáncer de próstata en comparación con la expresión en tejido normal. Lo mismo se observa cuando se compara la expresión de *CD147* en cáncer en general. Estos resultados están alineados con los trabajos que muestran a *CD147* como un factor que contribuye con el crecimiento tumoral y el desarrollo de metástasis<sup>66-71</sup>. Por otro lado, el análisis de sobrevida muestra que, al contrario de lo esperado, el nivel de expresión de *CD147* en el tumor prostático parece no afectar la sobrevida del paciente. Dado que la metástasis es la principal causa de muerte por cáncer de próstata, sería interesante comparar la sobrevida de los pacientes con tumores prostáticos metastásicos con altos versus bajos niveles de expresión de *CD147*. Lamentablemente, no se cuenta con la información necesaria en este conjunto de datos.



**Figura 7: Curva de supervivencia Kaplan Meier para expresión de *CD147* en cáncer de próstata**

Diferencia en supervivencia entre muestras de tumor primario de expresión alta y expresión baja de *CD147*. Datos de expresión obtenidos del *UCSC Toil RNAseq Reanalyze Compendium*<sup>108</sup> en el portal de *UCSC Xena*<sup>127</sup>. La expresión de *CD147* se muestra en escala logarítmica convertida mediante el cálculo  $\log_2$  (valor normalizado + 1). Significación estadística evaluada mediante prueba logrank. Supervivencia en muestras de la cohorte TCGA Prostate Cancer. No hay una diferencia significativa entre las curvas de supervivencia ( $p = 0,480$  y Log-rank = 0,500).

#### 4.2. Diseño de ARN guías para la edición del gen *CD147*

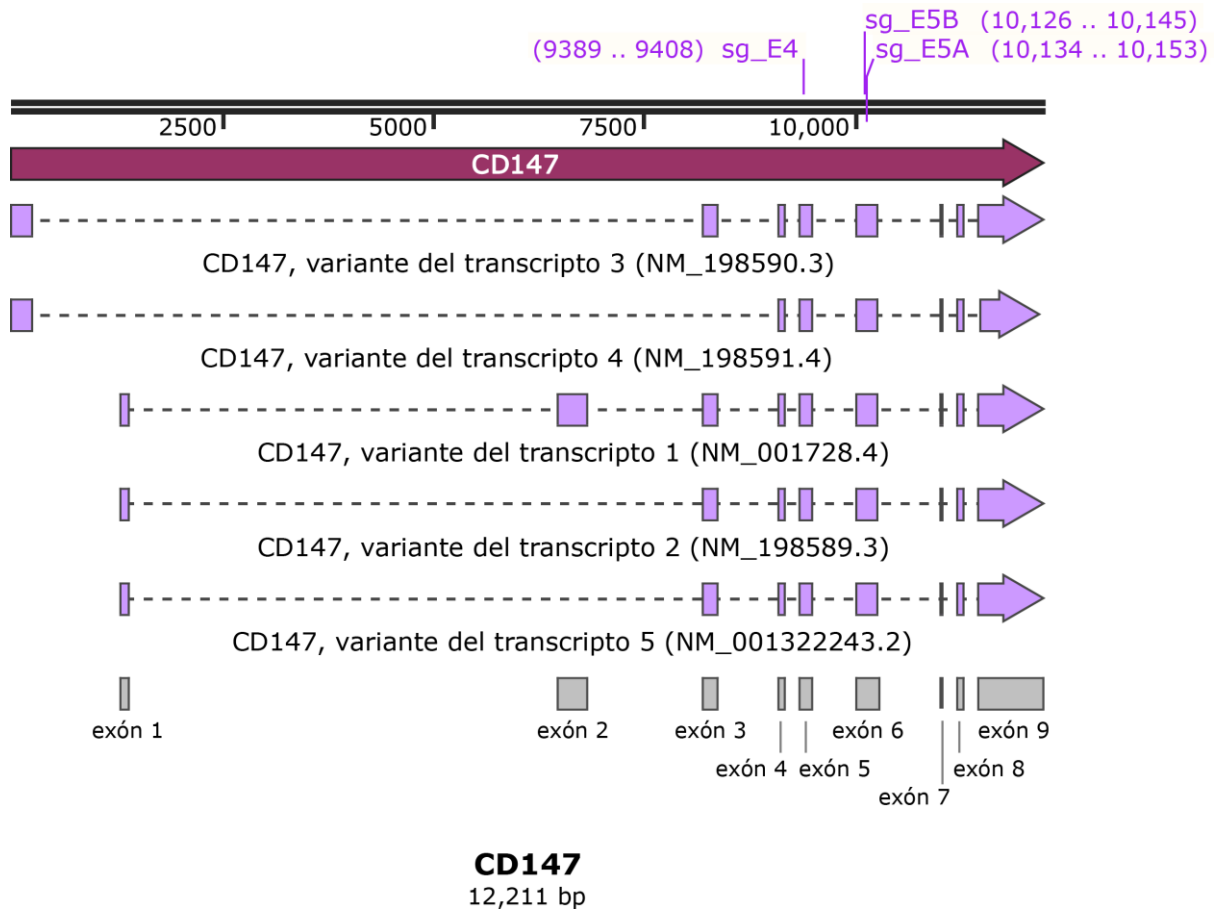
Una de las herramientas más ampliamente utilizadas para estudiar la función de una proteína en una célula es anular su función mediante la delección o silenciamiento del gen que la expresa. Para intentar comprender la función de la proteína *CD147* en el cáncer de próstata se generó una línea celular *knock-out* para el gen *CD147* (Ensembl: ENSG00000172270).

---

La línea celular *knock-out* se generó mediante edición genómica utilizando el sistema CRISPR-Cas9<sup>87</sup>. El método empleado requiere el clonado en el vector lentiCRISPR V2 (Addgene plasmid cat. 52961; Figura 5)<sup>112</sup> de un oligonucleótido de 20 nt de largo que contiene la secuencia objetivo del sistema CRISPR-Cas9. Una vez clonado en el vector, este oligonucleótido se transcribe junto a una secuencia de andamiaje produciendo una molécula única de ARN quimérico llamado ARN guía (single-guide RNA o sgRNA)<sup>88</sup> capaz de dirigir a la endonucleasa Cas9 a la región de ADN que se desea editar.

La endonucleasa Cas9 reconoce su secuencia objetivo mediante apareamiento de bases Watson-Crick en un trecho de 20 nt directamente *upstream* de un motivo 5'-NGG. Sin embargo, el reconocimiento de la secuencia objetivo por Cas9 tolera bases desapareadas entre el ARN guía y el ADN objetivo, siendo esta tolerancia dependiente del número, posición e identidad de las bases desapareadas<sup>135</sup>. Esta tolerancia puede llevar a que Cas9 reconozca secuencias distintas de la secuencia objetivo, resultando en que se edite un gen diferente (llamado gen *off-target*).

Para diseñar el oligonucleótido que forma la secuencia guía del sgRNA se utilizó la herramienta web *CRISPR design tool* (<http://crispr.mit.edu><sup>111</sup>). Esta herramienta se utilizó para encontrar ARN guías que tengan alta especificidad por un sitio en el gen *CD147* y para buscar potenciales sitios *off-target* de estos ARN guía. Para el diseño de los ARN guías se utilizó la secuencia de referencia NCBI NM\_198589.3 que corresponde a la variante 2 del ARN mensajero (ARNm) de *CD147* y codifica para la isoforma 2 de la proteína. En particular, se buscó ARN guías que tengan como blanco el exón 4 o el exón 5 de este ARNm debido a que estos se encuentran en todas las variantes de *CD147* (Figura 8). Se seleccionaron 3 ARN guía, uno cuyo blanco se encuentra en el exón 4 (llamado sg\_E4) cuyo *hit score* es de 93 y dos ARN guías cuyos blancos se encuentran en el exón 5 (llamados sg\_E5A y sg\_E5B) con *hit score* 96 y 94 respectivamente (Tabla 5).



**Figura 8: Secuencias de referencia para las variantes de ARN mensajero de *CD147*.**

Se muestran cinco variantes de ARN mensajero de *CD147* que codifican para las cuatro isoformas de *CD147*. Las variantes 2 y 5 codifican la isoforma 2 y las variantes 1, 3 y 4 codifican para las isoformas 1, 3 y 4 respectivamente. Se indican los exones de las isoformas 1 y 2. Los exones 4 a 9 son compartidos por todas las isoformas. Se indican los sitios de reconocimiento de los ARN guía diseñados (sg\_E4, sg\_E5A y sg\_E5B).

sgRNA	Exón	Secuencia guía	<i>hit score</i>
sg_E4	4	GTCGTCAGAACACATCAACG	93
sg_E5A	5	TGGTGCCGTTGCACCGGTAC	96
sg_E5B	5	CCCGGCCAGTACCGGTGCAA	94

**Tabla 5: ARN guías diseñados para la edición genómica del gen *CD147* utilizando un sistema CRISPR-Cas9.**

Los oligonucleótidos fueron diseñados con *CRISPR design tool* (<http://crispr.mit.edu> [27]). Se indica el número del exón de *CD147* que tiene como blanco cada guía diseñado, la secuencia guía de 20 nt complementaria a la secuencia objetivo y el *hit score*. El valor del *hit-score* corresponde aproximadamente a la probabilidad (en porcentaje) de que el ARN guía se una a su secuencia objetivo y no a una secuencia *off-target*.

La herramienta *CRISPR design tool* provee una lista de los potenciales genes *off-targets* de los ARN guías seleccionados. Utilizando esta lista se encontró un posible gen *off-target* para sg\_E4 y cuatro para cada uno de sg\_E5A y sg\_E5B (Tabla 6). En todos los casos las secuencias de ARN guía presentan un *hit score*  $\geq 93$  para la secuencia blanco en el gen *CD147* (Tabla 5) y un *hit score*  $\leq 0,5$  para las secuencias *off-targets*, indicando que son buenos candidatos para ser utilizados en el sistema de edición. Sin embargo, cabe mencionar que el sg\_E5A presenta como *off-target* el gen *AP3D1* ubicado en la misma región cromosómica que el gen *CD147* (cromosoma 19 p13.3). Esta situación no es la óptima para la edición genómica<sup>136</sup> pero se continuó con la utilización del sg\_E5A en paralelo con los otros dos sgRNA diseñados (sg\_E4 y sg\_E5B) que presentan mejores propiedades.

sgRNA	Secuencia <i>off-target</i>	<i>hit score</i>	Gen <i>off-target</i>	Ensembl	locus
sg_E4	GTTGTCACAGCACATCAAGGAGG	0.5	<i>TTF2</i>	ENSG00000116830	1p13.1
sg_E5A	TGTTGCCTTTGCAGCGGTACCAG	0.4	<i>AP3D1</i>	ENSG00000065000	19p13.3
	TGGAGGCCTTGCACCTGTACCGG	0.2	<i>CELSR2</i>	ENSG00000143126	1p13.3
	TGGCGCAGTTGCGCCGCTACCGG	0.2	<i>PPP1R3F</i>	ENSG00000049769	Xp11.23
	TGGGGCGGTGGCAGCGGTACCAG	0.1	<i>MRGPRX1</i>	ENSG00000170255	11p15.1; 11
sg_E5B	CCCGTCAGGTACCGGTGCACCAG	0.5	<i>AGAP1</i>	ENSG00000157985	2q37.2
	TCCAGCCAGCACCGATGCAAAGG	0.4	<i>FHAD1</i>	ENSG00000142621	1p36.21
	CCCTGCCAGCACCGCTGCATGAG	0.2	<i>RIPK4</i>	ENSG00000183421	21q22.3
	CCTGACCAGTACAGCTGCAAGGG	0.2	<i>MAPKBP1</i>	ENSG00000137802	15q15.1

**Tabla 6: Genes potencialmente *off-target* de los ARN guía diseñados.**

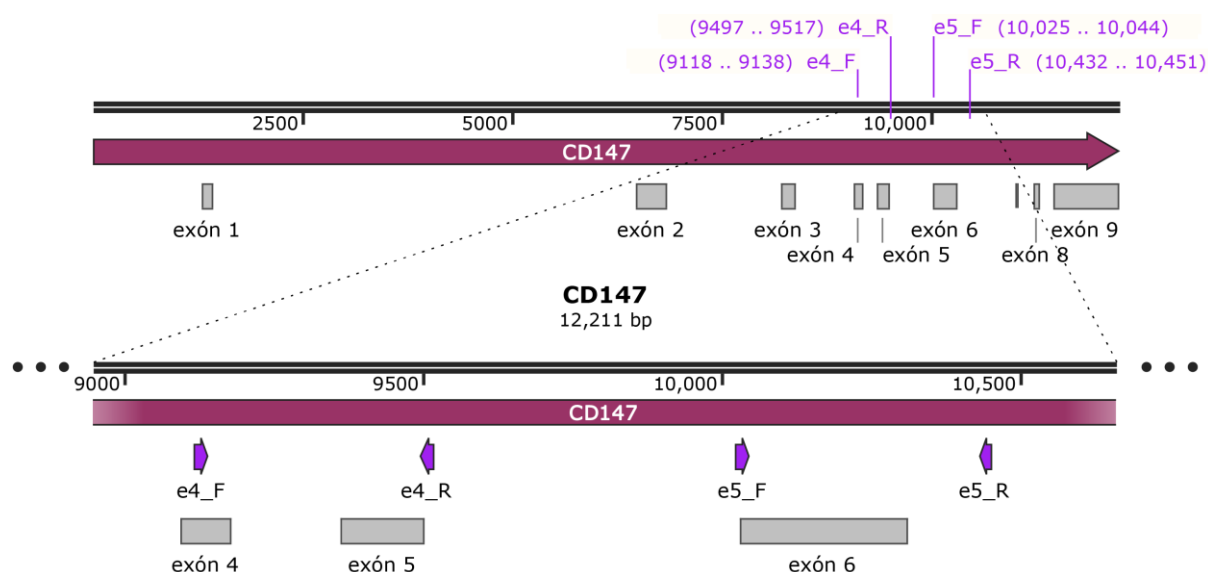
Obtenidos empleando la herramienta *CRISPR design tool* (<http://crispr.mit.edu> [27]) donde el *hit score* es el puntaje de 0 a 100% que corresponde aproximadamente con la afinidad de la secuencia guía por la secuencia *off-target*.

El sistema CRISPR-Cas9 empleado en este trabajo tiene sus pros y contras. Una ventaja es que se emplea un solo vector para expresar tanto el sgRNA como la Cas9, necesitando un único evento de transfección. Sin embargo, este sistema tiene la desventaja de que una vez insertado el vector en el genoma del hospedero se expresará de manera continua la endonucleasa y el sgRNA, aumentando la probabilidad de editar genes *off-targets* con el paso del tiempo. Una manera de sobrellevar este problema es verificar de manera rutinaria (cada 15 pasajes aproximadamente) el estado de la secuencia de los posibles genes *off-targets*.



