

Aproximación al estudio de las reglas de apareamiento microARN-ARNm

Tesis de Maestría en Ciencias Biológicas
Subárea - Biología Celular y Molecular
PEDECIBA - BIOLOGÍA
Lic. Juan Manuel Trinidad Barnech

Laboratorio de Interacciones Moleculares - Facultad de Ciencias
Departamento de Genética - Facultad de Medicina
Universidad de la República – Uruguay

Orientadora
Dra. María Ana Duhagon
Co-orientadora
Dra. Beatriz Garat

Indice

Resumen	5
Introducción.....	6
MicroARNs	6
Biogénesis de microARNs.....	7
Mecanismo de silenciamiento	7
Características de la interacción microARN-ARNm.....	9
Importancia de la región suplementaria en la interacción microARN-ARNm	11
Algoritmos de predicción de blancos de microARNs	13
Aproximaciones experimentales para el estudio de la interacción microARN-ARNm	15
Objetivo general.....	19
Objetivos específicos	19
Materiales y Métodos del objetivo específico 1	20
Datos transcriptómicos de microARNs y ARNm de TCGA	20
Identificación de interacciones microARN-ARNm en el PRAD-TCGA	20
Análisis de correlación entre la expresión de microARN y ARNm.....	21
Análisis de composición nucleotídica y enriquecimiento de motivos.....	22
Generación de programas ejecutables	22
Materiales y Métodos del objetivo específico 2	22
Electroforesis de proteínas y <i>Western-blot</i>	22
Plásmidos	25
Extracción y purificación de plásmidos	26
Transformación de bacterias	27
Digestión y ligación del cDNA del gen AGO2 plasmídico	28
PCR de colonias.....	28
Secuenciación del plásmido recombinante con AGO2-flag	29

<i>Líneas celulares</i>	29
<i>Propagación de los cultivos de líneas celulares</i>	30
Transfección de líneas celulares	30
<i>Descongelamiento de crioviales</i>	31
Cuantificación de ARNm de AGO2	31
Expresión de SaCas9 recombinante en <i>E. coli</i>	32
Purificación de SaCas9 recombinante.....	33
Diseño de sgARNs	34
Transcripción <i>in vitro</i> del sgARNs.....	35
Síntesis de ADN para ensayos de digestión <i>in vitro</i>	35
Purificación de productos de PCR.....	38
Electroforesis en gel de agarosa	38
Evaluación de la actividad de la SaCas9 <i>in vitro</i>	39
Resultados y discusión	40
Identificación de sitios de interacción microARN-ARNm en el transcriptoma PRAD-TCGA basado en la complementariedad de bases y las correlaciones de abundancia (OE 1).....	40
Validación de mecanismos represivos de los microARNs utilizando la actividad de inferida a partir de las correlaciones de expresión de microARN-ARNm en TCGA-PRAD.	43
Características de la interacción microARN-ARNm que involucra la región suplementaria.....	48
Diseño de Programa ejecutable para búsqueda de interacciones microARN-ARNm	55
Avances experimentales en metodologías de inmunoprecipitación de AGO2 para la validación futura de hipótesis surgidas de los análisis <i>in silico</i> de interacción microARN-ARNm (OE 2)	57

Anticuerpo Anti AGO2.....	59
Plásmido de expresión de AGO2 fusionada a un epítoto inmunogénico	61
Sistema CRISPR/Cas	67
Conclusiones y Perspectivas	75
Abordaje bioinformático	75
Abordaje experimental	76
Bibliografía	78
Material suplementario	83

Resumen

Los microARN son ARN monocatenarios pequeños (~21 nucleótidos) que regulan la expresión de ARN mensajeros (ARNm) a nivel postranscripcional esenciales en metazoos. En humanos el mecanismo de represión mediado por microARNs dominante actúa sin cortar el ARNm y requiere de la complementariedad perfecta de los nucleótidos 2-7 o 2-8 del microARN denominados región semilla. Dado que solo se requiere de 6-7 nucleótidos contiguos, es frecuente que existan secuencias complementarias en el transcriptoma no asociadas a la represión, lo que dificulta la identificación de ARNm blanco. En este trabajo evaluamos los mecanismos de interacción mediados por microARN utilizando un abordaje *in silico* utilizando la base de datos TCGA-PRAD (datos de tejido no perturbado) y avanzamos en la puesta a punto de abordajes experimentales para evaluar los resultados. Mediante el cálculo de correlaciones de expresión génica entre microARNs y ARNms logramos validar reglas establecidas y profundizar en el estudio de interacciones menos estudiadas. Resultado de esto, encontramos que la región suplementaria del microARN (nucleótidos 13-16) cumple un rol valioso en la represión al igual que los nucleótidos “puente”, identificados como aquellos entre la semilla y la región suplementaria (8-12 o 9-12, dependiendo de la semilla). Por último, avanzamos en la puesta a punto de tres aproximaciones experimentales para la inmunoprecipitación de la proteína AGO2, principal responsable del direccionamiento de la actividad del microARN, que podrán ser utilizadas para el estudio de los hallazgos antes mencionados.

Introducción

MicroARNs

Los microARN son ARN monocatenarios pequeños que regulan la expresión de ARN mensajeros (ARNm) a nivel postranscripcional, siendo fundamentales para la homeostasis fisiológica y el desarrollo en los metazoos (Bartel, 2018). Fueron descubiertos en *C. elegans* a mediados del año 2000 gracias a que los genes codificantes para los microARNs hasta entonces desconocidos *lin-4* y *let-7* cumplían un rol central en el desarrollo embrionario. Los investigadores descubrieron que estos genes, lejos de producir una proteína, generaban ARN no codificante incluyendo ARNs no codificantes pequeños de ~22 nt de largo (Lee et al., 1993; Reinhart et al., 2000). *lin-4* parecía ser específico de gusanos, sin embargo, *let-7* estaba presente en gusanos, moscas y humanos. Este descubrimiento impulsó el campo de microARN, y pronto se encontraron muchos microARNs adicionales en *C. elegans*, *D. melanogaster* y humanos (Berezikov, 2011). El descubrimiento de los pequeños ARNs explotó con el advenimiento de las tecnologías de secuenciación, donde además se descubrieron otro tipo de pequeños ARNs (Bartel, 2018). La vía de microARNs deriva de la aún más basal vía de ARN interferentes la cual, según estudios de genómica comparativa, parece haber estado presente en el último ancestro común de los eucariotas como mecanismo de defensa de virus y elementos móviles (Berezikov, 2011; Shabalina & Koonin, 2008). Sin embargo, dos proteínas esenciales para la vía de microARNs como DROSHA y PASHA (*Microprocessor*) son propias del grupo *Holozoa* por lo tanto la vía de los microARNs canónica parece restringirse a este grupo dejando fuera al reino *Fungi* y *Plantae* (Bråte et al., 2018; Moran et al., 2017). Otra diferencia central entre plantas y metazoarios es el mecanismo de reconocimiento del ARN blanco: en plantas se requiere extensa complementariedad, mientras que en metazoarios sólo complementariedad parcial (descrito más adelante) (Moran et al., 2017).

Biogénesis de microARNs

A pesar de que el proceso individual en cada especie puede diferir en algunos detalles, el proceso de biogénesis de los microARNs es ampliamente conservado evolutivamente (Figura 1A). El requisito central para un gen codificante de microARN es la correcta estructura secundaria. Sólo las horquillas que son reconocidas y procesadas por Drosha/Dicer pueden resultar en un silenciamiento funcional inducido por microARN (Berezikov, 2011). El proceso canónico comienza con los genes codificantes de microARNs siendo transcritos por la ARN polimerasa II y dando lugar a un transcrito primario (pri-microARNs) que presenta una estructura de tallo-horquilla. Posteriormente, los pri-microARNs son clivados por el complejo nuclear *Microprocessor* compuesto por DROSHA y DGCR8 (PASHA) para dar lugar al precursor del microARN (pre-microARNs) de aproximadamente 60 nucleótidos con estructura de horquilla. El precursor es exportado al citoplasma por la exportina 5 (XPO5) y procesados por DICER, generando los microARNs dúplex de aproximadamente 21–25 nucleótidos, que consisten en una hebra guía (referida como microARN) y una hebra pasajera (referida como microARN*). Existen algunos microARNs que escapan a la vía canónica de génesis sobre todo saltando el procesamiento por DROSHA. Por ejemplo, los microARNs derivados de intrones (mirtrons) (Figura 1B) donde el pri-microARNs es generado directamente por la maquinaria de *splicing*, evitando el procesamiento por DROSHA. Otra clase de microARNs que no utiliza la vía canónica son los shARNs (*short-hairpin* ARNs) endógenos que también utilizan una vía de maduración independiente de DROSHA. Una tercera clase comprende aquellos que se transcriben junto a, o como parte de, otro tipo de ARN pequeño (incluso ARN de transferencia), y dependen de los determinantes transcripcionales y del procesamiento del otro ARN pequeño para generar el sustrato de horquilla de DICER. Por último, también existen algunos pocos microARNs independientes de DICER como es el caso de miR-451 en humanos (Bartel, 2018).

Mecanismo de silenciamiento

El dúplex de microARN y microARN* es separado por una helicasa y una de las hebras es cargada en un complejo proteico, formando el denominado *RNA induced silencing*

complex (RISC), el cual es el responsable de la parte efectora del silenciamiento y está conformado principalmente por una de las proteínas de la familia Argonauta (AGO1-4), siendo la más estudiada la proteína AGO2. Este complejo regula la expresión de sus genes blanco mediante represión traduccional y/o degradación del ARNm, valiéndose de una de las hebras del microARN como guía y dependiendo del grado de complementariedad entre el microARN y su ARNm diana (Figura 1).