### 4.3. Diseño de cebadores

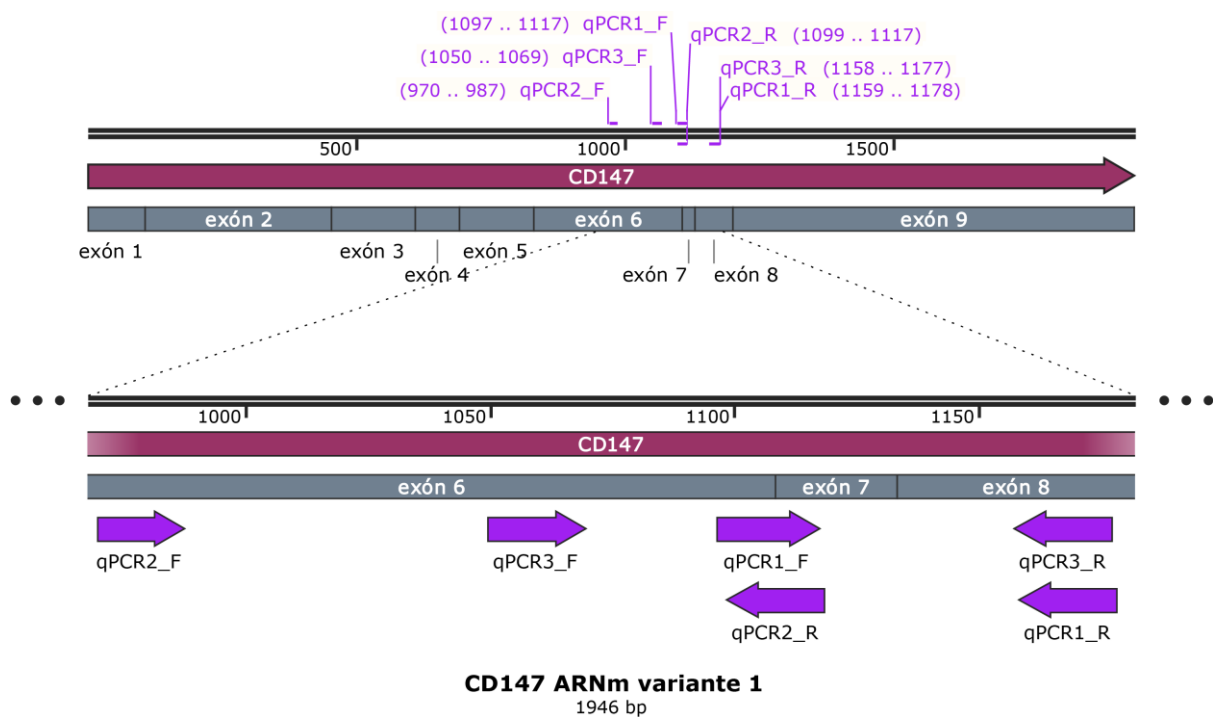
Para la verificación de la edición genómica mediante reacciones en cadena de la polimerasa (PCR) y secuenciación por Sanger, se diseñaron 2 pares de cebadores cuyos amplicones contienen los blancos genómicos de los ARN guía diseñados (exones 4 y 5 de la isoforma 2 de *CD147*; Figura 9), 9 pares para amplificar las secuencias *off-targets* de los ARN guía, y 3 pares para amplificar el ARN mensajero de *CD147* (Figura 10). Adicionalmente, el laboratorio cuenta con pares de cebadores que amplifican el gen *ACTB* ( $\beta$ 1actina 1 y  $\beta$ 1actina 2), y para amplificar el gen *cas9* (*cas9\_F* y *cas9\_R*) que también fueron utilizados en este trabajo (Tabla 7). Para el diseño de los cebadores usados en este trabajo se utilizó la herramienta *primer-BLAST* (<https://www.ncbi.nlm.nih.gov/tools/primer-blast>)<sup>110</sup>, usando como molde secuencias de referencia de la base de datos *RefSeq*<sup>109</sup>. Para los cebadores que amplifican el ARNm se buscó que los cebadores se unan a la región de unión exón-exón del ARNm, evitando de esta forma que amplifiquen ADN genómico.



**Figura 9: Sitios de unión de cebadores diseñados para amplificar el blanco genómico de los sgrNA.**

Cebadores diseñados con primer-BLAST<sup>110</sup> usando como molde la secuencia de referencia de *CD147* de GenBank (NG\_007468.1). El par de cebadores *e4\_F* - *e4\_R* genera un amplicón de largo 400 pb que contiene parte del exón 4 y la totalidad del exón 5. El par *e5\_F* - *e5\_R* genera un amplicón de largo 427 pb que contiene al exón 6. Los exones numerados corresponden a la variante 1 del ARNm de *CD147*.

Los cebadores diseñados para amplificar los blancos genómicos de los ARN guía se utilizaron en reacciones de PCR y para secuenciación por Sanger para verificar la edición génica. Los cebadores diseñados para amplificar cas9 se utilizaron en reacciones de PCR para verificar la transfección del vector lentiCRISPR v2 en células PC3. Finalmente, los cebadores diseñados para amplificar el ARNm de *CD147* se utilizaron en reacciones de PCR cuantitativo (qPCR) para verificar la expresión de *CD147* en las líneas celulares *knock-out* generadas.



**Figura 10: Sitios de unión de cebadores diseñados para amplificar ARNm de *CD147*.**

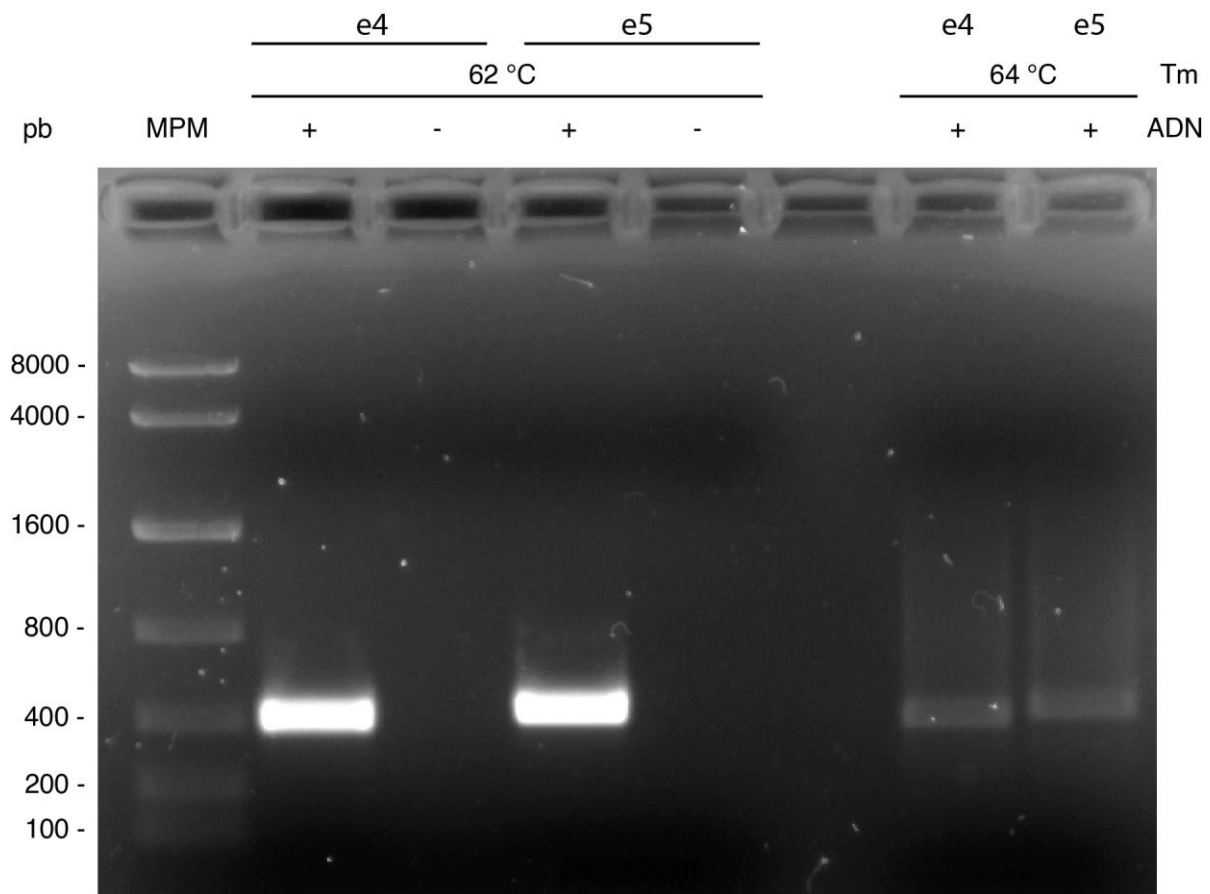
Cebadores diseñados con primer-BLAST<sup>110</sup> usando como molde la secuencia de referencia de la variante 1 del ARNm de *CD147* de GenBank (NM\_001728.4). Cebadores diseñados para amplificar ARNm en PCR cuantitativa (qPCR). Los cebadores qPCR1\_F y qPCR2\_R unen en la unión entre el exón 6 y el exón 7, evitando que amplifiquen ADN genómico.

Región	Nombre	Secuencia	Tamaño amplicón	Tm (°C)	DMSO
CD147 Exón 3 y 4	e4_F	GGGAGAGTACTCCTGCGTCTT	400 pb	62	-
	e4_R	CCTCCTTGGCTTCTCACCTTG			
CD147 Exón 5	e5_F	TCCCCTAGGCCCTCATGAAC	427 pb	62	-
	e5_R	CTTTCGGGGTCCCATCAACT			
TTF2	TTF2_F	CCTGAGACCGAGATGTGTTACG	427 pb	63	-
	TTF2_R	AAGCTATGCTCCCGTTTCC			
AP3D1	AP3D1_F	ATAGGCCTGTGGGGTCACTC	406 pb	63	-
	AP3D1_R	CAGCCACTCAGCACTGCAC			
CELSR2	CELSR2_F	CTGGAAGCAGTTCCTCAACAC	428 pb	60	-
	CELSR2_R	GGGTAAACCAGAGGGACACC			
PPP1R3F	PPP1R3F_F	CGTACGGCCCCTGTGGAG	520 pb	64	+
	PPP1R3F_R	CGAAGGAGCGTTTCAGCAC			
MRGPRX1	MRGPRX1_F	CCTGGTCAGCGGTATCTTCC	433 pb	65	+
	MRGPRX1_R	CGCAGCAGACTTCTCTTCC			
AGAP1	AGAP1_F	TGGTTCTACTGCGCTATTCCAG	440 pb	60	-
	AGAP1_R	GAGCTTGTTTTTGCCATCC			
FHAD1	FHAD1_F	TTGGCCTCCCAAAGTGTTG	441 pb	60	-
	FHAD1_R	CCTGACGCATACCTTGCTTG			
RIPK4	RIPK4_F	TGGAGGCAACTCAGTTCCAG	443 pb	60	-
	RIPK4_R	TTCCTCCCAAGGGTGAGATG			
MAPKBP1	MAPKBP1_F	GGTCCCCTGATGCTAGATGC	443 pb	62	-
	MAPKBP1_R	GCCTGTGGAACACAGAGAGC			
ARNm CD147 Exones 5, 6 y 7	qPCR1_F	ACGTCCTGGATGATGACGACG	82 pb	60	-
	qPCR1_R	CGGACGTTCTTGCCTTTGTC			
	qPCR2_F	CAGGCCATCATCAGCTC	148 pb	60	-
	qPCR2_R	CGTCGCATCATCCAGGAC			
	qPCR3_F	GCTGGTCACCATCATCTTCA	128 pb	60	-
qPCR3_R	GGACGTTCTTGCCTTTGTCA				
Promotor U6	hU6_F	GAGGGCCTATTTCCCATGATT			
ACTB	$\beta$ lactina 1	CACCCAGCACAAATGAAGATCA	131 pb	60	-
	$\beta$ lactina 2	CTCGTCATACTCCTGCTTGC			
cas9	cas9_F	GCCGAGAATATCATCCACCTG	108 pb	56	-
	cas9_R	CTCTTTGGTGCTGGTGTACCT			

**Tabla 7: Cebadores diseñados.**

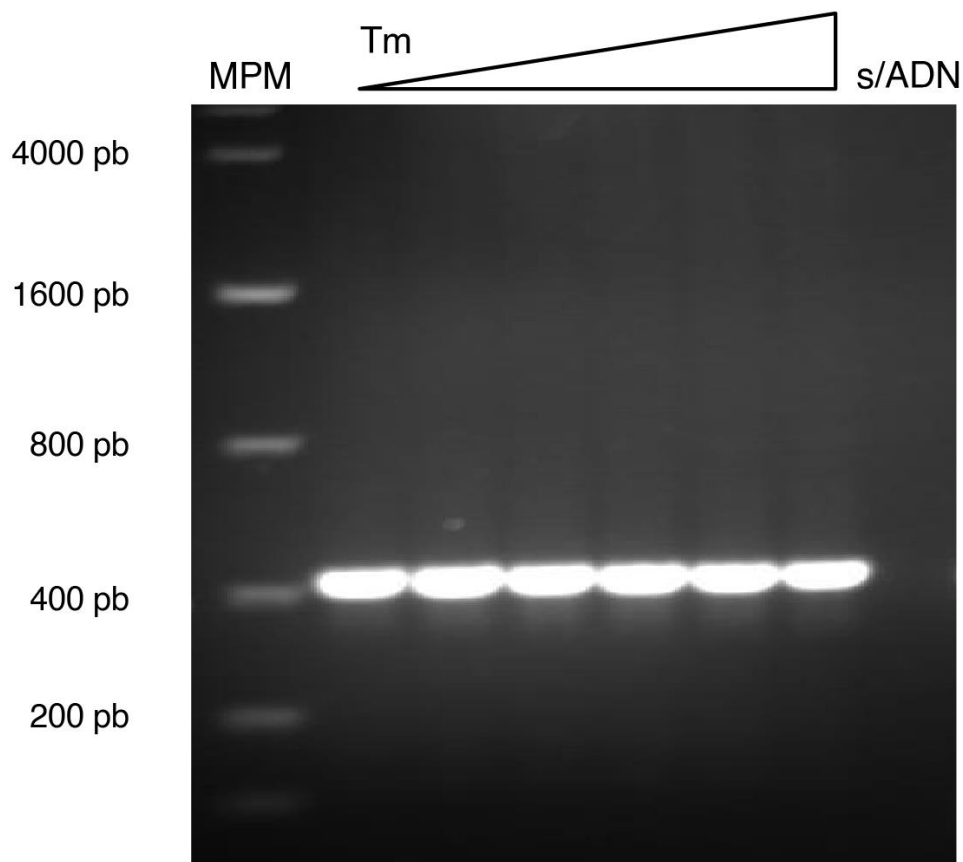
Se empleó *primer-BLAST*<sup>110</sup> usando secuencias de referencia de *GenBank* como molde. Se indican para cada par de cebadores la temperatura de fusión y presencia o ausencia de dimetilsulfóxido en la reacción en cadena de la polimerasa (PCR). Tm, temperatura de fusión; DMSO, dimetilsulfóxido.

Para cada par de cebadores se realizaron pruebas de PCR para determinar las condiciones óptimas de la reacción (Tabla 7). Cada producto de PCR se resolvió por electroforesis en gel de agarosa. Para los cebadores diseñados para amplificar los exones 4 o 5 del gen *CD147* se verificó que producen una única banda en condiciones de 62 °C T<sub>m</sub> y sin la adición de DMSO (Figura 11). El par de cebadores diseñado para amplificar el gen *TTF2* produce una única banda en 63 °C T<sub>m</sub> y sin DMSO (Figura 12). Para los tres pares de cebadores diseñados para amplificar el ARNm de *CD147* los mejores resultados se obtienen con una T<sub>m</sub> de 60 °C y sin la adición de DMSO.



**Figura 11: Puesta a punto de PCR para la amplificación de los exones 4 y 5 de *CD147*.**

Visualización en gel de agarosa del producto de PCR de la amplificación de los exones 4 y 5 de *CD147* probando temperaturas de fusión de 62 °C y 64 °C. Se utilizaron los pares de cebadores e4\_F y e4\_R para amplificar una región de largo 400 pb que contiene al exón 4 (e4). Se utilizaron los pares de cebadores e5\_F y e5\_R para amplificar una región de 427 pb que contiene al exón 5 (e5). Para ambos pares de cebadores se observa una única banda del tamaño esperado correspondiente, siendo mejores los resultados cuando se utiliza una T<sub>m</sub> de 62 °C. T<sub>m</sub>, temperatura de fusión. MPM, marcador de peso molecular (KAPA™ Express Ladder #KK6305). pb, pares de bases. ADN, presencia (+) o ausencia (-) de ADN molde en la mezcla de reacción.



**Figura 12: Puesta a punto de PCR para el par de cebadores TTF2\_F y TTF2\_R**

Visualización en gel de agarosa del producto de PCR de la amplificación del *off-target* TTF2 usando el par de cebadores TTF2\_F y TTF2\_R. Se probaron temperaturas de fusión de 60°C a 65°C. Para todas las temperaturas se observa una única banda del tamaño esperado (427 pb). Tm, temperatura de fusión. MPM, marcador de peso molecular (KAPA™ Express Ladder #KK6305). s/ADN, control negativo sin la adición de ADN molde a la mezcla de reacción.