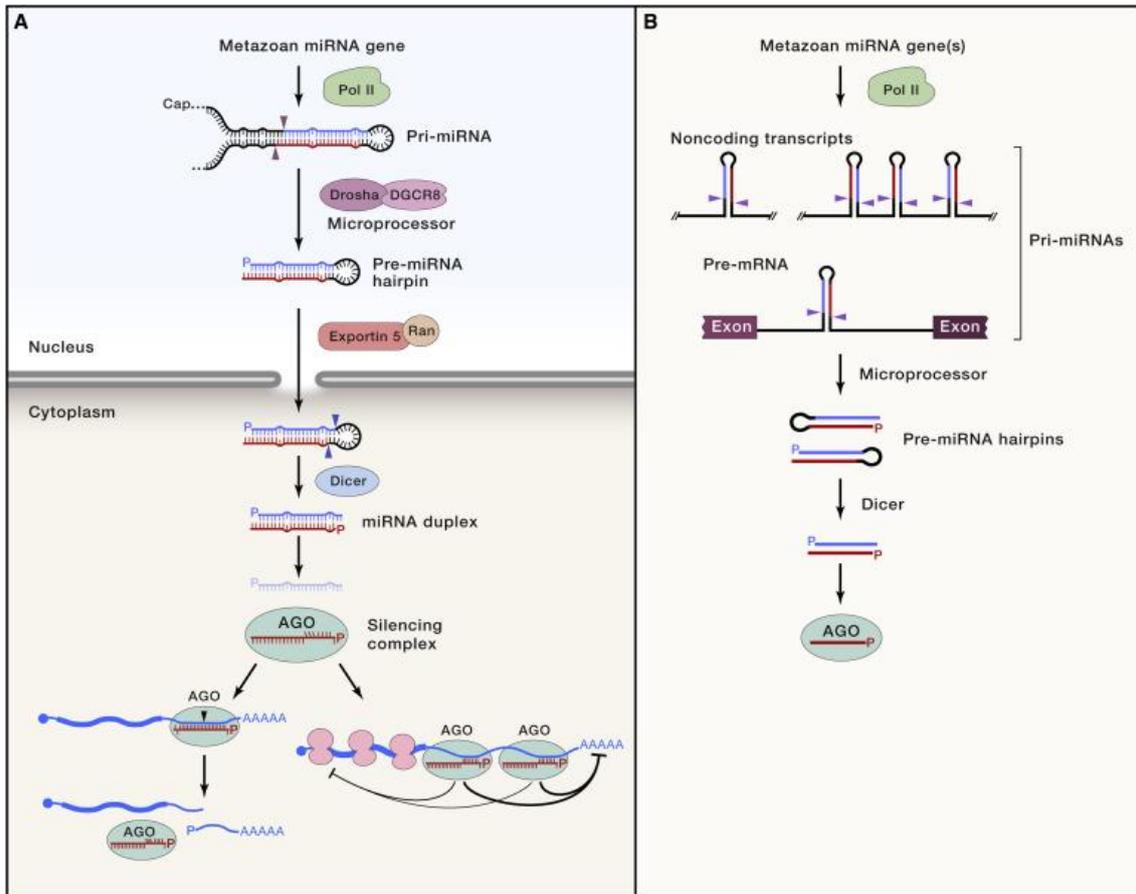


Figura 1. Biogénesis y función de los microARN canónicos de animales. 1) Biogénesis y función de un microARN canónico. *B)* Fuentes típicas por las que se generan de microARNs canónicos. Tomado de Bartel 2018.

En humanos el mecanismo de represión mediado por microARNs dominante actúa sin cortar el ARNm y no requiere extensa complementariedad (Bartel, 2018). De hecho, en los microARNs, la región semilla (nt 2-7 o 2-8, desde el extremo 5' del microARN) es la principal determinante para la eficacia y especificidad del direccionamiento (Agarwal et al., 2015; Grimson et al., 2007; McGeary et al., 2019). La identidad de los nucleótidos 2-8 (región semilla) es utilizada para agrupar a los microARNs en “familias de microARNs”,

donde todos los microARNs de una misma familia tienen la misma secuencia en la región semilla, denotando un vínculo funcional y no necesariamente evolutivo (aunque en la mayoría de los casos existe este vínculo) (Bartel, 2009). Los microARNs interactúan principalmente con las regiones 3'UTR y CDS de los ARNm diana y la represión afecta entre 20-50% de la expresión (Bartel, 2009; Grimson et al., 2007; Hafner et al., 2010; Helwak et al., 2013; Moore et al., 2015; Schnall-Levin et al., 2010). Mediante esta interacción los microARN junto con el complejo RISC silencian la expresión génica reprimiendo la traducción y promoviendo la degradación del ARNm blanco (Bartel, 2018; Jonas & Izaurralde, 2015). Existen distintos mecanismos por los cuales el complejo RISC regula a los ARNm blanco. Mientras que, en plantas la regulación reside en la capacidad de las proteínas AGO de cortar el ARNm, en mamíferos el mecanismo más aceptado es el reclutamiento de proteínas encargadas de quitar la caperuza 5' m7G o acortar la cola poliA de los ARNm (Bartel, 2018; Eulalio et al., 2008). La proteína encargada de ambos procesos es TNRC6 que interactúa con las PABP (*polyA-binding-protein*) además de reclutar complejos de deadenilasas (PAN2–PAN3, CCR4–NOT) encargados de acortar la cola poliA y de esta manera desestabilizar el ARNm desencadenando a su vez el mecanismo de *decapping* (Bartel, 2018).

Características de la interacción microARN-ARNm

Dado que la región de la semilla está compuesta por solo 6-7 nucleótidos contiguos, es frecuente que existan secuencias complementarias en el transcriptoma, pero no necesariamente asociadas a la represión mediada por microARN (Baek et al., 2008; Selbach et al., 2008; Yue et al., 2009). Existen 4 tipos de semillas canónicas que se ordenan en base a su eficacia media de la siguiente manera: 8mer >> 7mer-m8 > 7mer-A1 >> 6mer > no site. Es importante tener en cuenta que el contexto influye en gran medida en la represión, por lo tanto, a nivel individual, es decir si comparamos sitio a sitio este orden puede verse modificado (Bartel, 2009). A pesar de varios años de esfuerzo para comprender los factores y contextos subyacentes involucrados en la represión eficiente de ARNm, aún es un desafío identificar interacciones funcionalmente relevantes especialmente en mamíferos, tal como lo discute el Dr. David Bartel en su última revisión

sobre los microARNs (Bartel, 2018). En los estudios seminales sobre el mecanismo de interacción microARN-ARNm se describió que los nucleótidos fuera de la semilla pueden contribuir al reconocimiento del ARNm diana, aunque no se observó grandes contribuciones en la represión (Grimson et al., 2007). Por otro lado, estudios ómicos recientes sin modelos preestablecidos de interacción, vuelven a confirmar tanto que la semilla es la principal determinante de la interacción como que nucleótidos hacia el 3' del microARN contribuyen a la represión de un modo que parece ser más importante que lo previamente descrito (Broughton et al., 2016; Helwak et al., 2013; Moore et al., 2015). Los mecanismos de interacción que involucran la región 3' del microARN se clasifican como: suplementarios (si la semilla tiene complementariedad perfecta) o compensatorios (si la semilla tiene complementariedad imperfecta) (Bartel, 2018; Grimson et al., 2007) (Figura 2). La interacción suplementaria parece ser bastante específica de cada microARN e incluso permite distinguir blancos de microARNs de una misma familia (Broughton et al., 2016; Moore et al., 2015). Otro argumento a favor de la relevancia de la interacción suplementaria es el hallazgo de que los nucleótidos 13-16 del microARN son, después de los correspondientes a la semilla, los de mayor conservación evolutiva (Grimson et al., 2007) (Figura 2). También son los más importantes según los análisis de *Photoactivatable ribonucleoside enhanced cross-linking and immunoprecipitation* de AGO (AGO PAR-CLIP) realizados por otros investigadores (Hafner et al., 2010). Además de los sitios canónicos, existen otros, menos frecuentes y con poca evidencia de su actividad real como son las semillas centradas (Shin et al., 2010) o el extenso apareamiento 3' sin semilla en las CDS (Zhang et al., 2018), entre otras.

Muchos de los estudios antes mencionados en los cuales se describieron las reglas de la interacción microARN-ARNm involucran alguna perturbación o algún sesgo que puede conducir a conclusiones no generalizables. Entre ellos se encuentra el uso de pocos microARNs, la sobreexpresión de alguno de los componentes de RISC (generalmente AGO2 o microARNs), el uso de líneas celulares o plásmidos con *AGO2-tagged*, el análisis de las interacciones *in vitro* y en presencia de un número mínimo de componentes. Estos antecedentes nos llevaron a proponer el estudio de datos de tejidos no perturbados para evaluar las interacciones buscando esclarecer la existencia de reglas generales de reconocimiento de los microARNs y su eventual vínculo fisiológico

los primeros reportes, los autores demuestran que las interacciones suplementarias óptimas aumentan en gran medida la afinidad del microARN por su ARNm diana (Sheu-Gruttadauria et al., 2019). En este trabajo se descubren dos características de la interacción sumamente relevantes. La primera es que la semilla y la región suplementaria se pueden unir mediante un bucle no estructurado en el ARNm de 1-15 nucleótidos. La segunda es que en caso de existir complementariedad de bases en la región central del microARN (nucleótidos del 9 al 12), las moléculas de AGO2 mantienen una conformación estable que evita el apareamiento (Sheu-Gruttadauria et al., 2019). Estos resultados son consistentes con los análisis extensivos de PAR-CLIP donde se muestra que esta región del microARN parecería interaccionar preferentemente con la proteína AGO2 que con el ARNm (Hafner et al., 2010). Estas características no habían sido tomadas en cuenta en las evaluaciones previas de la región suplementaria y probablemente expliquen parte de la discordancia entre los hallazgos.

Se ha demostrado que la arquitectura de la interacción con la región 3' del microARN depende principalmente de su secuencia (Mcgeary et al., 2021; Sheu-Gruttadauria et al., 2019) pero los autores sugieren que, para comprender los detalles influyentes de dicha arquitectura, son necesarios análisis con mayor número de microARNs (Mcgeary et al., 2021).

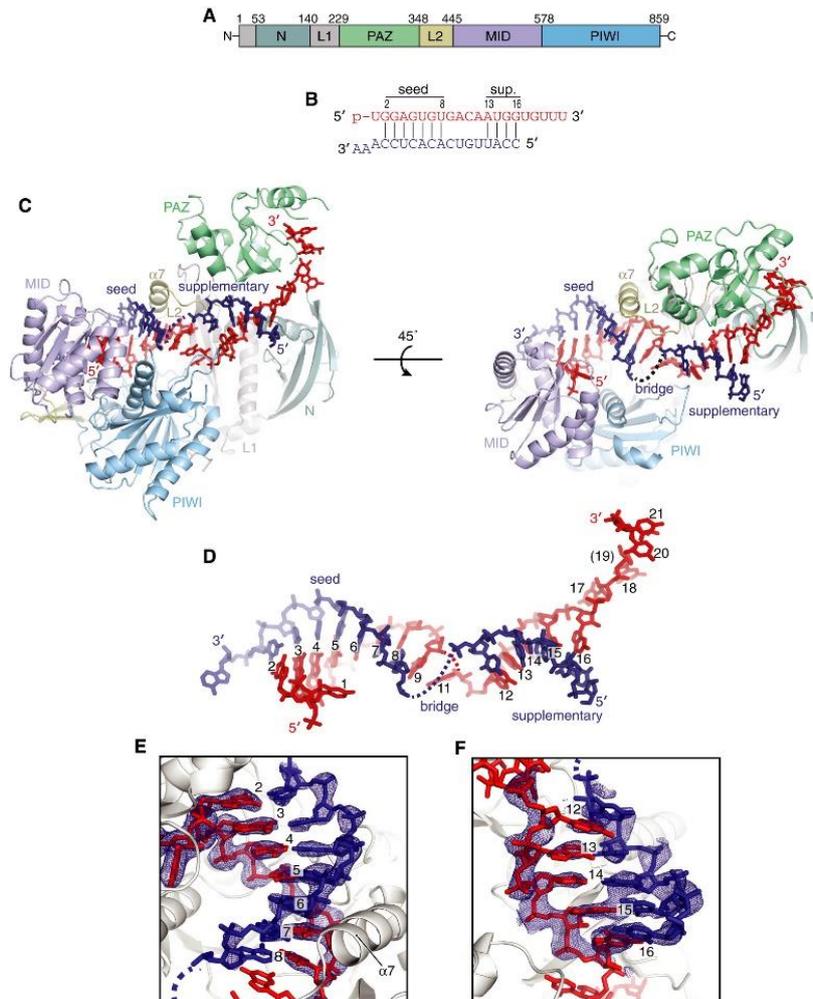


Figura 3. Estructura de la interacción semillas más región suplementaria. **A)** Esquema de la secuencia primaria de AGO2 con dominios coloreados. **B)** Esquema de microARN cristalizado y secuencias diana. Las líneas verticales indican los pares de bases observados en la estructura. microARN de color rojo; ARN diana de color azul oscuro. **C)** Vistas frontal y superior de AGO2 unido a ARN diana y semilla más región suplementaria. **D)** Dúplex microARN-diana con nucleótidos de microARN numerados desde el extremo 5' (AGO2 eliminado para mayor claridad). La línea discontinua indica nucleótidos desordenados en la región del puente. **E)** Primer plano de la región emparejada de semillas del dúplex microARN-diana. Mapa de densidad de electrones Fo-Fc calculado después de omitir nucleótidos alrededor de la región semilla (microARN y diana en las posiciones 1-9), contorneado a 3 sigma, mostrado como una malla azul. **F)** Primer plano de la región de pares suplementarios del dúplex microARN-diana. Mapa de densidad de electrones Fo-Fc calculado después de omitir nucleótidos en la región suplementaria (microARN y diana en las posiciones 12-16), contorneado a 2 sigma, mostrado como una malla azul. Tomada de Sheu-Gruttadauria et al. 2019.

Algoritmos de predicción de blancos de microARNs

Los algoritmos de predicción de sitios de interacción microARN-ARNm son ampliamente utilizados, mayoritariamente para definir posibles ARNm diana a ser luego validados

experimentalmente. A su vez, las predicciones permiten abordar a los ARNm diana en su conjunto y, por tanto, adentrarse en aspectos sistémicos como ontología génica y procesos biológicos vinculados a cada microARN (Mortazavi et al., 2021). Se han desarrollado diferentes herramientas computacionales para predecir los ARNm diana de microARNs. Sin embargo, el número de posibles interacciones microARN-ARNm consideradas por los diferentes métodos varía en gran medida, y refleja las diversas estrategias y prioridades de estos algoritmos de predicción (Agarwal et al., 2015). Para predecir de manera confiable los ARNm diana de un microARN, el primer paso es buscar sitios canónicos de 7 a 8 nt en las 3'UTR. Por sí solo, este simple paso de búsqueda de texto genera predicciones que responden mejor que las de algunos de los algoritmos populares (Bartel, 2018). Una de las razones subyacentes es que algunos algoritmos, en un intento por maximizar la sensibilidad de predicción, consideran interacciones no canónicas que carecen de las interacciones de 7–8 nt en 3'UTR. Consecuentemente, todos estos algoritmos han fallado cuando fueron evaluados experimentalmente (Agarwal et al., 2015).

Los algoritmos de predicción vigentes son: *DIANA-TarBase* (Karagkouni et al., 2017), *miRwalk* (Sticht et al., 2018), *miRDB* (Chen & Wang, 2020), *TargetScan* (Agarwal et al., 2015; McGeary et al., 2019) y *miRTarBase* (Huang et al., 2020) (en orden creciente según última actualización). El algoritmo que ha mostrado el mejor desempeño, e incluso es comparable con los resultados de experimentos de *in vivo crosslinking*, es *TargetScan* (Agarwal et al., 2015) previo a su reciente actualización (McGeary et al., 2019), la cual no ha sido comparada con otros algoritmos hasta el momento. De acuerdo con esta estrategia predictiva, existen catorce características relevantes que afectan la capacidad de un microARN para unirse a su ARNm. Estas se pueden dividir a su vez en tres categorías dependiendo si atañen al microARN, al sitio de interacción propiamente o al ARNm y sus regiones (UTRs o CDS). Si ponemos el foco en el microARN encontramos cuatro características: la cantidad de sitios que tiene ese microARN en todo el transcriptoma, la termodinámica del apareamiento de su región semilla, y con menor capacidad predictiva las identidades de los nucleótidos 1 y 8 del microARN. En cuanto al sitio de interacción, se valora: el contenido de AU cercano, la identidad del nucleótido 8 del sitio, la accesibilidad, la distancia desde el codón stop, el nivel de conservación evolutiva y la

posibilidad de apareamiento de la región suplementaria (sin incluir bucle/*loop* variable, como se detalla en el apartado previo). Por último, si nos enfocamos en el ARNm se valora: el largo de la región codificante, el largo de la 3'UTR, la cantidad de sitios 6mer-*offset* en la 3UTR y la cantidad de sitios 8mer en la región codificante (Agarwal et al., 2015).

Cabe destacar que, cualquiera que sea la herramienta de predicción que se haya utilizado, las interacciones microARN-ARNm siempre deben confirmarse experimentalmente debido a la alta frecuencia de falsos positivos. Las técnicas comúnmente empleadas para la validación experimental incluyen métodos clásicos como qRT-PCR, inmunoprecipitación, *Western blot*, ensayos de reportero, *microarrays*, y métodos de *high-throughput* como RNA-Seq, CLIP-Seq y CLASH. Incluso algunos algoritmos utilizan datos experimentales para su *ranking* de sitios final (Mortazavi et al., 2021). No obstante, debido a que el enfoque experimental también tiene falsos positivos y falsos negativos (Bartel, 2018), el modelo computacional y el enfoque experimental deben utilizarse en conjunto.

Aproximaciones experimentales para el estudio de la interacción microARN-ARNm

Se ha desarrollado una variedad de metodologías para estudiar la interacción de los microARNs con sus genes diana específicos. Las mismas pueden aplicarse a interacciones individuales entre un microARN y un gen diana o pueden ser escaladas a aproximaciones de alto procesamiento involucrando muchos microARNs. Asimismo, algunas aproximaciones se orientan a la confirmación de pares microARNA-ARNm predictivos, mientras que otras permiten su descubrimiento.