#### 4.4. Clonado de ARNs guías en el vector lentiCRISPR v2

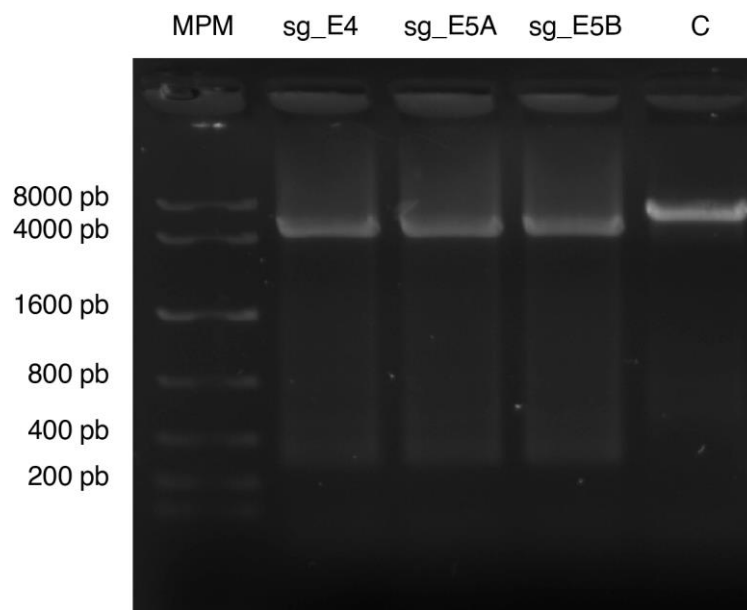
Cada ARN guía diseñado se clonó en el sitio BsmBI del vector lentiviral *lentiCRISPR* v2 (ver Figura 5). Debido a que los vectores lentivirales poseen repetidos largos terminales (LTR) que pueden ser reconocidos por la recombinasa bacteriana RecA, se emplearon bacterias *E. coli* de la cepa Stb13 para la expansión de los tres *lentiCRISPR\_v2* producidos. Esta cepa se caracteriza por ser RecA negativa ya que poseen la mutación *recA1*, favoreciendo la integridad del vector lentiviral durante su expansión en el cultivo bacteriano <sup>137</sup>.

sgRNA	Concentración (ng/μl)	A 260 nm / A 280 nm	A 260 nm / A 230nm
sg_E4	4684	1,61	1,72
sg_E5A	5003	1,64	1,82
sg_E5B	5150	1,07	1,36

**Tabla 8: Concentración y pureza del ADN plasmídico extraído de cultivos bacterianos transformados con el vector *lentiCRISPR v2***

Se extrajo ADN plasmídico a partir de colonias de bacterias transformadas con el vector *lentiCRISPR v2* portando uno de los tres sgRNAs diseñados (sg\_E4, sg\_E5A y sg\_E5B). Como medida de pureza se muestra la relación de absorbancia de luz a 260 nm y 280 nm y la relación de absorbancia a 260 nm y 230 nm. A, absorbancia.

Luego de la transformación de las bacterias *Stb13* y la selección con ampicilina de las colonias portadoras del *lentiCRISPR v2*, se tomó una colonia portadora de cada sgRNA y se cultivó en medio LB líquido. A partir de estos cultivos se extrajo ADN plasmídico para verificar la presencia del vector *lentiCRISPR v2*. El ADN plasmídico extraído se cuantificó (Tabla 8) y se utilizó como molde para la amplificación por PCR usando los cebadores hU6\_F y cas9\_R (Tabla 7 y Figura 13) como forma de corroborar su identidad. Como control del PCR se utilizó el vector *lentiCRISPR v2*. Para estas muestras se espera obtener un producto de PCR de tamaño 4640 pb y para el control se espera obtener un producto de 6525 pb. El producto de PCR en el control es mayor debido a que cuenta con un fragmento de aproximadamente 2000 pb que está ausente en las muestras ya que se pierde en el proceso de insertar el sgRNA en el vector. El producto de PCR se resolvió mediante electroforesis en gel de agarosa (Figura 13). Para los tres cultivos se observa una banda de entre 4000 pb y 8000 pb, cercano al tamaño esperado de 4640 pb, y para el control se observa una única banda de entre 4000 pb y 8000 pb, cercano al tamaño esperado de 6525 pb. Con estos resultados se verifica la presencia del vector *lentiCRISPR v2* en los tres cultivos.



**Figura 13: Presencia del vector lentiCRISPR v2 en cultivos de bacterias transformadas**

Amplificación mediante PCR de una región del ADN plasmídico purificado a partir de los cultivos de bacterias *Stbl3* transformadas con vectores *lentiCRISPR v2*. Se utilizan los cebadores hU6\_F y Cas9\_R. MPM, marcador de peso molecular (KAPA™ Express Ladder #KK6305), sg\_E4, sg\_E5A y sg\_E5B, ARN guía portado por cada vector purificado, C, control de amplificación.

Luego de verificar que el vector *lentiCRISPR v2* se mantiene en los cultivos bacterianos transformados, estos se expandieron y se extrajo el vector para obtener ADN en cantidad y pureza suficientes para transfectar cultivos celulares. Se cuantificó el ADN purificado mediante la cuantificación de la absorbancia de la muestra a 260 nm ( $A_{260}$ ), 280 nm ( $A_{280}$ ) y 230 nm ( $A_{230}$ ) (Tabla 9). Como indicador de la pureza del ADN purificado se utilizó las relaciones de absorbancia  $A_{260}/A_{280}$  y  $A_{260}/A_{230}$ , considerándose como adecuados valores mayores a 1,7. Para todos los extractos se obtuvieron valores mayores a 1,8 con la excepción del vector portador del sg\_E4 para el cual se midió una relación  $A_{260}/A_{280}$  de 1,71 indicando que los extractos son adecuadamente puros.

sgRNA	Concentración (ng/μl)	A 260 nm / A 280 nm	A 260 nm / A 230 nm
sg_E4	4320	1,71	2,05
sg_E5A	1586	1,87	2,22
sg_E5B	1437	1,89	2,25

**Tabla 9: Concentración y pureza de los vectores lentiCRISPR v2 purificados.**

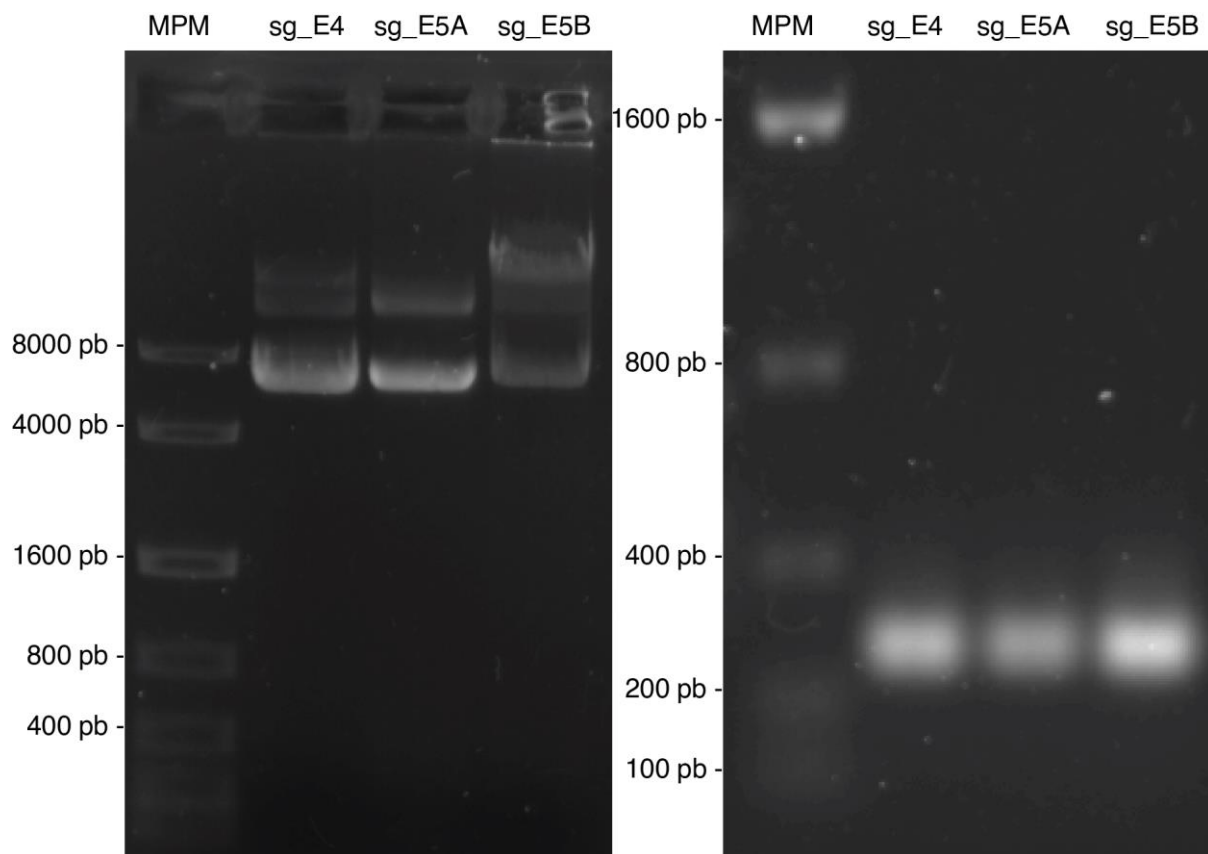
Se extrajo ADN plasmídico a partir de colonias de bacterias transformadas con el vector lentiCRISPR v2 portando uno de los tres ARN guía diseñados (sg\_E4, sg\_E5A y sg\_E5B). Como medida de pureza se muestra la relación de absorbancia de luz a 260 nm y 280 nm y la relación de absorbancia a 260 nm y 230 nm. A, absorbancia.

La integridad del ADN purificado se evaluó mediante electroforesis en gel de agarosa utilizando 500 ng de cada muestra (Figura 14). El vector se visualiza en el gel de agarosa como múltiples bandas, una banda mayor a 8000 pb que corresponde con el tamaño esperado del vector de aproximadamente 12000 pb, y una banda que migra entre los marcadores de 4000 pb y 8000 pb y que corresponde al vector circular superenrollado. No se observa degradación del vector.

Por último, para verificar la presencia de la secuencia codificante del ARN guía en los vectores purificados se realizó una PCR empleando al cebador hU6-F (ver Tabla 7 y Figura 5) como cebador directo y utilizando al oligonucleótido diseñado para cada ARN guía (ver Tabla 5) como cebador reverso. Para este par de cebadores hU6\_F-sgRNA se espera obtener un producto de PCR de 264 pb. Se resolvió el producto de PCR mediante electroforesis en gel de agarosa y se observó para los tres extractos una única banda de entre 200 pb y 400 pb, cercano al tamaño esperado de 264 pb (Figura 14).

Se transformaron bacterias Stbl3 con cada uno de los tres vectores diseñados previamente que portan el sistema CRISPR-cas9. A partir de cultivos de estas bacterias transformadas se logró amplificar y purificar los vectores diseñados manteniéndose la integridad del vector y obteniéndose ADN plasmídico de pureza adecuada.





**Figura 14: Verificación de la integridad de los plásmidos purificados.**

**Izquierda:** Integridad de plásmidos. Electroforesis en gel de agarosa de ADN plasmídico purificado a partir de cultivos de bacterias *Stbl3* transformadas con el vector *lentiCRISPR v2*. **Derecha:** Presencia de los ARN guía en los vectores purificados. Electroforesis en gel de agarosa del producto de PCR en los vectores purificados. Se utilizó el cebador hU6\_F como cebador directo. Para cada vector se utilizó como cebador reverso el mismo oligonucleótido que llevan en su secuencia como ARN guía para el sistema CRISPR-cas9. Para el par de cebadores hU6\_F-sgRNA se espera obtener un producto de PCR de 264 pb. MPM, marcador de peso molecular (KAPA™ Express Ladder #KK6305), sg\_E4, sg\_E5A y sg\_E5B, ARN guía portado por cada vector purificado.

#### 4.5. Generación de una línea celular de cáncer de próstata metastásica *CD147* negativa

Como se vio en el análisis de expresión génica, existe una posible asociación entre *CD147* y la progresión a la metástasis. Para estudiar esta relación, en este trabajo se seleccionó una línea celular PC3, que proviene de una metástasis en el hueso de un tumor prostático humano para editar el gen *CD147*.

Se transfectaron células PC3 mediante lipofección con los vectores *lentiCRISPR v2* generados previamente. La incorporación del vector *lentiCRISPR v2* confiere resistencia a la puromicina a las células. Se utilizó puromicina para seleccionar y aislar

cinco clones resistentes transfectados con el lentivirus portador del sg\_E4, cuatro clones con el lentivirus-sg\_E5A y un clon con el virus portador del lentivirus-sg\_E5B. A partir de estos clones se obtuvieron 4 líneas celulares viables: una capaz de expresar el sg\_E4 (llamado de aquí en más E4-3) y tres capaces de expresar el sg\_E5A (llamado de aquí en más E5A-1, E5A-3, y E5A-4). Lamentablemente el clon portador del sg\_E5B sufrió una contaminación con hongos y debió ser descartado, resultando en que no se pudo obtener ninguna línea celular viable que exprese este sgRNA.

Como control del sistema CRISPR-Cas9 se transfectaron células PC3 con el vector *lentiCRISPR v2* sin ARN guía (llamado de aquí en más vector\_V2). Se seleccionaron células resistentes a puromicina y se trabajó con el conjunto no clonal de células transfectadas.

La ausencia de expresión de *CD147* en las líneas celulares producidas se evaluó mediante *western blot* y RT-qPCR cuantitativo. Además, para identificar la edición del gen se secuenció la región de *CD147* reconocida por los ARN guía.

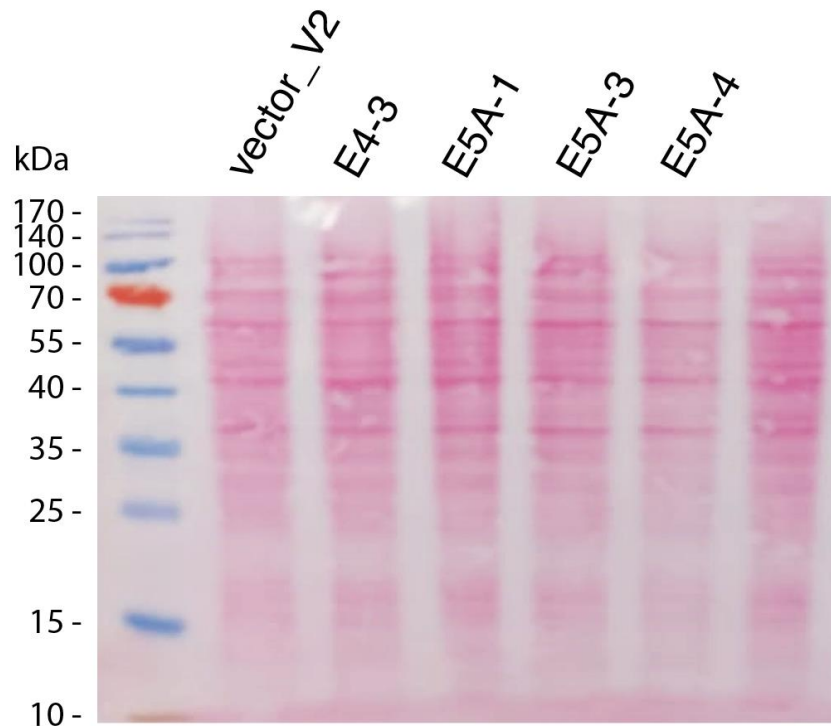
Para evaluar la expresión de la proteína CD147, se realizó un extracto de proteínas solubles de las líneas celulares mediante *buffer* de lisis celular RIPA. Se realizaron extractos de las 4 líneas celulares portadoras de ARN guía y de la línea vector\_V2. Los extractos proteicos se cuantificaron utilizando una dilución 1:10 del extracto de proteínas (Tabla 10) y para todos los extractos proteicos se obtuvo una concentración de proteínas mayor o igual a 20,0 µg/µl.