La correlación negativa entre el nivel de abundancia de los microARN y sus genes diana son una evidencia fundamental de la represión mediada por estos ARNs. La magnitud del coeficiente de correlación, así como la significación estadística se ha determinado de modo relativamente arbitrario dependiendo del tamaño y heterogeneidad de la muestra (Fort et al., 2018; L. Wang et al., 2018). Los niveles de abundancia pueden ser estudiados

por RT-qPCR de ambas moléculas, así como por análisis de RNA-Seq y small RNA-Seq en las mismas muestras. Con el avance en las tecnologías de secuenciación masiva se han generado grandes bases de datos como es el caso de *The Cancer Genome Atlas* (TCGA). Este repositorio es de los pocos que contiene datos pareados de RNA-Seq y sRNA-Seq. Esto permite estudiar ambas moléculas simultáneamente y hacer inferencias biológicas, incluso se han creado herramientas y bases de datos con el objetivo de evidenciar la interacción microARN-ARNm diana que se valen del cálculo de correlaciones de expresión como: *anamiR* (T. T. Wang et al., 2019), *miRCancerdb* (Ahmed et al., 2018), *Xenabrowser* (Goldman et al., 2020), entre otras.

Otra estrategia es el análisis de los ARNs asociados a AGO2, lo que circunscribe la búsqueda a los pares que están siendo sujetos de represión. Esto requiere de la inmunoprecipitación de AGO2 y el estudio de los ARN en el inmunoprecipitado (generalmente mediante secuenciación masiva). Esta técnica acoplada a secuenciación masiva es ampliamente utilizada para identificar blancos de microARNs al igual que para dilucidar mecanismos de interacción (Hafner et al., 2010; Hamilton et al., 2013; Helwak et al., 2013; Szczyrba et al., 2017). El éxito de esta aproximación depende de la especificidad y estabilidad de la interacción del anticuerpo a ser usado. Se han utilizado diversas metodologías para estabilizar la asociación y lograr mejor resolución. Una de las estrategias más utilizadas para aumentar el rendimiento de la inmunoprecipitación es el *crosslinking*, en la cual se puede incluir o no nucleótidos modificados (PAR-CLIP o CLIP, respectivamente) (Prado et al., 2016). A su vez, existe otra variante de esta técnica que, con el complejo RISC formado, genera mediante ligación de extremos un híbrido microARN-ARNm. En otras palabras, aquellos microARNs que están en conjunto con ARNm en RISC generan un único ARN híbrido (Helwak et al., 2013). Para la inmunoprecipitación de AGO2 humana se ha utilizado un anticuerpo monoclonal que reconoce epítopes de la proteína (Hamilton et al., 2016; Rüdél et al., 2008; Van Eijl et al., 2017) o un anticuerpo anti un epítipo específico como FLAG o Myc (Martinez-Sanchez & Murphy, 2013) cuando se expresa la proteína AGO2 fusionada a estos péptidos. Alternativamente, podría etiquetarse el gen endógeno por edición genómica. Este sistema cobró relevancia cuando dos equipos independientes perfeccionaron el sistema CRISPR/Cas, del inglés, *Clustered Regularly Interspaced Short Palindromic Repeats*)/Cas

(del inglés, *CRISPR-associated loci*), transformándolo en una tecnología eficaz y aplicable en la ingeniería genética (Doudna & Charpentier, 2014; Jinek et al., 2013). Uno de los principales aportes fue el desarrollo de un único ARN guía de la proteína Cas9 denominado *single guide RNA* (sgARN). Esta molécula de ARN quimérica conserva las funciones cruciales del dúplex *crRNA:tracrRNA* encargados naturalmente del direccionamiento del complejo CRISPR/Cas9 a la secuencia de ADN a escindir. El sgARN en su extremo 5' contiene la secuencia de nucleótidos complementaria al sitio blanco y gracias a esto la edición se puede dirigir a prácticamente cualquier *locus* dentro del genoma. De este modo, el sistema CRISPR/Cas9 se ha adoptado como técnica de edición genómica y se ha demostrado la generación eficiente de modificaciones sitio específicas en células humanas y en una amplia variedad de modelos eucariotas (Amoasii et al., 2018; Cong et al., 2013; Jinek et al., 2013). Una vez generada la escisión se desencadenan en la célula los mecanismos de reparación del ADN que serán aprovechados para incorporar una secuencia exógena (Ceccaldi et al., 2015). Para realizar un *knock-in* (*tagging* endógeno) se suele forzar esta maquinaria hacia la “reparación por recombinación homóloga” (HDR) presentándole a la célula un ADN donante con regiones flanqueantes de homología al sitio de corte (Bortesi & Fischer, 2015; Cong et al., 2013). A su vez, ese ADN donante tiene, entre las secuencias de homología, la secuencia exógena que se quiere insertar en el genoma. De esta manera, luego de la reparación por homología en algunas células se incorpora dicha secuencia exógena que en este caso es dirigida hacia el extremo N-terminal de AGO2 y codificará para una cola de histidinas con el fin de ser utilizada posteriormente para inmunoprecipitación. Las células editadas deben pasar a su vez por el proceso de *screening* y selección. Al momento de diseñar el experimento de edición genómica, existen diversas estrategias respecto a la introducción de la maquinaria CRISPR/Cas9 en la célula diana. En lo que refiere a la proteína Cas9 se puede recurrir ya sea a transfectar con plásmidos comerciales, con ARN mensajero para ser traducido *in vivo*, o directamente como proteína producida de forma recombinante (Fajrial et al., 2020).

Por último, también se utilizan aproximaciones para el estudio individual de microARNs como: microARN-biotinilados que luego serán IP mediante estreptoavidina, estudios de expresión génica luego de sobreexpresión/inhibición de microARNs de interés, ensayo de

genes reportero con las regiones reguladoras, con y sin el sitio de interés, o diversas variantes de los sitios (ubicación, secuencia, número de sitios, contexto de secuencias etc.), Kd de complejos tripartitos AGO2-microRNA-mRNA blanco, entre otras, incluyendo últimamente, *ribosome profiling* (Martinez-Sanchez & Murphy, 2013).

Objetivo general

Avanzar en la comprensión de las interacciones específicas de microARN-ARNm

Objetivos específicos

1. Realizar un análisis de las correlaciones de la abundancia de pares microARN-ARNm focalizando en las características suplementarias de las interacciones microARN-ARNm.
2. Poner a punto metodologías experimentales de inmunoprecipitación de AGO2 para la validación futura de hipótesis surgidas de los análisis *in silico* de interacción microARN-ARNm.

Materiales y Métodos del objetivo específico 1

Datos transcriptómicos de microARNs y ARNms de TCGA

Las secuencias de microARNs que fueron utilizadas para la búsqueda de las interacciones al igual que el puntaje de conservación evolutiva (*Family Conservation score*) se obtuvieron de la base de datos *TargetScan 7.2* (Agarwal et al., 2015). Para aumentar la fidelidad del análisis y con el objetivo de evitar microARNs no funcionales, elegimos microARNs con mayor puntaje de conservación evolutiva (puntuación 2 definida por *TargetScan*), referidos de ahora en más como “microARNs conservados” (Tabla Suplementaria 1). A su vez, se seleccionaron los microARNs y ARNm que se detectan en al menos el 80% de las muestras de tejido prostático (sano y tumoral) del proyecto TCGA-PRAD (obtenidas de *FireBrowser*). Este filtro busca evitar ARNm o microARNs de expresión espuria en el tejido cuyas correlaciones no sean confiables ya que posteriormente utilizaremos estas como *proxy* de la actividad de los microARNs (detallado más adelante).

Identificación de interacciones microARN-ARNm en el PRAD-TCGA

Las secuencias de las regiones de los ARNm de todos los genes codificadores para proteínas de *Homo sapiens* se obtuvieron usando [BioMart](#) y el genoma de referencia GRCh38.13 (*Ensembl*). Para la búsqueda de las interacciones microARN-ARNm se diseñó un *script* escrito en el lenguaje de programación *Python* (disponible en [github](#), *Script1.0*) con el cual identificamos todas las secuencias 100% complementarias a las semillas canónicas de los 143 microARNs seleccionados. Se utilizó la nomenclatura de *TargetScan* para nombrar las interacciones canónicas microARN-ARNm, la cual comprende las semillas 6mer, 7mer-A1, 7mer-m8 y 8mer (Tabla suplementaria 1). Si bien se ha definido una semilla denominada *offset-6mer*, la misma apareja con las posiciones 3-8 y no con las posiciones 2-7 de los microARNs maduros, por lo que no satisface la definición de *TargetScan* (Agarwal et al., 2015; McGeary et al., 2019). A su vez, se analizó la existencia de una región suplementaria con complementariedad perfecta en la región 3' de las semillas identificadas en el paso anterior para las interacciones canónicas; en base a la literatura buscamos exclusivamente las posibles interacciones entre el ARNm y las posiciones 13-16 microARN (Sheu-Gruttadauria et al., 2019). Estos dos módulos de

secuencia (semilla y región suplementaria, consulte la Figura 6) se pueden separar mediante un puente de nucleótidos de longitud variable en el microARN y el ARNm. Dado que limitamos la región suplementaria a las posiciones 13-16 del microARN, el largo de este puente en el microARN será de 5 nt y 4 nt para semillas de 6mer/7mer-A1 y de 7mer-m8/8mer respectivamente (Figura 6, panel inferior). Asimismo, basados en los estudios funcionales previos (Sheu-Gruttadauria et al., 2019), nuestra aproximación permite una longitud máxima de 15 nucleótidos de puente semilla-suplementaria en el ARNm. Usando expresiones regulares en el lenguaje de programación *Python*, se identificó los sitios de ARNm para los 8 tipos de interacciones de microARN indicados anteriormente (6mer, 7mer-A1, 7mer-m8, 8mer y 6mer+suppl, 7mer-A1+suppl, 7mer-m8+suppl, 8mer+suppl). Para cada gen, solo se incluyó la variante de *splicing* con el mayor número de interacciones con cada microARN estudiado. Además, dado que los ARNm pueden tener más de un sitio complementario para un microARN dado, para evaluar la contribución represiva de los diferentes tipos de semillas, se seleccionó los ARNm que presentaban solo una interacción microARN-ARNm predicha para cada microARN. Estos análisis se realizaron en forma independiente para las regiones CDS, 3'UTR y 5'UTR, *Script3*).