Línea celular	Concentración (µg / µL) dilución 1:10	Concentración real (µg / µL)
vector_V2	2,66	26,6
E4-3	2,56	25,6
E5A-1	3,64	36,4
E5A-3	2,22	22,2
E5A-4	2,00	20,0

**Tabla 10: Concentración de proteína extraída de las líneas celulares generadas**

Concentración, en microgramos por microlitro, de proteína soluble extraída de cultivos de líneas celulares generadas. Concentración medida sobre una dilución 1:10 del extracto total.

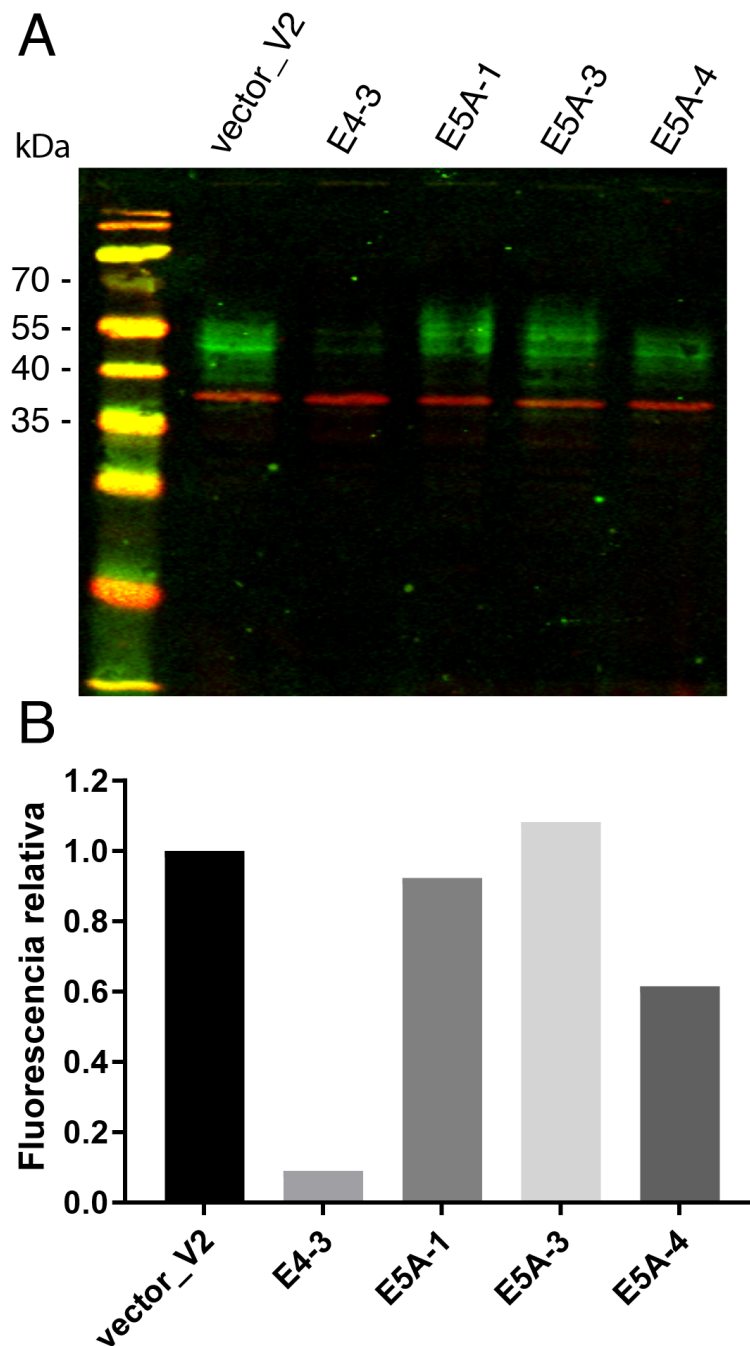
Los extractos fueron sometidos a electroforesis en gel de poliacrilamida (SDS PAGE) utilizando 60 µg de proteína total para cada carril. Luego de transferir la proteína a membrana de nitrocelulosa, se verificó la transferencia de proteínas mediante tinción de Ponceau (Figura 15).



**Figura 15: Detección de proteínas totales transferidas a membrana de nitrocelulosa.**

Tinción de Ponceau sobre membrana de nitrocelulosa. Las bandas del primer carril a la izquierda corresponden al marcador de peso molecular (PageRuler™ Prestained Protein Ladder #SM0671). Se cargaron 60 µg de proteínas solubles de cada extracto.

Para detectar la proteína CD147, la membrana de nitrocelulosa se incubó con un anticuerpo anti-CD147 y con un anti-GAPDH que se usó como control de carga del gel. Ambos fueron detectados usando anticuerpos secundarios fluorescentes. La proteína CD147 tiene un peso molecular de 27 kDa<sup>138</sup>, pero debido a que en la célula se encuentra glicosilada su peso molecular varía y por lo tanto en las imágenes obtenidas CD147 se visualiza como un conjunto de bandas de entre 35 kDa y 65 kDa<sup>139</sup>. En todas las líneas celulares analizadas se detectan CD147 y GAPDH, aunque en la línea E4-3 la fluorescencia correspondiente a CD147 es visiblemente menor que en las otras líneas celulares (Figura 16A).



**Figura 16: Expresión proteica de CD147 en líneas celulares generadas.**

Detección de CD147 en líneas celulares transfectadas con el sistema CRISPR-cas9 mediante *western blot*. **A.** Detección por fluorescencia de CD147 (verde) y el control de carga GAPDH (rojo). Marcador de peso molecular: PageRuler™ Prestained Protein Ladder 170 kDa (#SM0671). **B.** Relación de la fluorescencia de CD147 entre la muestra y la línea control vector\_V2. Medida semicuantitativa de la fluorescencia de CD147 normalizada con la del control de carga GAPDH.

Para comparar la expresión de CD147 entre las muestras, se utilizó el programa ImageJ<sup>120</sup> para obtener una medida semicuantitativa de la intensidad de cada señal

fluorescente. En base a estas medidas, se calculó la relación en el nivel de expresión de CD147 entre las muestras y la línea control vector\_V2 (Figura 16B).

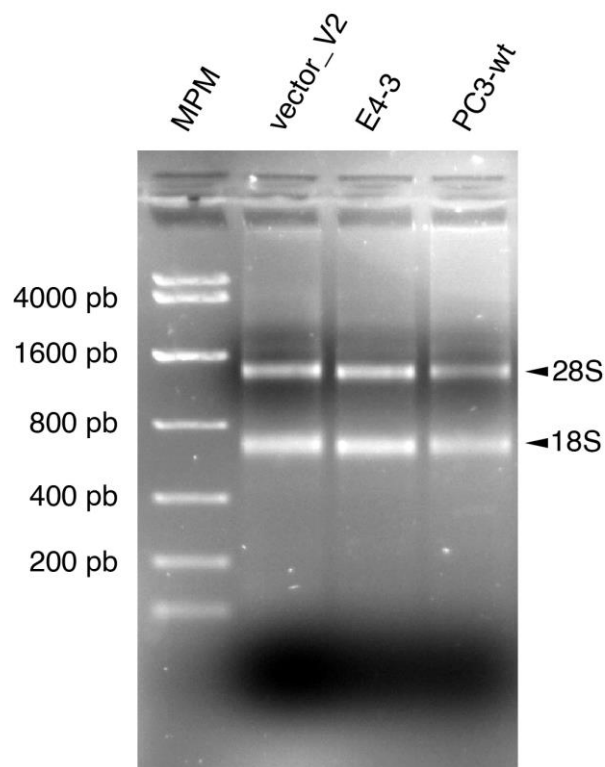
La expresión de CD147 detectada en las líneas E5A-1 y E5A-3 es similar a la expresión en la línea control, mientras que en las líneas E4-3 y E5A-4 se observa una marcada disminución en los niveles de la proteína en comparación con la línea vector\_V2. La mayor reducción se observa en la línea E4-3, donde el nivel de expresión de CD147 es aproximadamente 0,09 veces el de la línea control. En ninguna de las líneas celulares generadas se observó una pérdida total de expresión de CD147. Sin embargo, el bajo nivel de proteína en la línea E4-3 evidencia que la edición génica tuvo un efecto en estas células, aunque no haya generado un *knock-out* total.

Para comprender a que nivel de la expresión del gen sucede la reducción observada, se evaluó la expresión de ARNm del gen *CD147*. Se prosiguió el análisis únicamente con la línea E4-3 ya que en la misma se observó la mayor reducción de la expresión de proteína CD147. Se purificó ARN a partir de extractos celulares de las líneas E4-3, vector\_V2 y de células PC3 sin modificar (PC3-wt). Se cuantificó el ARN purificado total (Tabla 11) y se verificó su integridad mediante electroforesis en gel de agarosa (Figura 17). Para los tres extractos se observan en el gel de agarosa dos bandas nítidas correspondientes al ARN ribosomal 28S y 18S indicando una buena integridad del ARN purificado. A partir del ARN purificado se obtuvo ADN complementario (ADNc) del ARNm mediante retrotranscripción.

Línea celular	Concentración (ng/μl)
vector_V2	523,0
E4-3	806,0
PC3-wt	731,4

**Tabla 11: Cuantificación de ARN purificado a partir de extractos celulares.**

Concentración del ARN purificado a partir de la línea celular editada E4-3 y de las líneas celulares sin editar vector\_V2 y PC3-wt. ng/μl nanogramos por microlitro.



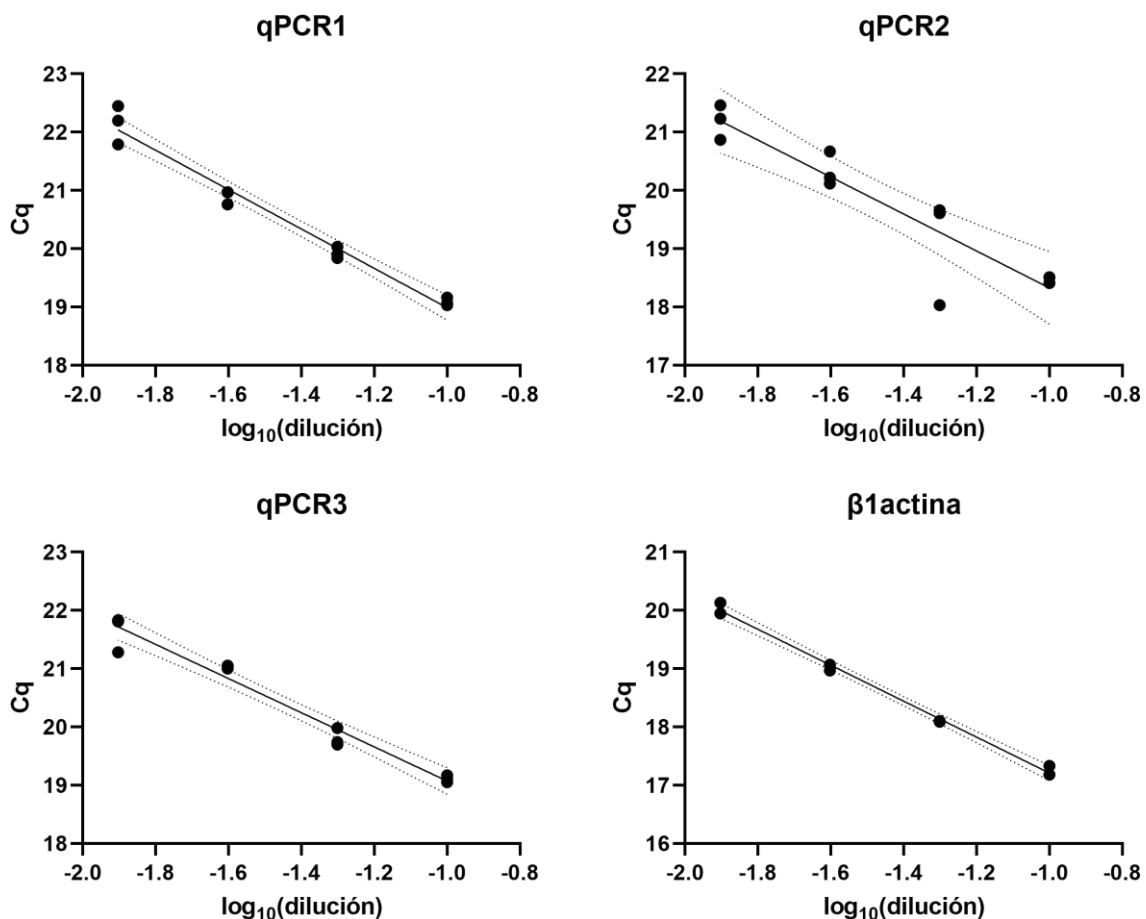
**Figura 17: Integridad del ARN extraído de las líneas celulares generadas.**

Integridad del ARN purificado a partir de la línea celular editada E4-3 y de las líneas celulares sin editar vector\_V2 y PC3-wt. Gel de agarosa con 2 $\mu$ g de ARN en cada carril. Las dos bandas claramente visibles corresponden a las subunidades 18S y 28S del ARN ribosomal.

La expresión del gen *CD147* se evaluó mediante qPCR utilizando como molde el ADNc obtenido. Como se mencionó previamente, en este trabajo se diseñaron tres pares de cebadores para amplificar el ARNm de *CD147* (Tabla 7). Dos de estos cebadores (qPCR1 y qPCR2) se diseñaron para que aparezcan en la unión exón-exón de los genes de interés, evitando que amplifiquen posibles contaminantes de ADN genómico. Adicionalmente, el laboratorio cuenta con un par de cebadores que amplifica ARNm de *ACTB* ( $\beta$ 1actina 1 y  $\beta$ 1actina 2) que se utilizó en este trabajo como referencia.

Para evaluar la eficiencia de los cebadores se generaron curvas de calibración a partir de reacciones de qPCR utilizando diluciones seriadas del ADNc. Se realizaron reacciones en triplicado para cada par de cebadores excepto para el par de cebadores  $\beta$ 1actina para el cual se hizo la reacción en duplicado para optimizar el uso del espacio en la placa de qPCR. Se midió el número de ciclos de amplificación

necesarios para superar el umbral de detección de fluorescencia (ciclos de umbral, Cq) y se generaron las curvas de calibración (Figura 18). Se calculó la eficiencia de cada par de cebadores en base a la relación entre el número de Cq y el logaritmo de las diluciones. Para todos los casos se calculó el cuadrado del coeficiente de correlación de Pearson ( $R^2$ ) para evaluar la existencia de una relación lineal entre el Cq y la concentración de la muestra (Tabla 12). Se calculó una eficiencia de 98% para el par de cebadores qPCR1 ( $R^2 = 0,97$ ), 107% para el par qPCR2 ( $R^2 = 0,84$ ), 119% para el par qPCR3 ( $R^2 = 0,96$ ) y 111% para el par de cebadores  $\beta$ 1actina ( $R^2 = 0,99$ ). En base a estos cálculos, se seleccionó el par de cebadores qPCR1 para amplificar *CD147*. Este par de cebadores es el que más se acerca a una eficiencia ideal de 100%.



**Figura 18: Curvas de calibración de los cebadores diseñados para qPCR**

Curvas de calibración generadas a partir de medidas en triplicado de Cq para cada par de cebadores. Se muestra la línea de mejor ajuste generada por un modelo de regresión lineal simple. Las líneas punteadas corresponden a los límites de un intervalo de confianza 95% para la línea de mejor ajuste.