Análisis de correlación entre la expresión de microARN y ARNm

A continuación, se calculó el coeficiente de correlación de Spearman normalizado (Z-score centrada en el microARN) para la abundancia del ARNm y microARN de todos los pares de microARN-ARNm identificados. Se calculó las correlaciones con la biblioteca *Pandas* de *Python* (McKinney Wes, 2010) (disponible en [github](#), *Script1.1*) utilizando los datos normalizados de TCGA RNA-seq (ARNms) y small-RNAseq (microARNs) (546 muestras de tejido sano y tumoral en conjunto) obtenidas de [FireBrowse](#) (Figura suplementaria 1).

A menos que se especifique lo contrario, las diferencias estadísticas se evaluaron mediante la prueba T de dos colas y los valores p se ajustaron mediante la corrección de Bonferroni, utilizando la biblioteca *SciPy* versión 1.6.2. La prueba exacta de Fisher se realizó utilizando la biblioteca *SciPy* versión 1.6.2. En todos los casos se consideraron significativos los valores de p ajustados <0.05. Todos los gráficos se generaron utilizando la biblioteca *Seaborn* 0.11.2.

Análisis de composición nucleotídica y enriquecimiento de motivos

Los análisis de composición nucleotídica se realizaron mediante *script* hechos en *Python*. Los estudios de enriquecimiento de motivo los realizamos con *MEME suite* (Modo Discriminativo) usando los 39 microARNs con actividad suplementaria frente a los 143 microARN.

Generación de programas ejecutables

Se desarrolló archivos ejecutables para buscar sitios de interacción de microARN/ARNm para sistema operativo *Windows 10* utilizando *PySimpleGUI 4.55.1* (interfaz gráfica de usuario) y *PyInstaller 4.10*. Disponibles en *GitHub*.

Materiales y Métodos del objetivo específico 2

Electroforesis de proteínas y *Western-blot*

Debido al peso molecular de las proteínas gliceraldehído 3-fosfato deshidrogenasa (GAPDH, control 36 kDa) y AGO2 (97 kDa) se preparó geles de poliacrilamida 12% según se indica en la Tabla 1. Para ello, primero se mezcla los componentes del gel agregando en última instancia el persulfato de amonio (APS) 10% y la tetrametiletilendiamina (TEMED). Inmediatamente después, se llenó el casete molde con esta mezcla hasta 3/4 de su altura y se cubre con 1mL de etanol 70% para que la superficie del gel quede pareja. Se esperó 10-20 min y se quitó el etanol. Se procedió a mezclar los componentes del gel de carga (Tabla 2), se agregó en última instancia el APS 10% y el TEMED, y se vertió sobre el gel de corrida. Inmediatamente se insertó el peine evitando la formación de burbujas y se espera 10-20 min para que la acrilamida del gel de carga polimerice. Se retiró cuidadosamente el peine.

Tabla 1 Reactivos para gel de corrida de poliacrilamida

Gel de corrida 12 %		
	1 gel	2 geles
acrilamida:bisacrilamida 30%	2,0 mL	4,0 mL
Tris-HCl 1.5M pH 8.8 SDS 0.4%	1.3 mL	2.5 mL
APS 10%	30 μ L	60 μ L
TEMED	6 μ L	10 μ L

H2O purificada	1.6 mL	3,3 mL
----------------	--------	--------

Tabla 2 Reactivos para gel de carga de poliacrilamida

Gel de carga		
	1 gel	2 geles
acrilamida:bisacrilamida 30:0,8	243.5 μ L	487 μ L
Tris-HCl 0.5M pH 6.8 SDS 0.4%	468.5 μ L	937 μ L
APS 10%	15 μ L	25 μ L
TEMED	3 μ L	5 μ L
H2O purificada	1.14 mL	2.28 mL

Preparación de la muestra

La extracción de proteína se realizó utilizando el *buffer* RIPA (*Thermo Scientific RIPA Lysis and Extraction Buffer, REF: 89900*) siguiendo las indicaciones del fabricante con el agregado de inhibidores de proteasas (*cOmplete™, EDTA-free Protease Inhibitor Cocktail, REF. 11836170001*). Para cada extracción se utilizó la línea celular DU145 cultivadas en placas de 10 cm hasta una confluencia máxima de 80%. La concentración de proteína se cuantificó mediante el *kit* comercial *BCA™ Protein Assay Kit Prod# 23225 (PIERCE)*. Luego se procedió a mezclar la muestra con el volumen necesario de *buffer* de carga de proteínas (4X), se hirvió 5 minutos y se dejó enfriar en hielo antes de cargar. Se utilizó la línea celular DU145.

Condiciones de la corrida electroforética

Se conectó la cuba electroforética a la fuente de poder y se estableció un voltaje constante de 100V (o 20mA) mientras las muestras atraviesan el gel de carga. Luego, se elevó el voltaje a 120V (o 25 mA) cuando las muestras penetren en el gel de corrida, y se dejó migrar hasta que el frente de corrida, indicado por el azul de bromofenol, alcanzó la parte inferior del gel.

Transferencia y *Western blot*

Se comenzó preparando 4 hojas de papel *Whatman* de 3 mm del mismo tamaño del gel y se recortó también una membrana de nitrocelulosa del mismo tamaño. Se procedió a sumergir totalmente los papeles y la membrana en la solución de transferencia y se preparó el sándwich de transferencia en el siguiente orden: la esponja, 2 papeles de filtro *Whatman*, el gel y la membrana. El gel debe quedar sobre el polo negativo (negro). Se eliminó todas las burbujas de aire que puedan haberse formado haciendo rodar una

pipeta o similar sobre el sándwich, se depositó los otros 2 papeles y la esponja, y luego se cerró el casete ajustando con la pestaña blanca. Se colocó el casete dentro de la cuba de transferencia de manera que el gel quede enfrentado al cátodo (polo negativo, color negro) y cubrió todo el sándwich con *buffer* de transferencia. Se comenzó la migración conectando la cuba a la fuente de poder y se realizó la transferencia durante 1 hora a una corriente constante de 400 mA. Al finalizar la transferencia, se tomó la membrana con una pinza e incubó en la solución de bloqueo (leche 5% en PBS) durante 1 hora con agitación constante a temperatura ambiente. Se incubó la membrana con el anticuerpo primario en dilución adecuada en *buffer* de bloqueo o PBS-Tween 0.05% durante 1 hora en agitación constante a temperatura ambiente. Luego, se realizó tres lavados con PBS-Tween 0.1% e incubó la membrana con la dilución adecuada del anticuerpo secundario en *buffer* de bloqueo o PBS-Tween 0.05% durante 1 hora en agitación constante a temperatura ambiente. Por último, se lavó la membrana 2 veces con PBS-Tween 0.1% y luego 1 vez con PBS 1X y se reveló con el escáner de fluorescencia para membranas *LiCor Odyssey Imaging System 9120*, utilizando los canales 700 y 800 nm. Se utilizó el marcador de peso molecular *AccuRuler RGB Prestained Protein Ladder* Cat. No.: 02101-250.

Se utilizaron los anticuerpos primarios comerciales monoclonales: Anti-AGO2 hecho en rata (MABE253 de Millipore, dilución 1:500) y el Anti-GAPDH hecho en conejo (dilución 1:1000, *Cell Signaling Technology #2118*) incubados en leche 5% y PBS-T 0.05% respectivamente durante 12 hs a 4°C. Los anticuerpos secundarios utilizados fueron IGS: anti-rata (dilución 1:10.000, *Cross-Adsorbed Goat anti-Rat, DyLight™ 800, Invitrogen*) y anti-conejo (dilución 1:10.000, *Cross-Adsorbed Donkey anti-Rabbit, DyLight™ 680, Invitrogen*) incubados en PBS-T 0.05% una hora a temperatura ambiente.

Plásmidos

El vector comercial con la secuencia codificante para AGO2, pFRT/FLAG/HA-DEST EIF2C2, se obtuvo del repositorio [Addgene](#) (código #19888). El vector receptor con la secuencia codificante para el péptido Tag, p3XFLAG-CMV-14, fue gentilmente cedido por la Dra. Mercedes Rodríguez Teja (Departamento de Genética, Facultad de Medicina).

Para la expresión de SaCas9 se utilizó el plásmido p6XHis_NLS-SaCas9 obtenido del repositorio *Addgene* (código #101086). Para generar el ADN molde a cortar con la SaCas9 se utilizó como molde de PCR el plásmido pEZX-MR02 (N. de catálogo: CmiR0001-R02, [GeneCopoeia](#)), previamente purificado en el laboratorio.

Amplificación de plásmidos en bacterias

Los plásmidos provenientes de *Addgene* son enviados dentro de bacteria transformadas en un *stab* de medio de cultivo. Por lo tanto, se precedió picando bacterias del *stab* con una punta de micropipeta, y se sumergió el mismo en un tubo de ensayo con 2mL de medio LB con el antibiótico ampicilina (100mg/mL). Se dejó a 37°C, a 200 rpm overnight (ON). Posteriormente volúmenes crecientes (50-200 µL) del pre-cultivo se esparcieron en placas de Petri con el mismo medio de cultivo y antibiótico, se dejaron en estufa de cultivo ON a 37 °C. Se seleccionó la placa en la que se observan colonias aisladas y se repitió el proceso de picar (en este caso una colonia) y generar un pre-cultivo para utilizar en la Miniprep (Ausubel, 2011).

Extracción y purificación de plásmidos

Se realizó la extracción y purificación de los plásmidos con el inserto, a partir de las colonias aisladas pre-cultivadas, mediante un protocolo de Miniprep de Lisis Alcalina (Ausubel, 2011). Se creció las colonias hasta fase estacionaria en 1,5mL de Luria Bertani (LB) con 100 µg/mL de Ampicilina. Las bacterias se centrifugaron a 12000g por 30 segundos y se re-suspendieron agitando con un mezclador de tipo Vortex en 1mL de *buffer* SET, que contiene 20% sacarosa, 50mM Tris HCl pH 4.6 y 6.5mM EDTA. Luego, se centrifugó a 12000g por 1 minuto, se resuspendió el *pellet* en 150 µL de *buffer* SET y se agregó 5 µL de ARNasa A agitando con un mezclador de tipo Vortex. Se agregó 350 µL de Mezcla Lítica, que contiene 1% SDS y 0,2N NaOH, y se dejó a temperatura ambiente por 5 minutos. Pasado el tiempo, se agregó 250 µL de NaAc 3M pH 4.8, se agitó levemente por inversión y se dejó en hielo por 30 minutos. Finalizado el tiempo, se separó el ADN plasmídico de los restos celulares mediante centrifugación a 12000g por 20 minutos a 4°C. El sobrenadante, que contiene el ADN plasmídico en suspensión se retiró y se

precipitó mediante el agregado de un volumen de isopropanol por 15 minutos a temperatura ambiente. Finalmente, se centrifugó a 12000g por 20 minutos a temperatura ambiente, se lavó el *pellet* con EtOH 70% y centrifugación a 12000g por 5 minutos, se secó al aire y se resuspendió en 40 µL de agua estéril. Se cuantificó la cantidad ADN plasmídico extraído utilizando un espectrómetro de gota y se evaluó la pureza por la relación 260/280.

Para amplificar el plásmido ligado con el inserto de interés se realizó Maxiprep utilizando el kit comercial [QIAGEN Plasmid Kits](#) siguiendo las instrucciones del fabricante.

Transformación de bacterias

En esta tesis se utilizó bacterias quimio y electrocompetentes preparadas en el laboratorio previamente.

Químicamente competentes

Se realizó la transformación de bacterias *E. coli* XL-1 químicamente competentes con los productos de ligación mediante el protocolo de *shock* térmico. Brevemente, 100 µL de las bacterias que se encontraban almacenadas a -80°C y se descongeló en hielo por 10 minutos. Luego, se agregó los 6 µL totales de la ligación en condiciones de esterilidad y se dejó 20 minutos en hielo. El *shock* térmico fue de 42°C por 90 segundos y luego en hielo por 2 minutos. Finalmente, se adicionó 4 volúmenes de medio de cultivo LB estéril a temperatura ambiente, se dejó a 37°C durante 1 hora con agitación fuerte y se plaqueó la totalidad de las bacterias en placas de LB Agar con 100 µg/mL de Ampicilina. Las placas se dejaron por 24 horas en estufa a 37°C.

Eléctricamente competentes

Antes de comenzar el procedimiento se enfrió la celda de electroporación en hielo. Por otro lado, se calentó a 37 °C 1 mL de medio SOC. Se comenzó descongelando las bacterias electro competentes en hielo (30-40 min). Se agregó entre 1 y 2 µL de plásmido recién sacado de la ligación (en el caso de estar ligando un inserto, entre 5 pg y 0.5 µg) y se mezcló suavemente. Se transfirieron las bacterias con el plásmido a la celda de electroporación fría y se aplicó el pulso en el electroporador (2.5 kV, 20 µF y el pulso de control a 200 ohms). Inmediatamente se retiró la cubeta del electroporador y se agregó 1 mL de medio SOC a 37°C, sin esterilidad, en el momento. Se pasó con pipeta en

esterilidad desde la cubeta a un tubo *Eppendorff*. Se dejó a 37°C con agitación entre 30-60 min. Finalmente, se centrifugó a 4000 g dejar 100 µL de volumen, se resuspendió el precipitado y se plaqueó en LB.

Digestión y ligación del cDNA del gen *AGO2* plasmídico

El fragmento de ADN correspondiente a ADN copia (cDNA, proveniente de la transcripción reversa del ARNm madura, por lo que no tiene intrones) del gen *AGO2* humano, fue escindida del plásmido comercial pFRT/FLAG/HA-DEST EIF2C2. Para obtener este fragmento de ADN se realizó una digestión doble del plásmido con las enzimas *EcoRI* y *XbaI* generando un producto de 2807 pb. Se utilizó el *buffer* Tango en concentración 2X y el doble de enzima *XbaI* en la incubación (1hora a 37°C). El producto obtenido se purificó mediante columna desde gel de agarosa utilizando [Zymoclean Gel DNA Recovery Kits](#) (Cat. No.: D4007). El producto purificado se utilizó para la ligación en el vector pCMV-14 utilizando las mismas enzimas de restricción. Previo a la ligación el plásmido digerido se desfosforiló para evitar re-circularización utilizando la enzima [FastAP Thermosensitive Alkaline Phosphatase](#) (Cat. No.: EF0651) siguiendo las indicaciones del fabricante. El plásmido desfosforilado y el producto purificado se ligaron utilizando [T4 DNA Ligase](#) (REF: EL0011) siguiendo las instrucciones del fabricante, con una relación 6:1 de inserto/plásmido.

PCR de colonias

Se picaron 10 colonias por placa, se sumergió cada punta en 100 µL de agua estéril y se realizó al mismo tiempo un aislamiento en placas de LB Agar con 100 µg/mL de Ampicilina. Los tubos se hirvieron a 100°C por 5 minutos, se centrifugaron a 16.000g por 15 minutos a temperatura ambiente y se tomó de cada uno 10 µL de molde para la PCR. La reacción se realizó con la enzima *Dream Taq Polimerasa* (*ThermoFisher Scientific*, Cat No.: EP0702) en las condiciones recomendadas y se utilizaron un cebador interno a la secuencia de *AGO2* y el cebador M13 presente en el plásmido sin inserto (Tabla 3). Aquellas colonias que contienen el inserto deberían amplificar un producto de 1088 pb. Se analizó el tamaño de los productos amplificados por electroforesis en gel de agarosa 2% en *buffer* TAE 1X, utilizando bromuro de etidio (BrEt) para su visualización.

Tabla 3 Secuencias de cebadores para PCR colonia

Nombre	Secuencia nucleotídica
M13	CCCAGTCACGACGTTGTAAAACG
AGO2AmpCorto	GGCATATCCTTGGATGGGGG

Secuenciación del plásmido recombinante con AGO2-flag

Los plásmidos purificados se enviaron a secuenciar a la compañía *Macrogen* utilizando 6 *primers* complementarios a la secuencia del plásmido y/o del gen *AGO2* (Tabla 4). Las secuencias se recibieron en forma de electroferogramas y se analizaron mediante alineamiento utilizando el programa *SnapGene Version 6.0.2*.

Tabla 4 Cebadores utilizados para secuenciación

Nombre	Secuencia nucleotídica
M13	CCCAGTCACGACGTTGTAAAACG
Seq 1	CGCAAATGGGCGGTAGGCGTG
Seq 2	CCTGCGTGAGCTTGCAGGCG
Seq 3	AGGTCGGCGCCCGATCGGCA
Seq 4	GCCGGGGATGGGAAGAAGCC
Seq 5	GGGCGAGACCACCAAGCACT

Líneas celulares

La línea celular RWPE1 1 (ATCC CRL-11609) se adquirió del repositorio celular de la compañía ATCC (Manassas, Virginia, EEUU). Es una línea celular humana epitelial proveniente de tejido prostático normal. Estas células presentan una morfología de tipo epitelial y gran adhesión en cultivo. Se cultivaron en el medio de cultivo *Keratinocyte Serum Free Medium (K-SFM)*, Kit Catalog Number 17005-042 el cual debe ser suplementado con 0.05 mg/mL de *bovine pituitary extract (BPE)* y 5 ng/mL de *human recombinant epidermal growth factor (EGF)*.

La línea celular DU145 se adquirió del repositorio celular de la compañía ATCC (Manassas, Virginia, EEUU). Es una línea celular humana proveniente de un carcinoma de próstata,

específicamente de un sitio de metástasis en el cerebro de un hombre adulto caucásico de 69 años. Estas células presentan una morfología de tipo epitelial y son tumorigénicas en ratones inmunodeprimidos. No responden a hormonas y no expresan antígeno prostático (PSA). El análisis citogenético evidencia que sus células son triploides, presentando mayoritariamente 61 y 62 cromosomas (<http://www.iclc.it/listanuova.html>). Se cultivaron en medio de cultivo RPMI 1640 (*Catalog number: 11875093*).

Propagación de los cultivos de líneas celulares

Se observó los cultivos de manera rutinaria bajo un microscopio óptico invertido para evaluar el crecimiento y apariencia de las células. El medio se cambió cada 2 o 3 días en función de la densidad celular y el color del medio, el cual contiene un indicador colorimétrico de pH que pasa de color rosado-rojo (pH 7.4) a un tono naranja-amarillento (pH 6.5). Una vez que las células alcanzaron un 80% de confluencia aproximadamente, se realizó el sub-cultivos para mantener el cultivo en crecimiento. Brevemente, se aspiró el medio utilizando una bomba de vacío y se lavó la monocapa de células con 3mL de PBS 1X con el fin de remover trazas de suero que contienen inhibidores de tripsina. Luego, se adicionó una solución de tripsina 0,25% (w/v) y EDTA 0.53 mM en volumen suficiente para cubrir la monocapa celular (entre 0.5mL – 1mL para frascos de 25cm³) y se incubó por 5 a 15 minutos a 37°C. Durante este lapso, se observó bajo el microscopio y, una vez que la monocapa se desprendió completamente, se les adicionó 1.5 volúmenes de medio para inactivar la tripsina y el EDTA. Se centrifugó a 300g durante 5 minutos, se aspiró el sobrenadante con bomba de vacío y se resuspendieron las células en medio para deshacer los acúmulos celulares. Finalmente, se tomó una alícuota de la suspensión celular que contiene el número de células que deseamos subcultivar y se llevó a 5mL con medio en un frasco de cultivo de 25cm³.