Cebadores	Pendiente	Eficiencia	R <sup>2</sup>
qPCR1	-3,377	0,978	0,9745
qPCR2	-3,172	1,067	0,8365
qPCR3	-2,931	1,194	0,9630
β1actina	-3,083	1,110	0,9944

**Tabla 12: Eficiencia de cebadores diseñados para qPCR**

Eficiencia calculada como  $Eficiencia = 10^{(-\frac{1}{pendiente})} - 1$  donde pendiente refiere a la pendiente de la relación lineal entre el ciclo de umbral y la concentración del ADN molde. R<sup>2</sup>, cuadrado del coeficiente de correlación de Pearson calculado para la relación lineal entre el ciclo de umbral y la concentración del ADN molde.

Una vez determinadas las eficiencias de amplificación de los cebadores, se realizó qPCR utilizando el par de cebadores qPCR1 para amplificar *CD147* y el par de cebadores β1actina como gen de referencia. Se realizaron tres réplicas técnicas de la reacción de qPCR para cada una de las líneas E4-3, vector\_V2 y PC3-wt (Tabla 13). Se midió C<sub>q</sub> y se utilizó el *software EcoStudy*<sup>125</sup> para calcular la expresión de *CD147* en las líneas E4-3 y vector\_V2 relativo a la expresión de *CD147* en las células PC3-wt. Dado que la eficiencia de los cebadores no es 100%, se utilizó la metodología de cuantificación relativa (RQ) para calcular la expresión relativa de *CD147*. Para cada medida de RQ el *software* determina un límite superior y un límite inferior del valor de RQ en base a la incertidumbre de las medidas. Para la línea E4-3 se calculó una expresión relativa de 0,252 respecto a la línea PC3-wt con un límite superior de RQ de 0,299 y un límite inferior de 0,205, mientras que para la línea vector\_V2 se calculó un RQ de 1,202 con un límite superior de 1,421 y un límite inferior de 0,983 (Figura 19).

Estos resultados muestran que el ARNm de *CD147* es aproximadamente 4 veces menor en la línea E4-3 que en la línea *wildtype* mientras que en la línea vector\_V2 no se observa una diferencia medible. Tomado en conjunto con los resultados de expresión de proteína, se evidencia que la expresión del gen *CD147* se encuentra reducida en esta línea celular.

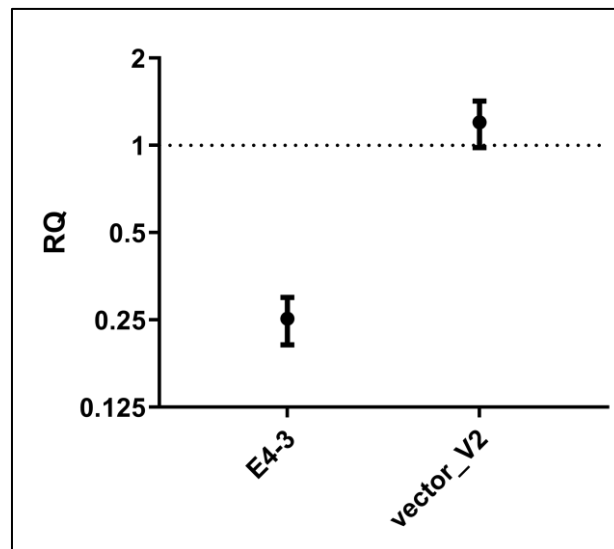


qPCR1	Ciclo de umbral (Cq)			Promedio de Cq
	Réplica 1	Réplica 2	Réplica 3	
E4-3	21,49	21,34	21,48	21,44
vector_V2	19,12	19,08	19,00	19,07
PC3-wt	20,37	19,91	20,34	20,21

$\beta$ 1actina	Ciclo de umbral (Cq)			Promedio de Cq
	Réplica 1	Réplica 2	Réplica 3	
E4-3	18,29	18,01	17,55	17,95
vector_V2	18,04	17,67	17,92	17,88
PC3-wt	18,69	19,00	18,32	18,67

**Tabla 13: Amplificación de *CD147* y *ACTB* por qPCR**

Número de ciclos de qPCR necesarios para superar el umbral de detección de la fluorescencia. qPCR realizado con tres réplicas técnicas para cada muestra. La tabla superior indica los resultados para el par de cebadores qPCR1 que amplifica el ARNm de *CD147* y la tabla inferior indica resultados para el par de cebadores que amplifica ARNm de *ACTB*.



**Figura 19: Cuantificación relativa de expresión de *CD147* en líneas celulares.**

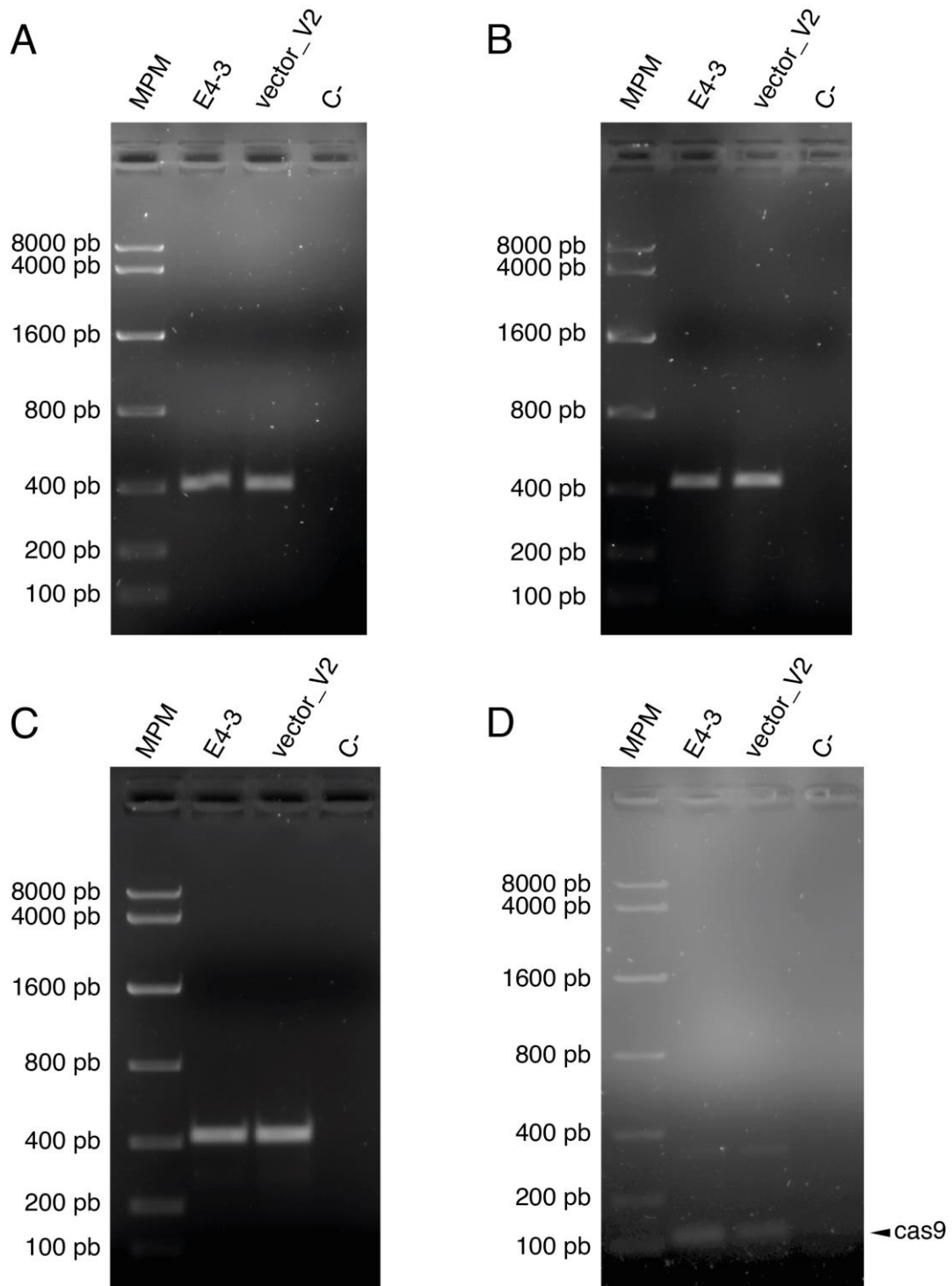
*Fold change* de la expresión de *CD147* de las líneas E4-3 y vector\_V2 en relación a la expresión en la línea sin editar PC3-wt. Expresión relativa de ARNm de *CD147* normalizada con la expresión relativa de *ACTB*. Las barras de error muestran los límites de incertidumbre para el cálculo de la expresión relativa de cada muestra. RQ, cuantificación relativa.

---

En este trabajo se utilizó el sistema CRISPR-Cas9 para generar un *knock-out* del gen *CD147* en líneas celulares PC-3. Esto se logra mediante la generación de mutaciones en *CD147* que resultan en la pérdida de función del gen. El sistema CRISPR-Cas9 genera cortes de doble hebra en una región específica del gen lo que activa el sistema de reparación no homóloga (NHEJ) que repara el daño al ADN, frecuentemente con errores. El resultado de esto es generalmente la aparición de una inserción o deleción de nucleótidos (indels) en la secuencia del gen. Si el largo de estos indels no es un múltiplo de 3, causa que ocurra un corrimiento del marco de lectura del gen, produciendo grandes cambios en la secuencia de codones y posiblemente resultando en la aparición de un codón de terminación prematuro. Generalmente, los ARN mensajeros que llevan un codón de terminación prematuro son degradados en la célula mediante la vía de degradación llamada *Nonsense-mediated decay* (NMD).

La degradación de ARN mensajeros por la vía NMD, podría explicar la reducida expresión de ARNm en la línea E4-3. De ser este el caso debería haber un codón de terminación prematuro en la secuencia del gen *CD147*.

Para identificar la edición génica en la línea celular E4-3 se secuenciaron mediante Sanger los exones 4 y 5 de *CD147*. Para la secuenciación se purificó ADN genómico a partir de las líneas celulares E4-3 y el control vector\_V2. A partir del ADN obtenido se amplificó mediante PCR las regiones de interés. Se utilizó el par de cebadores e4\_F y e4\_R para amplificar la región que contiene al exón 4 de *CD147* (Figura 20A), y el par de cebadores e5\_F y e5\_R se utilizaron para amplificar la región del exón 5 (Figura 20B). A su vez, se amplificó la región del potencial *off-target* del sg-E4 utilizando el par de cebadores TTF2\_F y TTF2\_R para verificar la ausencia de edición (Figura 20C). Por otra parte, para verificar la presencia del gen cas9 en las líneas celulares, se amplificó por PCR el gen utilizando los cebadores cas9\_F y cas9\_R (Figura 20D). La amplificación por PCR se verificó mediante electroforesis en gel de agarosa. En la electroforesis se observa una única banda bien definida del tamaño esperado (ver Tabla 7) para todos los amplicones excepto para los amplicones de cas9 para los cuales aparece una segunda banda de aproximadamente 400 pb debida a amplificación inespecífica de ADN genómico (Figura 20). A partir del gel de agarosa se purificaron y secuenciaron los amplicones correspondientes a *CD147* y *TTF2* para enviarlos a secuenciar.



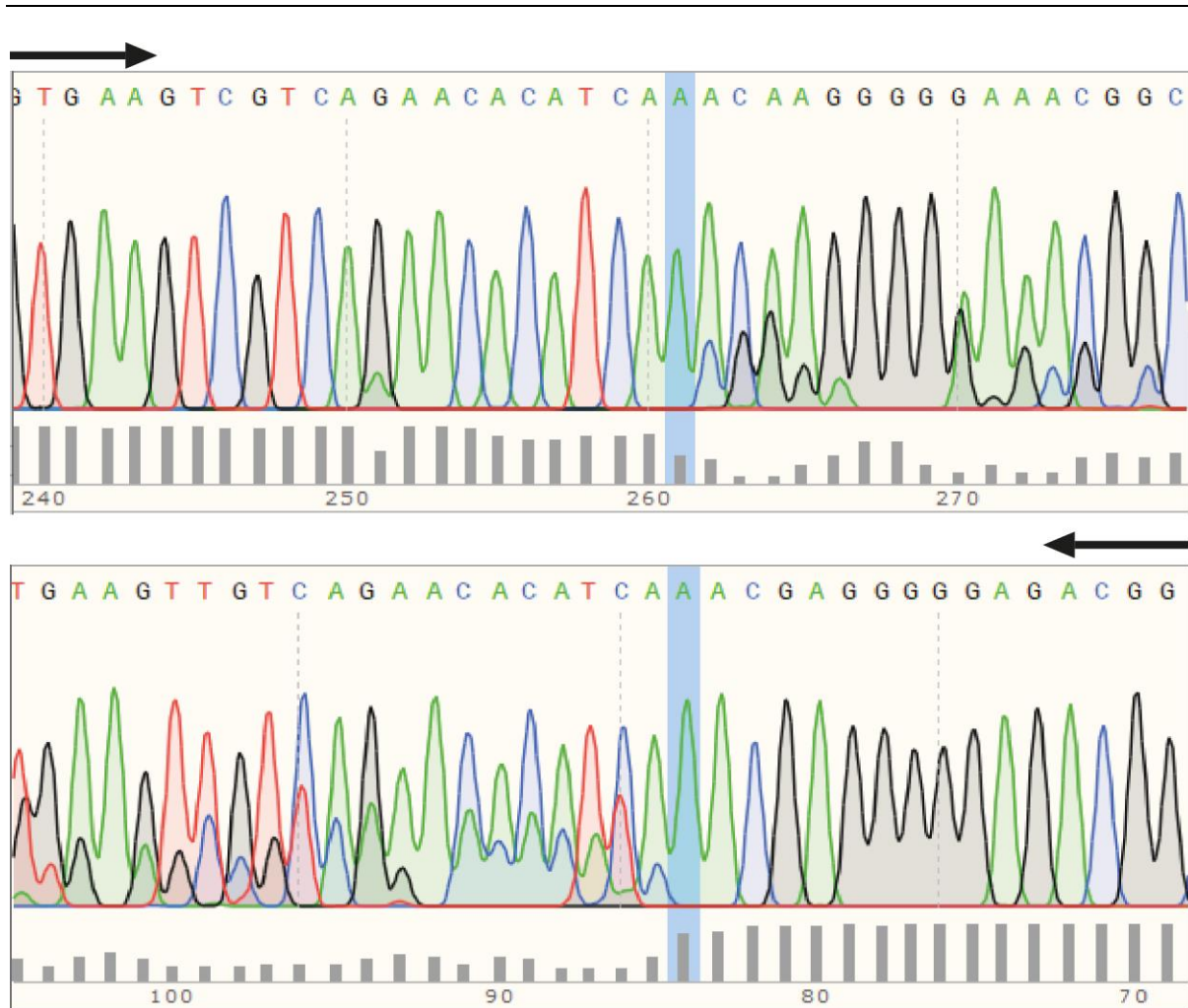
**Figura 20: Verificación de amplificación por PCR de regiones de interés para secuenciación.**

Electroforesis en gel de agarosa para los productos de PCR **A.** exón 4 de *CD147*. **B.** exón 5 de *CD147*. **C.** gen *TTF2*, *off-target* de *sg\_E4* **D.** gen *cas9*.