Transfección de líneas celulares

Utilizamos [Lipofectamine 3000](#) (REF: L3000-008) siguiendo las indicaciones del fabricante con la siguiente modificación. Previo a comenzar el protocolo se sometió a las células a una hora con medio mínimo esencial [opti-MEM](#) (REF: 31985-062) sin agregados. Se realizó el protocolo indicado por el fabricante y a las 10 horas se puso medio completo.

Descongelamiento de crioviales

Las células se almacenaron en freezer a -140°C en crioviales de 1.4 mL en el medio correspondiente, suplementado con 5% de DMSO y 10% extra de FBS, aproximadamente un millón de células en fase exponencial de crecimiento. El descongelamiento se realizó en baño de agua a 37°C y, rápidamente, el contenido fue trasvasado a tubos estériles de 15 mL a los que se le añadió 8mL del medio correspondiente precalentado a 37°C . Se centrifugó a 300g durante 5 minutos y luego se removió el medio sobrenadante utilizando una bomba de vacío. Las células fueron resuspendidas en medio y trasvasado a frascos de cultivo celular estériles de 25 cm³. Los frascos se almacenaron horizontalmente en una incubadora a 37°C , a 1 atmósfera de presión, 5% de CO₂ y 95% de humedad.

Cuantificación de ARNm de AGO2

Se extrajo ARN total de cultivos exponenciales de la línea celular RWPE-1 sin transfectar y las 3 clonas transfectadas aisladas generadas en este trabajo. Se produjo cDNA utilizando oligodT y el kit SuperScript III (Invitrogen, REF: 18080-044). Con un par de cebadores listados en la Tabla 5 se realizó una PCR cuantitativa con el sistema *SensiFASTTM SYBR^R Hi-ROX Kit* (BIO-92005, *meridian bioscience*), en el equipo *Step-One Plus* de *Applied Biosystems*. Para la normalización de los datos por cantidad de ARN total se utilizó el gen TBP. Las condiciones del ciclado fueron las recomendadas por el protocolo de *SYBER Green*. Los resultados obtenidos se analizaron por el método de $2^{-\Delta\Delta\text{Ct}}$ (Livak & Schmittgen, 2001). Los cebadores utilizados fueron diseñados en el laboratorio con el objetivo de que el amplicón se encuentre en el entorno de las 100 pb.

Tabla 5 Cebadores para qPCR

Nombre	Secuencia nucleotídica
AGO2_F	TACAAGGCACAGCCAGTAATC
AGO2_R	CGTTATCTCCACCTTTAGACCTTT
TBP_F	GATCAAACCCAGAATTGTTCTCC
TBP_R	ATGTGGTCTTCCTGAATCCCTTT

Expresión de SaCas9 recombinante en *E. coli*

Se utilizó la cepa *Rosetta™ 2(DE3) pLysS (Novagen)* de *E. coli* transformada con el plásmido inducible comercial [p6XHis NLS-SaCas9](#). Inicialmente se realizó un pre-cultivo bacteriano de la cepa *Rosetta 2(DE3)* transformada en medio 2xYT (Ampicilina 100 mg/mL). Se dejó a 37°C, a 200 rpm overnight (ON). Al día siguiente, en dos matraces de 2L se preparó el cultivo, con 500mL de medio 2xYT (Ampicilina 100 mg/mL) y 10mL del pre-cultivo. Los matraces se dejaron en un agitador a 200 rpm y 37°C hasta que el cultivo alcanzó una densidad óptica (DO) entre 0.4 y 0.6 unidades de absorbancia (UA). Se añadió 500µL del inductor isopropil-β-D-1-tiogalactopiranosido (IPTG) 1M al medio de cultivo y se mantuvo en el incubador con agitación durante 24 horas a 18°C. Se traspasó el cultivo a tubos Falcon de 50 mL, y se los centrifugó por 20 minutos a 10000 rpm y 4°C. Posteriormente se lisaron las células químicamente; para ello se resuspendió el *pellet* en *buffer* fosfato en un volumen 1/20 respecto al cultivo y se prosiguió según el protocolo de lisis química explicado a continuación.

Protocolo empleado para lisis química de las bacterias transformantes luego de la inducción:

Primero, se resuspendió el *pellet* de 100 mL de cultivo en 5 mL de *buffer* fosfato obteniendo una concentración final de 1X. Luego, se agregó 50 µL de PMSF 100 mM para una concentración final de 1mM. Se agregó 100 µL de Lisozima 50mg/mL para una concentración final de 1mg/mL. Se agregaron perlas de vidrio y se agitó la muestra durante 30 minutos a temperatura ambiente. A continuación, se agregó 4 µL de ADNasa 10mg/mL y ARNasa 10mg/mL quedando en una concentración final de 8 µg/mL. Se agregó 5 µL de MgCl₂ 1M para una concentración final de 1mM. Para finalizar, se agitó la muestra durante 30 minutos a temperatura ambiente y se quitaron las perlas de vidrio.

Finalmente se centrifugó el lisado a 10000 g por 3 minutos a 4°C, se guardó el *pellet* (fracción insoluble) a 20°C y el sobrenadante a 4° (fracción soluble) hasta la etapa de purificación.

Purificación de SaCas9 recombinante

La purificación de la SaCas9 se llevó a cabo a través de cromatografía de afinidad por columna de níquel ([His-Trap HP](#), GE™, REF: GE29-0510-21) debido a que nuestra proteína recombinante presenta una etiqueta de seis residuos de histidina en su extremo N-terminal. Se utilizó un equipo de FPLC (del inglés, *Fast Protein Liquid Chromatography*) de Cytiva y modelo [ÄKTA Start](#), configurando la colección de fracciones cada 4 mL de fase móvil. Previo a la etapa de purificación, se preparó la columna de níquel *His-Trap* mediante el procedimiento conocido como “*stripping and recharging*”. Para su preparación, en primer lugar, se pasaron diez volúmenes de agua por la columna utilizando una bomba peristáltica y una jeringa [Microliter Hamilton™](#). Posteriormente, se pasó por la columna diez volúmenes de *buffer* de unión (*binding buffer*) con 50 mM de EDTA. En segundo lugar, se realizó la limpieza de la columna pasando diez volúmenes de hidróxido de sodio (NaOH) 1M, diez volúmenes del *buffer* de unión y diez volúmenes de agua, en orden cronológico. El tercer y último paso, es la recarga de la columna, para ello se pasó dos volúmenes de sulfato de níquel (NiSO₄) y diez volúmenes de agua. En este punto la columna está lista para ser utilizada, y se almacenó en etanol 20% a 4°C. Previo al uso se retiró el etanol y se equilibró la columna en *buffer* de unión. Todas las soluciones incorporadas a la columna se filtraron con filtros de 0,45 µm y se trabajó a un máximo de 1 mL/min. Una vez que la columna estaba lista, se cargó con el producto de lisis correspondiente a la fracción soluble, descrito anteriormente. Luego se realizaron lavados con *buffer* de unión conteniendo 30 mM de imidazol. En este lavado, se colectó aquella fracción de proteínas con baja afinidad a la columna (FT, de *flow-through*). Posteriormente se realizaron tres lavados, el primer lavado se realizó en 53 mM de imidazol (5/95% *buffer* elución/unión), el segundo en 82 mM (11/89% *buffer* elución/unión) y el tercero con 124 mM de imidazol (20/80% *buffer* elución/unión). Para el lavado final o elución de la proteína SaCas9, se utilizó 300 mM de imidazol correspondiéndose con el 100% del *buffer* de elución. Posteriormente se utilizó una columna de gel filtración [PD10 desalting column, Sephadex™](#) (MERCK, Cat. No: GE17-0851-01). Se utilizaron centrífugas *Macrosep Advance Centrifugal Devices with Omega Membrane 30K* (Prod. ID: MAP030C38) para llevar a la proteína a una concentración que permitiera su utilización en las posteriores aplicaciones. El producto obtenido se

conservó en 40% de glicerol y se almacenó a -20°C. Por último, se hizo un gel SDS-PAGE con los distintos lavados, incluyendo la elución final, con la finalidad de evaluar la pureza de la proteína purificada.

Diseño de sgARNs

Se utilizó el software informático [CRISPOR](#) (Concordet & Haeussler, 2018) para el diseño de sgARNs capaces de dirigir a SaCas9 a un corte doble hebra (DSB, por sus siglas en inglés) en regiones de interés en los genes *mCherry* (control) y *AGO2*. Una vez elegido el sitio de interés, el software asiste en el diseño de dos oligonucleótidos parcialmente complementarios que mediante un proceso de extensión darán lugar al molde de ADN que será utilizado para la síntesis del sgARN por transcripción *in vitro* (IVT) (Tabla 6 y Tabla 7). Para la secuencia del gen *AGO2* humano se utilizó la base de datos NCBI y para el caso del gen control *mCherry* se utilizó la secuencia contenida en el plásmido pEZX-MR02 (N. de catálogo: CmiR0001-R02, [GeneCopoeia](#)). Se seleccionó los oligonucleótidos más adecuados para ejercer el corte de interés en *mCherry* y *AGO2*; y se enviaron a sintetizar a la empresa [MACROGEN](#).