---

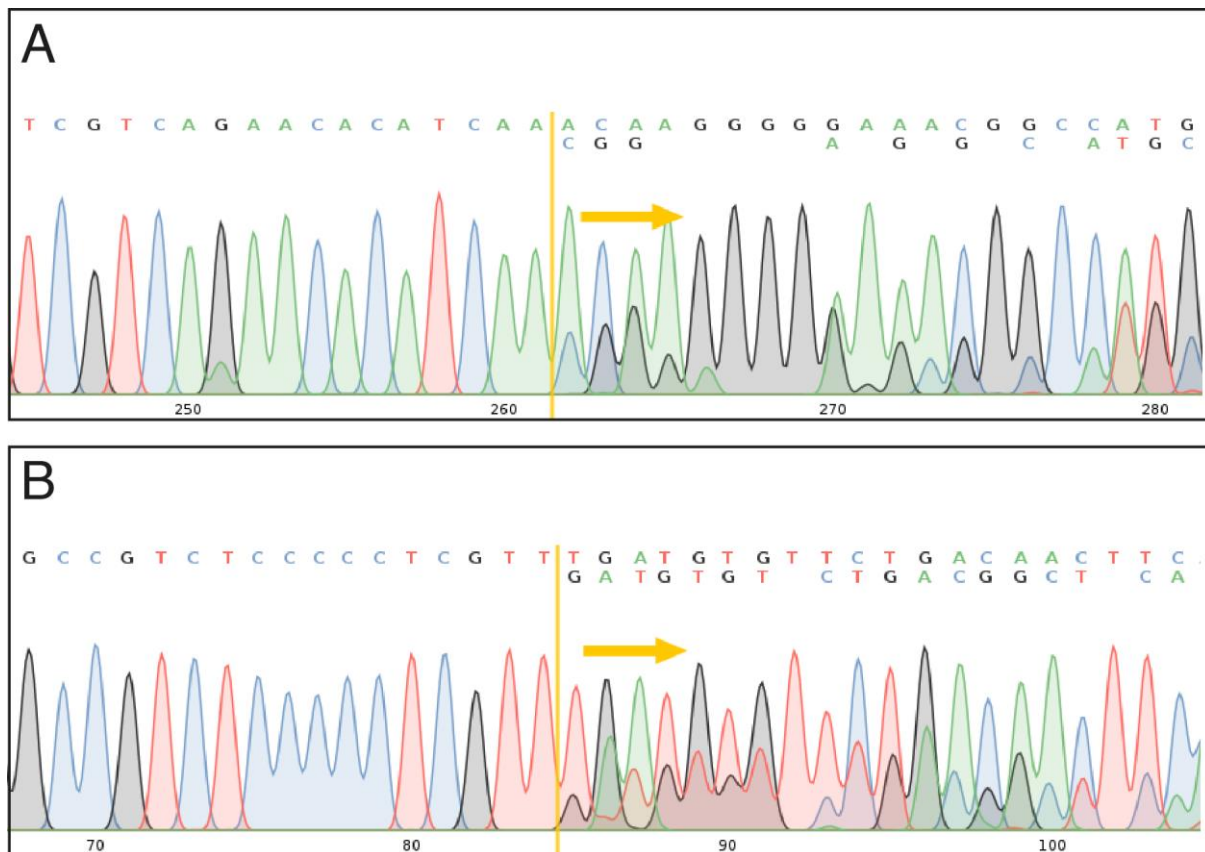
La calidad de las secuencias obtenidas se evaluó mediante cromatograma generado por el secuenciador. Para todos los amplicones secuenciados los cromatogramas evidencian una buena calidad de la secuenciación, observándose picos claros, altos, bien separados entre sí y con poco ruido de fondo (no mostrado) excepto en la secuenciación del exón 4 de la línea E4-3. Los cromatogramas de las secuencias de E4-3 muestran una buena calidad de la secuenciación en la primera parte de la secuencia, pero presentan picos dobles a partir de la posición 262 en la secuencia *forward* y 85 en la secuencia *reverse* (Figura 21). La posición a partir de la cual disminuye la calidad de la secuenciación corresponde al mismo sitio genómico en ambos casos. Además, para ambas secuencias, cada posición del cromatograma donde hay picos dobles uno de los picos es idéntico a uno de los picos de la siguiente posición. Estos resultados son indicativos de que existen al menos dos secuencias en la muestra secuenciada, una de las cuales diverge de la otra por la presencia de una inserción o una delección (indel) de un nucleótido.

Para verificar la presencia de un indel, se utilizó la herramienta CRISP-ID <sup>121</sup>. Esta herramienta permite caracterizar los indels generados por el sistema CRISPR-Cas9 y puede identificar hasta tres alelos a partir de datos de secuenciación Sanger. Utilizando esta herramienta se identificaron dos alelos en la secuencia del exón 4 en la línea E4-3 (Figura 22A y B). Estos alelos se diferencian por la inserción de un nucleótido en uno de los alelos.



**Figura 21: Cromatogramas de la secuencia del exón 4 de *CD147* de la línea E4-3.**

Se muestran los cromatogramas para la secuenciación directa y reversa del exón 4 de *CD147* en la línea celular E4-3. En la parte inferior de cada imagen, el gráfico representa la calidad de la lectura del secuenciador en cada posición. Para ambas secuencias la calidad de lectura es buena en la primera parte y decae rápidamente en la posición indicada por la barra azul. La posición de la barra azul corresponde al mismo sitio genómico en ambas secuencias indicando que la muestra puede contener múltiples alelos en este sitio.



**Figura 22: Resultados de CRISP-ID para las secuencias del exón 4 de *CD147* en la línea celular editada E4-3**

Se muestra el cromatograma para la secuenciación directa (**A**) y reversa (**B**) del exón 4 de *CD147* en la línea celular E4-3. Se indica el sitio donde el programa CRISP-ID detectó una posible divergencia entre dos alelos que difieren por la inserción o deleción de nucleótidos. El sitio donde se detecta la divergencia corresponde al mismo sitio genómico para ambas secuencias.

Utilizando el *software* MEGA X <sup>122</sup> se alinearon las secuencias obtenidas con secuencia de referencia correspondientes y se verificó que no existen modificaciones en las secuencias de las líneas celulares modificadas excepto en la secuencia del exón 4 en la línea E4-3. Utilizando esta última secuencia, se hizo un alineamiento con la secuencia correspondiente de la línea control vector\_V2 (Figura 23A), confirmando que existe una inserción de una adenina en la posición 43 del exón 4 del gen de *CD147* en la línea E4-3. Esta posición corresponde con el sitio de corte que tiene como blanco el sg\_E4 usado en esta línea celular, indicando que es el producto de la edición genómica.

**A**

```

      1           10           20           30           40 *           50
vector_V2 GGCCTCCCAGAGTGAAGGCTGTGAAGTCGTCAGAACACATCA-ACGAGGG
E4-3      GGCCTCCCAGAGTGAAGGCTGTGAAGTCGTCAGAACACATCAaACGAGGG

      51           60           70           80           90           100
vector_V2 GGAGACGGCCATGCTGGTCTGCAAGTCAGAGTCCGTGCCACCTGTCACTG
E4-3      GGAGACGGCCATGCTGGTCTGCAAGTCAGAGTCCGTGCCACCTGTCACTG

      101          110          120          130          138
vector_V2 ACTGGGCCTGGTACAAGATCACTGACTCTGAGGACAAG
E4-3      ACTGGGCCTGGTACAAGATCACTGACTCTGAGGACAAG

```

**B**

```

      1                                           12
vector_V2 GG CCT CCC AGA GTG AAG GCT GTG AAG TCG TCA GAA CAC
          - P P R V K A V K S S E H
E4-3      GG CCT CCC AGA GTG AAG GCT GTG AAG TCG TCA GAA CAC
          - P P R V K A V K S S E H
.....

      13                                           25
vector_V2 ATC AAC GAG GGG GAG ACG GCC ATG CTG GTC TGC AAG TCA
          I N E G E T A M L V C K S
E4-3      ATC AaA CGA GGG GGA GAC GGC CAT GCT GGT CTG CAA GTC
          I K R G G D G H A G L Q V
.....

      26                                           38
vector_V2 GAG TCC GTG CCA CCT GTC ACT GAC TGG GCC TGG TAC AAG
          E S V P P V T D W A W Y K
E4-3      AGA GTC CGT GCC ACC TGT CAC TGA CTG GGC CTG GTA CAA
          R V R A T C H * L G L V Q
.....

      39                                           45
vector_V2 ATC ACT GAC TCT GAG GAC AAG
          I T D S E D K
E4-3      GAT CAC TGA CTC TGA GGA CAA G
          D H * L * G Q
.....

```

**Figura 23: Secuencias del exón 4 de CD147 en las líneas E4-3 y vector\_V2**

**A.** Alineamiento de la secuencia nucleotídica del exón 4 de E4-3 con la de vector\_V2. El asterisco (\*) indica la posición donde difieren las dos secuencias por la inserción de un nucleótido en la secuencia obtenida de E4-3. **B.** Traducción de las secuencias del exón 4 de CD147. Se muestran los codones traducidos a secuencia aminoacídica. Los dos primeros nucleótidos del exón 4 forman un codón con el último nucleótido del exón 3 (no incluido en la figura). La inserción de un nucleótido en la secuencia de E4-3 lleva a un corrimiento del marco de lectura y resulta en cambios aminoacídicos a partir del aminoácido en la posición 14, y en la aparición de codones de terminación prematuros indicados con un asterisco (\*).

La inserción de una adenina en la posición 43 del exón 4 causa un corrimiento en el marco de lectura del gen y resulta en cambios aminoacídicos a partir del aminoácido 14. Entre otros cambios, se observa la aparición de codones de terminación prematuros en este exón en las posiciones 33, 41 y 43 (Figura 23B). Este resultado es consistente con los resultados observados en el análisis de expresión de CD147 e indican que la reducción de la expresión observada se puede deber a la degradación del ARNm por la vía de NMD. Además, la presencia de dos alelos en la secuencia del exón 4 en E4-3 es consistente con lo observado en el análisis de la expresión de proteína de CD147. Las células PC-3 son casi triploides y el cromosoma 13, donde se encuentra localizado CD147, puede variar entre monosómico y trisómico<sup>140</sup>. Es posible que en estas células ocurrió una edición parcial del gen resultando en que cuenten con al menos un alelo funcional del gen que produce proteína, aunque en menor cantidad que en las líneas no editadas. Alternativamente, podría ser que las células PC3 no son viables si no cuentan con al menos un alelo funcional de CD147, resultando en una selección negativa de las células mutantes *knock-out*. Otra posible explicación es que la edición ocurriera luego de la selección de clones y que la población de esta línea celular no sea homogénea. Esto podría ocurrir porque el sistema de edición génica utilizado se expresa continuamente. De ser este el caso, existirían células en la población que no fueron editadas y que por lo tanto expresan la proteína CD147.



---

## 5. Conclusiones y perspectivas.

Existe discusión sobre el rol de CD147 en cáncer de próstata. Muchos trabajos encuentran una correlación positiva entre la expresión de CD147 y la malignidad. En la mayoría de estos trabajos la expresión de CD147 se determina de manera semicuantitativa mediante inmunohistoquímica. Esta técnica se basa en el uso de anticuerpos para detectar la presencia de la proteína CD147 en células de biopsias de tejido prostático. Mediante observación en microscopio, se cuenta manualmente el porcentaje de células marcadas por el anticuerpo y se clasifican las muestras de manera cualitativa en “positivas” o “negativas”. En algunos trabajos se toma en cuenta la intensidad del marcado y se clasifica el tejido de manera semicuantitativa en cuatro niveles de expresión (negativo, bajo, intermedio y alto).

Esta metodología está sujeta a sesgos debidos a diferencias entre observadores o en la sensibilidad de los anticuerpos usados. A su vez, distintos trabajos usan distintos criterios para determinar si una muestra tiene expresión positiva o negativa de CD147 y tampoco existe un criterio universal para definir qué se considera alto o bajo con respecto a las variables patológicas (puntaje de Gleason, estadio TNM, valor PSA, etc.). Esto dificulta comparar resultados entre distintos estudios y puede explicar algunas de las diferencias entre sus conclusiones <sup>81,82</sup>.

Al no ser cuantitativo, esta metodología puede esconder variaciones en los patrones de expresión de CD147. A modo de ejemplo, una célula con expresión “positiva” de CD147 puede tener diferencias cuantitativas en su expresión con respecto a otra célula igualmente clasificada y esta diferencia no sería detectada por esta metodología. De esta manera un tejido puede presentar un pequeño número de células que expresan niveles altos de CD147 y ser marcado como negativo en su expresión de CD147, mientras que otro tejido puede tener muchas células con baja expresión de CD147 y ser marcado como positivo.

En este trabajo se utilizaron datos de expresión cuantitativos obtenidos por la cuantificación de ARNm mediante RNAseq. Se realizó un meta-análisis de expresión génica del gen *CD147* utilizando bases de datos *online* de acceso libre. Se encontró que la expresión de *CD147* se encuentra elevada tanto en muestras de tumores de cáncer de próstata como en otros tipos de tumores, cuando se compara con tejido no

---

tumoral. Aunque se encontró una correlación entre el nivel de expresión de *CD147* y el estado de metástasis, esta correlación desaparece cuando se analizan muestras pareadas. Para cáncer de próstata, no se contó con suficiente cantidad de muestras en la base de datos y no se logró realizar un análisis comparativo entre metástasis y tumor sólido.

Los resultados apoyan las observaciones anteriores de que la expresión de *CD147* se encuentra elevada en cáncer de próstata<sup>74-78</sup>. A pesar de que se encontró una correlación entre la expresión de *CD147* y metástasis para cáncer en general, esta correlación no se pudo demostrar para cáncer de próstata. A diferencia de observaciones anteriores<sup>74-76,80</sup>, en este trabajo no se encontró una relación entre el nivel de expresión de *CD147* y la supervivencia de los pacientes.

Este análisis tiene la fortaleza de que mide la expresión de *CD147* de manera cuantitativa utilizando una base de datos de expresión que fue normalizada para disminuir los efectos causados por diferencias metodológicas en el muestreo y secuenciación. Por otro lado, como se mencionó anteriormente, existen problemas en la comparación entre muestras provenientes del tumor y muestras de tejido normal adyacente. En particular para *CD147* se ha visto que las células tumorales pueden usar *CD147* para inducir la expresión de *CD147* en células del tejido circundante. Esto implica que el tejido normal cercano al tumor puede tener una expresión de *CD147* mayor a lo que tendría en un tejido normal de un paciente sano y esto podría estar escondiendo diferencias en la expresión entre estos tejidos. Sin embargo, a pesar de ese efecto, se observó una diferencia significativa en la expresión de *CD147* entre el tejido normal y el tumor. A su vez, el uso de la base de datos GTEx como una línea de base de expresión normal disminuyó este efecto de la influencia tumoral y se observó una diferencia aún mayor.

Es importante mencionar que en el análisis de las muestras pareadas no se tomó en cuenta el momento de toma de muestra. Si las muestras se tomaron cuando el paciente ya había desarrollado metástasis, entonces no sorprendería que el tumor sólido tenga valores elevados de expresión de genes relacionados al desarrollo de metástasis.

---

En el presente trabajo se generó una línea celular tumoral prostática metastásica *knock-out* para el gen *CD147*, conjuntamente con su línea control transfectada con el vector vacío. Utilizando un sistema de edición CRISPR-Cas9 se logró generar una línea celular que porta una mutación *frameshift* en el exón 4 del gen *CD147*. Esta mutación resulta, entre otros cambios, en la introducción de un codón de terminación prematuro en el exón 4. En la línea celular producida se observó una reducción de la expresión de *CD147* tanto a nivel de transcripto de ARNm como de proteína.

A nivel de secuencia, se observaron al menos dos alelos del exón 4 de *CD147* en la línea celular, uno de los cuales porta la mutación *frameshift*. Una posible explicación para esta observación es que la línea celular es heterocigota portando tanto alelos funcionales como no funcionales. Alternativamente podría ser que la población celular analizada es heterogénea, conviviendo en el cultivo células con ambos alelos mutados y células *wildtype*. Sin embargo, esta alternativa es poco probable ya que se empleó *cloning rings* para aislar los clones luego de la selección.

A futuro, nos planteamos verificar si las células son heterogéneas o si se trata de una línea heterocigota. Proponemos aislar células individuales mediante la dilución del cultivo de células hasta alcanzar una concentración tal que se pueda sembrar una célula por pocillo en una placa de 96 pocillos. Luego, se dejarán crecer los cultivos hasta alcanzar 100% de confluencia y se procederá a secuenciar el exón 4 de *CD147*.

La línea celular generada en este trabajo se produjo con el objetivo de estudiar el rol de *CD147* como modulador de MMP y su papel en cáncer de próstata metastásico. Con este objetivo, nos proponemos a futuro evaluar la capacidad de migración e invasión de la línea *CD147 knock-out* y compararla con la línea control y la línea PC3 *wildtype*. La capacidad de migración de las líneas celulares producidas se evaluará mediante un ensayo de cicatrización y se comparará con la línea PC3 *wildtype*. La capacidad de invasión de la línea *CD147 knock-out* y de la línea control se determinará mediante ensayos de *transwell* recubiertos con membrana basal. Por último, proponemos analizar mediante RT-qPCR y *western blot* la expresión de algunas MMPs (por ejemplo, MT1-MMP, MMP2, y MMP9) en las líneas generadas, ya que está ampliamente descrito que la proteína *CD147* regula la actividad de ellas

---

## 6. Bibliografía

1. Pernar CH, Ebot EM, Wilson KM, Mucci LA. The Epidemiology of Prostate Cancer. *Cold Spring Harb Perspect Med.* el 3 de diciembre de 2018;8(12).
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians.* 2018;68(6):394–424.
3. Fitzmaurice C, Abate D, Abbasi N, Abbastabar H, Abd-Allah F, Abdel-Rahman O, et al. Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-Years for 29 Cancer Groups, 1990 to 2017. *JAMA Oncol.* diciembre de 2019;5(12):1749–68.
4. Situación Epidemiológica del Uruguay en relación al Cáncer [Internet]. Comisión Honoraria de Lucha contra el Cáncer; mayo de 2021 [citado el 27 de octubre de 2021]. Disponible en: <https://www.comisioncancer.org.uy/Ocultas/Situacion-Epidemiologica-del-Uruguay-en-relacion-al-Cancer--Mayo-2021-uc108>
5. Schatten H. Brief Overview of Prostate Cancer Statistics, Grading, Diagnosis and Treatment Strategies. En: Schatten H, editor. *Cell & Molecular Biology of Prostate Cancer: Updates, Insights and New Frontiers* [Internet]. Cham: Springer International Publishing; 2018 [citado el 29 de octubre de 2021]. p. 1–14. (Advances in Experimental Medicine and Biology). Disponible en: [https://doi.org/10.1007/978-3-319-95693-0\\_1](https://doi.org/10.1007/978-3-319-95693-0_1)
6. Amin MB, Edge SB, Greene FL, Byrd DR, Brookland RK, Washington MK, et al. *AJCC Cancer Staging Manual.* 8a ed. Springer International Publishing; 2018. 1032 p.
7. Gleason DF. Classification of prostatic carcinomas. *Cancer Chemother Rep.* marzo de 1966;50(3):125–8.
8. Weinberg RA. Chapter 14: Moving Out: Invasion and Metastasis. En: *The biology of cancer.* Second edition. New York: Garland Science, Taylor & Francis Group; 2014.
9. Madzharova E, Kastl P, Sabino F, auf dem Keller U. Post-Translational Modification-Dependent Activity of Matrix Metalloproteinases. *Int J Mol Sci.* el 24 de junio de 2019;20(12):3077.
10. Quintero-Fabián S, Arreola R, Becerril-Villanueva E, Torres-Romero JC, Arana-Argáez V, Lara-Riegos J, et al. Role of Matrix Metalloproteinases in Angiogenesis and Cancer. *Front Oncol* [Internet]. 2019 [citado el 12 de mayo de 2020];9. Disponible en: <https://www.frontiersin.org/articles/10.3389/fonc.2019.01370/full>
11. Gonzalez-Avila G, Sommer B, Mendoza-Posada DA, Ramos C, Garcia-Hernandez AA, Falfan-Valencia R. Matrix metalloproteinases participation in the

- 
- metastatic process and their diagnostic and therapeutic applications in cancer. *Critical Reviews in Oncology/Hematology*. el 1 de mayo de 2019;137:57–83.
12. Young D, Das N, Anowai A, Dufour A. Matrix Metalloproteases as Influencers of the Cells' Social Media. *Int J Mol Sci*. el 7 de agosto de 2019;20(16):3847.
  13. Ågren MS, auf dem Keller U. Matrix Metalloproteinases: How Much Can They Do? *Int J Mol Sci*. el 12 de abril de 2020;21(8):2678.
  14. Kessenbrock K, Plaks V, Werb Z. Matrix Metalloproteinases: Regulators of the Tumor Microenvironment. *Cell*. el 2 de abril de 2010;141(1):52–67.
  15. Shay G, Lynch CC, Fingleton B. Moving targets: Emerging roles for MMPs in cancer progression and metastasis. *Matrix Biology*. el 1 de mayo de 2015;44–46:200–6.
  16. Roy R, Morad G, Jedinak A, Moses MA. Metalloproteinases and their roles in human cancer. *The Anatomical Record*. 2020;303(6):1557–72.
  17. Allard JB, Duan C. IGF-Binding Proteins: Why Do They Exist and Why Are There So Many? *Frontiers in Endocrinology*. el 9 de abril de 2018;9:117.
  18. Gong Y, Chippada-Venkata UD, Oh WK. Roles of Matrix Metalloproteinases and Their Natural Inhibitors in Prostate Cancer Progression. *Cancers (Basel)*. el 27 de junio de 2014;6(3):1298–327.
  19. Zheng G, Lyons JG, Tan TK, Wang Y, Hsu T-T, Min D, et al. Disruption of E-Cadherin by Matrix Metalloproteinase Directly Mediates Epithelial-Mesenchymal Transition Downstream of Transforming Growth Factor- $\beta$ 1 in Renal Tubular Epithelial Cells. *Am J Pathol*. agosto de 2009;175(2):580–91.
  20. De Wever O, Derycke L, Hendrix A, De Meerleer G, Godeau F, Depypere H, et al. Soluble cadherins as cancer biomarkers. *Clin Exp Metastasis*. el 19 de octubre de 2007;24(8):685–97.
  21. Egeblad M, Werb Z. New functions for the matrix metalloproteinases in cancer progression. *Nat Rev Cancer*. marzo de 2002;2(3):161–74.
  22. Gobin E, Bagwell K, Wagner J, Mysona D, Sandirasegarane S, Smith N, et al. A pan-cancer perspective of matrix metalloproteases (MMP) gene expression profile and their diagnostic/prognostic potential. *BMC Cancer*. el 14 de junio de 2019;19(1):581.
  23. Binder MJ, Ward AC. The Role of the Metzincin Superfamily in Prostate Cancer Progression: A Systematic-Like Review. *Int J Mol Sci*. el 30 de marzo de 2021;22(7):3608.
  24. Decock J, Thirkettle S, Wagstaff L, Edwards DR. Matrix metalloproteinases: protective roles in cancer. *J Cell Mol Med*. junio de 2011;15(6):1254–65.

25. López-Otín C, Palavalli LH, Samuels Y. Protective roles of matrix metalloproteinases. *Cell Cycle*. el 15 de noviembre de 2009;8(22):3657–62.
26. Soria-Valles C, Gutiérrez-Fernández A, Guiu M, Mari B, Fueyo A, Gomis RR, et al. The anti-metastatic activity of collagenase-2 in breast cancer cells is mediated by a signaling pathway involving decorin and miR-21. *Oncogene*. el 5 de junio de 2014;33(23):3054–63.
27. Andarawewa KL, Boulay A, Masson R, Mathelin C, Stoll I, Tomasetto C, et al. Dual stromelysin-3 function during natural mouse mammary tumor virus-ras tumor progression. *Cancer Res*. el 15 de septiembre de 2003;63(18):5844–9.
28. Brasse D, Mathelin C, Leroux K, Chenard M-P, Blaise S, Stoll I, et al. Matrix metalloproteinase 11/stromelysin-3 exerts both activator and repressor functions during the hematogenous metastatic process in mice. *Int J Cancer*. el 1 de septiembre de 2010;127(6):1347–55.
29. Wu E, Mari BP, Wang F, Anderson IC, Sunday ME, Shipp MA. Stromelysin-3 suppresses tumor cell apoptosis in a murine model. *J Cell Biochem*. 2001;82(4):549–55.
30. Lim M, Martinez T, Jablons D, Cameron R, Guo H, Toole B, et al. Tumor-derived EMMPRIN (extracellular matrix metalloproteinase inducer) stimulates collagenase transcription through MAPK p38. *FEBS Letters*. 1998;441(1):88–92.
31. Guo H, Zucker S, Gordon MK, Toole BP, Biswas C. Stimulation of matrix metalloproteinase production by recombinant extracellular matrix metalloproteinase inducer from transfected Chinese hamster ovary cells. *J Biol Chem*. el 3 de enero de 1997;272(1):24–7.
32. Biswas C, Zhang Y, DeCastro R, Guo H, Nakamura T, Kataoka H, et al. The human tumor cell-derived collagenase stimulatory factor (renamed EMMPRIN) is a member of the immunoglobulin superfamily. *Cancer Res*. el 15 de enero de 1995;55(2):434–9.
33. Muramatsu T. Basigin (CD147), a multifunctional transmembrane glycoprotein with various binding partners. *J Biochem*. mayo de 2016;159(5):481–90.
34. Muramatsu T, Miyauchi T. Basigin (CD147): a multifunctional transmembrane protein involved in reproduction, neural function, inflammation and tumor invasion. *Histol Histopathol*. julio de 2003;18(3):981–7.
35. Grass GD, Toole BP. How, with whom and when: an overview of CD147-mediated regulatory networks influencing matrix metalloproteinase activity. *Biosci Rep [Internet]*. el 15 de enero de 2016 [citado el 9 de abril de 2020];36(1). Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4718507/>
36. Biswas C. Collagenase stimulation in cocultures of human fibroblasts and human tumor cells. *Cancer Lett*. septiembre de 1984;24(2):201–7.

37. Kataoka H, DeCastro R, Zucker S, Biswas C. Tumor cell-derived collagenase-stimulatory factor increases expression of interstitial collagenase, stromelysin, and 72-kDa gelatinase. *Cancer Res.* el 1 de julio de 1993;53(13):3154–8.
38. Caudroy S, Polette M, Nawrocki-Raby B, Cao J, Toole BP, Zucker S, et al. EMMPRIN-mediated MMP regulation in tumor and endothelial cells. *Clin Exp Metastasis.* diciembre de 2002;19(8):697–702.
39. Kanekura T, Chen X, Kanzaki T. Basigin (CD147) is expressed on melanoma cells and induces tumor cell invasion by stimulating production of matrix metalloproteinases by fibroblasts. *Int J Cancer.* el 1 de junio de 2002;99(4):520–8.
40. Yang J-M, Xu Z, Wu H, Zhu H, Wu X, Hait WN. Overexpression of extracellular matrix metalloproteinase inducer in multidrug resistant cancer cells. *Mol Cancer Res.* abril de 2003;1(6):420–7.
41. Sameshima T, Nabeshima K, Toole BP, Yokogami K, Okada Y, Goya T, et al. Glioma cell extracellular matrix metalloproteinase inducer (EMMPRIN) (CD147) stimulates production of membrane-type matrix metalloproteinases and activated gelatinase A in co-cultures with brain-derived fibroblasts. *Cancer Lett.* el 1 de septiembre de 2000;157(2):177–84.
42. Tang Y, Kesavan P, Nakada MT, Yan L. Tumor-Stroma Interaction: Positive Feedback Regulation of Extracellular Matrix Metalloproteinase Inducer (EMMPRIN) Expression and Matrix Metalloproteinase-Dependent Generation of Soluble EMMPRIN. *Mol Cancer Res.* el 1 de febrero de 2004;2(2):73–80.
43. Schmidt R, Bültmann A, Ungerer M, Joghetaei N, Bülbül O, Thieme S, et al. Extracellular matrix metalloproteinase inducer regulates matrix metalloproteinase activity in cardiovascular cells: implications in acute myocardial infarction. *Circulation.* el 14 de febrero de 2006;113(6):834–41.
44. Zhu P, Ding J, Zhou J, Dong W-J, Fan C-M, Chen Z-N. Expression of CD147 on monocytes/macrophages in rheumatoid arthritis: its potential role in monocyte accumulation and matrix metalloproteinase production. *Arthritis Research & Therapy.* el 23 de junio de 2005;7(5):R1023.
45. Amit-Cohen B-C, Rahat MM, Rahat MA. Tumor cell-macrophage interactions increase angiogenesis through secretion of EMMPRIN. *Front Physiol.* el 12 de julio de 2013;4:178.
46. Grass GD, Bratoeva M, Toole BP. Regulation of invadopodia formation and activity by CD147. *J Cell Sci.* el 1 de febrero de 2012;125(3):777–88.
47. Li R, Huang L, Guo H, Toole BP. Basigin (murine EMMPRIN) stimulates matrix metalloproteinase production by fibroblasts. *Journal of Cellular Physiology.* 2001;186(3):371–9.

48. Sun J, Hemler ME. Regulation of MMP-1 and MMP-2 Production through CD147/Extracellular Matrix Metalloproteinase Inducer Interactions. *Cancer Res.* el 3 de marzo de 2001;61(5):2276–81.
49. Toole BP. The CD147-HYALURONAN Axis in Cancer. *The Anatomical Record.* 2020;303(6):1573–83.
50. Tang W, Chang SB, Hemler ME. Links between CD147 Function, Glycosylation, and Caveolin-1. *MBoC.* el 1 de septiembre de 2004;15(9):4043–50.
51. Von Ungern-Sternberg SNI, Zerneck A, Seizer P. Extracellular Matrix Metalloproteinase Inducer EMMPRIN (CD147) in Cardiovascular Disease. *International Journal of Molecular Sciences.* febrero de 2018;19(2):507.
52. Cui H-Y, Guo T, Wang S-J, Zhao P, Dong Z-S, Zhang Y, et al. Dimerization is essential for HAb18G/CD147 promoting tumor invasion via MAPK pathway. *Biochem Biophys Res Commun.* el 16 de marzo de 2012;419(3):517–22.
53. Egawa N, Koshikawa N, Tomari T, Nabeshima K, Isobe T, Seiki M. Membrane type 1 matrix metalloproteinase (MT1-MMP/MMP-14) cleaves and releases a 22-kDa extracellular matrix metalloproteinase inducer (EMMPRIN) fragment from tumor cells. *J Biol Chem.* el 8 de diciembre de 2006;281(49):37576–85.
54. Taylor PM, Woodfield RJ, Hodgkin MN, Pettitt TR, Martin A, Kerr DJ, et al. Breast cancer cell-derived EMMPRIN stimulates fibroblast MMP2 release through a phospholipase A(2) and 5-lipoxygenase catalyzed pathway. *Oncogene.* el 22 de agosto de 2002;21(37):5765–72.
55. Aoki M, Koga K, Hamasaki M, Egawa N, Nabeshima K. Emmprin, released as a microvesicle in epithelioid sarcoma, interacts with fibroblasts. *International Journal of Oncology.* el 1 de junio de 2017;50(6):2229–35.
56. Redzic JS, Kendrick AA, Bahmed K, Dahl KD, Pearson CG, Robinson WA, et al. Extracellular Vesicles Secreted from Cancer Cell Lines Stimulate Secretion of MMP-9, IL-6, TGF- $\beta$ 1 and EMMPRIN. *PLOS ONE.* el 1 de agosto de 2013;8(8):e71225.
57. Landras A, Reger de Moura C, Jouenne F, Lebbe C, Menashi S, Mourah S. CD147 Is a Promising Target of Tumor Progression and a Prognostic Biomarker. *Cancers (Basel)* [Internet]. el 16 de noviembre de 2019 [citado el 5 de junio de 2020];11(11). Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6896083/>
58. Guindolet D, Gabison EE. Role of CD147 (EMMPRIN/Basigin) in Tissue Remodeling. *The Anatomical Record.* 2020;303(6):1584–9.
59. Wang S-J, Cui H-Y, Liu Y-M, Zhao P, Zhang Y, Fu Z-G, et al. CD147 promotes Src-dependent activation of Rac1 signaling through STAT3/DOCK8 during the motility of hepatocellular carcinoma cells. *Oncotarget.* el 1 de enero de 2015;6(1):243–57.



- 
60. Bougatef F, Quemener C, Kellouche S, Naïmi B, Podgorniak M-P, Millot G, et al. EMMPRIN promotes angiogenesis through hypoxia-inducible factor-2 $\alpha$ -mediated regulation of soluble VEGF isoforms and their receptor VEGFR-2. *Blood*. el 24 de diciembre de 2009;114(27):5547–56.
  61. Tang Y, Nakada MT, Kesavan P, McCabe F, Millar H, Rafferty P, et al. Extracellular matrix metalloproteinase inducer stimulates tumor angiogenesis by elevating vascular endothelial cell growth factor and matrix metalloproteinases. *Cancer Res*. el 15 de abril de 2005;65(8):3193–9.
  62. Khayati F, Pérez-Cano L, Maouche K, Sadoux A, Boutalbi Z, Podgorniak M-P, et al. EMMPRIN/CD147 is a novel coreceptor of VEGFR-2 mediating its activation by VEGF. *Oncotarget*. 2015;6(12):9766–80.
  63. Wilson MC, Meredith D, Fox JEM, Manoharan C, Davies AJ, Halestrap AP. Basigin (CD147) is the target for organomercurial inhibition of monocarboxylate transporter isoforms 1 and 4: the ancillary protein for the insensitive MCT2 is EMBIGIN (gp70). *J Biol Chem*. el 22 de julio de 2005;280(29):27213–21.
  64. Kirk P, Wilson MC, Heddle C, Brown MH, Barclay AN, Halestrap AP. CD147 is tightly associated with lactate transporters MCT1 and MCT4 and facilitates their cell surface expression. *EMBO J*. el 1 de agosto de 2000;19(15):3896–904.
  65. Le Floch R, Chiche J, Marchiq I, Naiken T, Ilc K, Murray CM, et al. CD147 subunit of lactate/H<sup>+</sup> symporters MCT1 and hypoxia-inducible MCT4 is critical for energetics and growth of glycolytic tumors. *Proc Natl Acad Sci U S A*. el 4 de octubre de 2011;108(40):16663–8.
  66. Riethdorf S, Reimers N, Assmann V, Kornfeld J-W, Terracciano L, Sauter G, et al. High incidence of EMMPRIN expression in human tumors. *International Journal of Cancer*. 2006;119(8):1800–10.
  67. Nabeshima K, Iwasaki H, Koga K, Hojo H, Suzumiya J, Kikuchi M. Emmprin (basigin/CD147): matrix metalloproteinase modulator and multifunctional cell recognition molecule that plays a critical role in cancer progression. *Pathol Int*. julio de 2006;56(7):359–67.
  68. Li Y, Xu J, Chen L, Chen L, Zhong W-D, Zhang Z, et al. HAb18G (CD147), a cancer-associated biomarker and its role in cancer detection. *Histopathology*. mayo de 2009;54(6):677–87.
  69. Xin X, Zeng X, Gu H, Li M, Tan H, Jin Z, et al. CD147/EMMPRIN overexpression and prognosis in cancer: A systematic review and meta-analysis. *Sci Rep*. el 9 de septiembre de 2016;6:32804.
  70. Gabison EE, Hoang-Xuan T, Mauviel A, Menashi S. EMMPRIN/CD147, an MMP modulator in cancer, development and tissue repair. *Biochimie*. abril de 2005;87(3–4):361–8.

71. Fan H, Yi W, Wang C, Wang J. The clinicopathological significance and prognostic value of EMMPRIN overexpression in cancers: evidence from 39 cohort studies. *Oncotarget*. el 31 de julio de 2017;8(47):82643–60.
72. Reimers N, Zafrakas K, Assmann V, Egen C, Riethdorf L, Riethdorf S, et al. Expression of extracellular matrix metalloproteases inducer on micrometastatic and primary mammary carcinoma cells. *Clin Cancer Res*. el 15 de mayo de 2004;10(10):3422–8.
73. Klein CA, Seidl S, Petat-Dutter K, Offner S, Geigl JB, Schmidt-Kittler O, et al. Combined transcriptome and genome analysis of single micrometastatic cells. *Nat Biotechnol*. abril de 2002;20(4):387–92.
74. Zhong W, Han Z, He H, Bi X, Dai Q, Zhu G, et al. CD147, MMP-1, MMP-2 and MMP-9 protein expression as significant prognostic factors in human prostate cancer. *Oncology*. 2008;75(3–4):230–6.
75. Han Z, Bi X, Qin W, He H, Dai Q, Zou J, et al. CD147 expression indicates unfavourable prognosis in prostate cancer. *Pathol Oncol Res*. septiembre de 2009;15(3):369–74.
76. Han Z, He H, Bi X, Qin W, Dai Q, Zou J, et al. Expression and clinical significance of CD147 in genitourinary carcinomas. *J Surg Res*. el 15 de mayo de 2010;160(2):260–7.
77. Bi X-C, Liu J-M, Zheng X-G, Xian Z-Y, Feng Z-W, Lou Y-X, et al. Over-expression of extracellular matrix metalloproteinase inducer in prostate cancer is associated with high risk of prostate-specific antigen relapse after radical prostatectomy. *Clin Invest Med*. el 1 de diciembre de 2011;34(6):E358.
78. Madigan MC, Kingsley EA, Cozzi PJ, Delprado WJ, Russell PJ, Li Y. The role of extracellular matrix metalloproteinase inducer protein in prostate cancer progression. *Cancer Immunol Immunother*. septiembre de 2008;57(9):1367–79.
79. Hao J, Madigan MC, Khatri A, Power CA, Hung T-T, Beretov J, et al. In vitro and in vivo prostate cancer metastasis and chemoresistance can be modulated by expression of either CD44 or CD147. *PLoS One*. 2012;7(8):e40716.
80. Zhong W, Liang Y-X, Lin SX, Li L, He H-C, Bi X-C, et al. Expression of CD147 is associated with prostate cancer progression. *Int J Cancer*. el 15 de enero de 2012;130(2):300–8.
81. Bauman TM, Ewald JA, Huang W, Ricke WA. CD147 expression predicts biochemical recurrence after prostatectomy independent of histologic and pathologic features. *BMC Cancer*. el 25 de julio de 2015;15(1):549.
82. Grupp K, Höhne TS, Prien K, Hube-Magg C, Tsourlakis MC, Sirma H, et al. Reduced CD147 expression is linked to ERG fusion-positive prostate cancers but lacks substantial impact on PSA recurrence in patients treated by radical prostatectomy. *Exp Mol Pathol*. octubre de 2013;95(2):227–34.

83. Pérttega-Gomes N, Vizcaíno JR, Miranda-Gonçalves V, Pinheiro C, Silva J, Pereira H, et al. Monocarboxylate transporter 4 (MCT4) and CD147 overexpression is associated with poor prognosis in prostate cancer. *BMC Cancer*. el 25 de julio de 2011;11:312.
84. Gaj T, Gersbach CA, Barbas CF. ZFN, TALEN and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol*. julio de 2013;31(7):397–405.
85. Lewis BP, Green RE, Brenner SE. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc Natl Acad Sci U S A*. el 7 de enero de 2003;100(1):189–92.
86. Li H, Yang Y, Hong W, Huang M, Wu M, Zhao X. Applications of genome editing technology in the targeted therapy of human diseases: mechanisms, advances and prospects. *Signal Transduct Target Ther*. el 3 de enero de 2020;5:1.
87. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. *Nature Protocols*. noviembre de 2013;8(11):2281–308.
88. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*. el 17 de agosto de 2012;337(6096):816–21.
89. Doudna JA, Charpentier E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science*. el 28 de noviembre de 2014;346(6213):1258096.
90. Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J Bacteriol*. diciembre de 1987;169(12):5429–33.
91. Mojica FJ, Díez-Villaseñor C, Soria E, Juez G. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol Microbiol*. abril de 2000;36(1):244–6.
92. Jansen R, Embden JDA van, Gaastra W, Schouls LM. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol*. marzo de 2002;43(6):1565–75.
93. Mojica FJM, Díez-Villaseñor C, García-Martínez J, Soria E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol*. febrero de 2005;60(2):174–82.
94. Pourcel C, Salvignol G, Vergnaud G. CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology (Reading)*. marzo de 2005;151(Pt 3):653–63.

- 
95. Bolotin A, Quinquis B, Sorokin A, Ehrlich SD. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology (Reading)*. agosto de 2005;151(Pt 8):2551–61.
  96. Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct*. el 16 de marzo de 2006;1:7.
  97. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Science*. el 23 de marzo de 2007;315(5819):1709–12.
  98. Deveau H, Barrangou R, Garneau JE, Labonté J, Fremaux C, Boyaval P, et al. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J Bacteriol*. febrero de 2008;190(4):1390–400.
  99. Mojica FJM, Díez-Villaseñor C, García-Martínez J, Almendros C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology (Reading)*. marzo de 2009;155(Pt 3):733–40.
  100. Garneau JE, Dupuis M-È, Villion M, Romero DA, Barrangou R, Boyaval P, et al. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*. el 4 de noviembre de 2010;468(7320):67–71.
  101. Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*. el 31 de marzo de 2011;471(7340):602–7.
  102. The Nobel Prize in Chemistry 2020 [Internet]. NobelPrize.org. 2020 [citado el 11 de diciembre de 2021]. Disponible en: <https://www.nobelprize.org/prizes/chemistry/2020/press-release/>
  103. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, et al. RNA-guided human genome engineering via Cas9. *Science*. el 15 de febrero de 2013;339(6121):823–6.
  104. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. *Science*. el 15 de febrero de 2013;339(6121):819–23.
  105. Pickar-Oliver A, Gersbach CA. The next generation of CRISPR–Cas technologies and applications. *Nat Rev Mol Cell Biol*. agosto de 2019;20(8):490–507.
  106. C MAZ, Gaj T. Next-Generation CRISPR Technologies and Their Applications in Gene and Cell Therapy. *Trends in Biotechnology* [Internet]. el 1 de diciembre de 2020 [citado el 10 de diciembre de 2020];0(0). Disponible en: [https://www.cell.com/trends/biotechnology/abstract/S0167-7799\(20\)30287-0](https://www.cell.com/trends/biotechnology/abstract/S0167-7799(20)30287-0)

107. Goldman M, Craft B, Hastie M, Repečka K, McDade F, Kamath A, et al. The UCSC Xena platform for public and private cancer genomics data visualization and interpretation. *bioRxiv*. el 26 de septiembre de 2019;326470.
108. Vivian J, Rao AA, Nothaft FA, Ketchum C, Armstrong J, Novak A, et al. Toil enables reproducible, open source, big biomedical data analyses. *Nature Biotechnology*. abril de 2017;35(4):314–6.
109. O’Leary NA, Wright MW, Brister JR, Ciuffo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*. el 4 de enero de 2016;44(D1):D733-745.
110. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*. el 18 de junio de 2012;13:134.
111. Optimized CRISPR Design [Internet]. 2017 [citado el 31 de julio de 2019]. Disponible en: <https://web.archive.org/web/20171008132606/http://crispr.mit.edu/>
112. Sanjana NE, Shalem O, Zhang F. Improved vectors and genome-wide libraries for CRISPR screening. *Nature Methods*. agosto de 2014;11(8):783–4.
113. Addgene: lentiCRISPR v2 [Internet]. [citado el 26 de mayo de 2020]. Disponible en: <https://www.addgene.org/52961/#>
114. Birnboim HC, Doly J. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res*. el 24 de noviembre de 1979;7(6):1513–23.
115. Kaighn ME, Narayan KS, Ohnuki Y, Lechner JF, Jones LW. Establishment and characterization of a human prostatic carcinoma cell line (PC-3). *Invest Urol*. julio de 1979;17(1):16–23.
116. Laemmli UK, Beguin F, Gujer-Kellenberger G. A factor preventing the major head protein of bacteriophage T4 from random aggregation. *J Mol Biol*. el 14 de enero de 1970;47(1):69–85.
117. Towbin H, Staehelin T, Gordon J. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc Natl Acad Sci U S A*. septiembre de 1979;76(9):4350–4.
118. Mahmood T, Yang P-C. Western Blot: Technique, Theory, and Trouble Shooting. *N Am J Med Sci*. septiembre de 2012;4(9):429–34.
119. Salinovich O, Montelaro RC. Reversible staining and peptide mapping of proteins transferred to nitrocellulose after separation by sodium dodecylsulfate-polyacrylamide gel electrophoresis. *Anal Biochem*. el 1 de agosto de 1986;156(2):341–7.

- 
120. Schindelin J, Rueden CT, Hiner MC, Eliceiri KW. The ImageJ ecosystem: An open platform for biomedical image analysis. *Molecular Reproduction and Development*. 2015;82(7–8):518–29.
  121. Dehairs J, Talebi A, Cherifi Y, Swinnen JV. CRISP-ID: decoding CRISPR mediated indels by Sanger sequencing. *Scientific Reports*. el 1 de julio de 2016;6(1):28973.
  122. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol*. el 1 de junio de 2018;35(6):1547–9.
  123. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. el 11 de noviembre de 1994;22(22):4673–80.
  124. Svec D, Tichopad A, Novosadova V, Pfaffl MW, Kubista M. How good is a PCR efficiency estimate: Recommendations for precise and robust qPCR efficiency assessments. *Biomol Detect Quantif*. el 11 de marzo de 2015;3:9–16.
  125. EcoStudy Software [Internet]. United States: Illumina, Inc; 2013 [citado el 23 de abril de 2019]. Disponible en: [https://support.illumina.com/downloads/eco\\_real\\_time\\_pcr\\_software\\_v5-0.html](https://support.illumina.com/downloads/eco_real_time_pcr_software_v5-0.html)
  126. Pfaffl MW. Quantification strategies in real-time PCR. En: *A-Z of quantitative PCR*. La Jolla, CA: International University Line; 2014. p. 87–120. (IUL biotechnology series; vol. 5).
  127. Goldman MJ, Craft B, Hastie M, Repečka K, McDade F, Kamath A, et al. Visualizing and interpreting cancer genomics data via the Xena platform. *Nature Biotechnology*. junio de 2020;38(6):675–8.
  128. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. el 4 de agosto de 2011;12(1):323.
  129. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*. el 5 de diciembre de 2014;15(12):550.
  130. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. enero de 2009;10(1):57–63.
  131. Hoadley KA, Yau C, Hinoue T, Wolf DM, Lazar AJ, Drill E, et al. Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell*. el 5 de abril de 2018;173(2):291-304.e6.
  132. Chai H, Brown RE. Field effect in cancer-an update. *Ann Clin Lab Sci*. 2009;39(4):331–7.

- 
133. Heaphy CM, Griffith JK, Bisoffi M. Mammary field cancerization: molecular evidence and clinical importance. *Breast Cancer Res Treat.* el 1 de noviembre de 2009;118(2):229–39.
  134. Zeng WZD, Glicksberg BS, Li Y, Chen B. Selecting precise reference normal tissue samples for cancer research using a deep learning approach. *BMC Medical Genomics.* el 31 de enero de 2019;12(1):21.
  135. Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology.* septiembre de 2013;31(9):827–32.
  136. Wu X, Kriz AJ, Sharp PA. Target specificity of the CRISPR-Cas9 system. *Quant Biol.* el 1 de junio de 2014;2(2):59–70.
  137. Kowalczykowski SC. Biochemical and biological function of Escherichia coli RecA protein: behavior of mutant RecA proteins. *Biochimie.* el 1 de febrero de 1991;73(2):289–304.
  138. Bai Y, Huang W, Ma L-T, Jiang J-L, Chen Z-N. Importance of N-Glycosylation on CD147 for Its Biological Functions. *Int J Mol Sci.* el 15 de abril de 2014;15(4):6356–77.
  139. Kendrick AA, Schafer J, Dzieciatkowska M, Nemkov T, D'Alessandro A, Neelakantan D, et al. CD147: a small molecule transporter ancillary protein at the crossroad of multiple hallmarks of cancer and metabolic reprogramming. *Oncotarget.* el 27 de diciembre de 2016;8(4):6742–62.
  140. Ohnuki Y, Marnell MM, Babcock MS, Lechner JF, Kaighn ME. Chromosomal Analysis of Human Prostatic Adenocarcinoma Cell Lines. *Cancer Res.* el 1 de marzo de 1980;40(3):524–34.