Tabla 6 Reactivos y condiciones utilizados para generar el ADN molde de las IVT

Reactivos	Volumen	Concentración Final
5X Phusion HF Buffer	20 μ L	1X
dNTP Mix (10mM c/u)	2 μ L	200 μ M c/u
Oligonucleótido guía (10 μ M)	5 μ L	0,5 μ M
Oligonucleótido universal (10 μ M)	5 μ L	0,5 μ M
Phusion DNA Polymerase (2 U/ μ L)	1 μ L	0,02 U/ μ L
H ₂ O milliQ csp 100 μ l	67 μ L	-

Tabla 7 Programa de termociclado utilizado para generar el ADN molde para las IVT

Etapas	Temperatura	Tiempo	N° Ciclos
Desnaturalización inicial	98°C	30 segundos	1
Desnaturalización	98°C	10 segundos	35
Annealing	60°C	30 segundos	35
Extensión	72°C	15 segundos	35

Extensión Final	72°C	10 minutos	1
-----------------	------	------------	---

Transcripción *in vitro* del sgARNs

La síntesis de los sgARNs se llevó a cabo mediante ensayos de IVT, utilizando el ADN molde generado y purificado en el apartado anterior. Para esto, se utilizó el *HiScribe™ T7 High Yield RNA synthesis Kit* (Cat. No.: E2040S, *New England Biolabs, NEB*). Se preparó la reacción de IVT según la Tabla 8 y se incubaron las reacciones a 37°C durante 16 horas. Luego, los sgARNs obtenidos se purificaron utilizando el *Monarch® RNA Cleanup Kit* (Cat. No.: T2040S, *New England Biolabs, NEB*) y se cuantificó su concentración y pureza midiendo la absorbancia a una longitud de onda de 260 y 260/280nm respectivamente en un *NanoDrop LITE* (*Thermo Scientific*). Los sgARNs se almacenaron a -80°C.

Tabla 8 Reactivos y condiciones utilizados en las IVT

Reactivos	Volumen	Concentración final
10X <i>buffer</i> de reacción	2 µL	1X
ATP (100mM)	2 µL	10mM
UTP (100mM)	2 µL	10mM
CTP (100mM)	2 µL	10mM
GTP (100mM)	2 µL	10mM
ADN molde	-	1 µg
T7 ARN polimerasa <i>mix</i>	2 µL	-
H ₂ O RNAsa <i>free csp</i> para 20 µL	-	-

Síntesis de ADN para ensayos de digestión *in vitro*

mCherry

El ADN blanco para el ensayo de actividad de la SaCas9 del gen *mCherry* se obtuvo mediante PCR a partir del plásmido pEZX-MR02. Los cebadores utilizados (Tabla 9) se diseñaron para amplificar un producto de 360pb. Para estas reacciones de PCR se utilizó la enzima *DreamTaq DNA Polymerase* (N° de catálogo: EP0705, *ThermoFisher Scientific*),

siguiendo las instrucciones detalladas en la Tabla 10 y Tabla 11. El producto se purificó mediante el método de extracción acuosa con fenol-cloroformo seguido de una precipitación con etanol. La concentración de ADN obtenida (Abs. 260 nm) y a pureza (Abs. 260/280nm) fueron medidas en *Nanodrop LITE (Thermo Scientific)*. Los fragmentos amplificados se analizaron mediante una electroforesis en gel de agarosa 1%.

Tabla 9 Oligonucleótidos cebadores utilizados para amplificar el gen mCherry.

Nombre	Secuencia Nucleotídica
mCherry_F	CACGAGTTCGAGATCGAGGG
mCherry_R	AGCCCATGGTCTTCTTCTGC

Tabla 10 Reactivos y condiciones utilizados en la PCR del gen mCherry

REACTIVO	VOLUMEN	CONCENTRACIÓN FINAL
10X DreamTaq <i>Buffer</i>	5 µL	1X
dNTP Mix (10mM each)	1 µL	0,2 mM each
<i>Forward</i> cebador (10 µM)	1 µL	0,2 µM
<i>Reverse</i> cebador (10 µM)	1 µL	0,2 µM
ADN plásmido (30 ng)	5 µL	-
DreamTaq DNA <i>Polymerase</i> (5U/ µL)	0,25 µL	0,025U/ µL
H ₂ O milliQ csp 50 µL	36,75 µL	-

Tabla 11 Programa de termociclado utilizado para amplificar el gen mCherry

Etapas	Temperatura	Tiempo	Nro. de ciclos
Desnaturalización inicial	95°C	3 minutos	1
Desnaturalización	95°C	30 segundos	35
<i>Annealing</i>	60°C	30 segundos	35
Extensión	72°C	1 minuto	35
Extensión final	72°C	10 minutos	1

AGO2

Se llevó a cabo la reacción de PCR para amplificar la región 5' de *AGO2* a partir de ADN genómico humano y utilizarla como diana de corte de Cas9 *in vitro*, guiada por los sgARNs diseñados para la edición endógena. Los cebadores utilizados para la reacción se detallan en la Tabla 12. Se realizó la PCR utilizando la enzima [KAPA HiFi HotStart DNA Polymerase \(Roche\)](#) debido al alto contenido de GC de la región a amplificar. En la Tabla 13 y Tabla 14 se detalla el procedimiento y el programa de termociclado utilizado.

Tabla 12 Cebadores utilizados en las tres reacciones de PCR del gen *AGO2*

Nombre	Secuencia nucleotídica
AGO2_F	GAGACCGCCTTCTACTCTGC
AGO2_R	GTCCGCTGATCTGATCCCTG

Tabla 13 Condiciones utilizadas en las PCRs del gen *AGO2* para las distintas enzimas empleadas

Reactivos	Volumen	Concentración final
5X KAPA HiFi GC Buffer	5 µL	1X
dNTP Mix (10mM each)	1 µL	0,3 mM each
Forward cebador (10 µM)	0,75 µL	0,3 µM
Reverse cebador (10 µM)	0,75 µL	0,3 µM
ADN genómico (13ng)	5 µL	-
KAPA HiFi HotStart DNA Pol. (1 U/µL)	0,5 µL	0.5 U/ µL
H ₂ O milliQ csp para 25 µL	12 µL	-

Tabla 14 Programa de termociclado utilizado para amplificar el gen *AGO2*

Etapas	Temperatura	Tiempo	Nº de ciclos
Desnaturalización inicial	95°C	5 minutos	1
Desnaturalización	98°C	20 segundos	35
Annealing	60°C	30 segundos	35
Extensión	72°C	1 minuto y 30 segundos	35

Extensión final	72°C	10 minutos	1
-----------------	------	------------	---

Purificación de productos de PCR

Se comenzó añadiendo 1 volumen de fenol-cloroformo al tubo de PCR y homogeneizó hasta formar una emulsión. Se centrifugó a 12000 g durante 90 segundos a temperatura ambiente. Luego, se transfirió la fase acuosa (arriba) a un nuevo tubo, luego se descartó la fase orgánica (abajo) y la interfase. Se repitieron estos pasos dos veces para obtener un mayor rendimiento. Se continuó añadiendo un volumen de cloroformo y se repitió los pasos de centrifugado y transferencia de fase acuosa. Para la precipitación, se agregó un volumen de acetato de sodio a una concentración final de 0.3 M y se homogeneizó la muestra mediante un mezclador de tipo Vortex. Se procedió a precipitar la muestra con dos volúmenes de etanol 95%, se homogeneizó y se almacenó en el *freezer* durante dos horas para la formación del precipitado. Se centrifugó el ADN a máxima velocidad durante 20 minutos a 0°C y se removió el líquido sobrenadante. Para finalizar, se realizó un lavado con etanol 70% y se centrifugó a 12000g durante dos minutos, se repitió este paso una vez. Se descartó el líquido sobrenadante y se diluyó precipitado en 20 µL de agua milli-Q.

Electroforesis en gel de agarosa

En este trabajo se prepararon geles de 1 y 2%. En el caso de ADN, se mezcló la agarosa con buffer TAE 1X. Se procedió a calentar hasta generar una solución homogénea. Posteriormente se agregó BrEt, se vertió el gel en la cuba, se colocó el peine para la formación de los pocillos y luego de salificado se llenó la cuba con TAE 1X. Se procedió a cargar las muestras las cuales previamente fueron preparadas con *buffer* de carga para ADN. La corrida electroforética se realizó a 100V. Al finalizar la corrida, se reveló el gel en un transiluminador ultravioleta. Para la visualización de los sgARNs, se siguió el mismo procedimiento con algunas modificaciones: se preparó el gel añadiendo 1 g de agarosa en 47.5 mL de buffer MOPS (1X), se disolvió la agarosa calentando, y se agregaron 2.5 mL de formaldehído y BrEt. Las muestras fueron preparadas con *buffer* de carga para ARN 1X, y se calentaron por 10 minutos a 65°C.

Evaluación de la actividad de la SaCas9 *in vitro*

Se evaluó la capacidad de la proteína SaCas9 de ejercer un DSB específico *in vitro* en los genes *mCherry* y *AGO2*. Para esto, se realizó una incubación de los componentes generados para cada sistema durante 1 hora a 37°C, según Burle-Caldas et al., 2018. Los resultados de las digestiones se analizaron en geles de agarosa. En la Tabla 15 se detalla el procedimiento empleado.

Tabla 15 Ensayo de actividad SaCas9

Reactivos	Concentración inicial
SaCas9	6 µg
sgARN	6 µg
ADN blanco	2 µg
<i>Buffer SaCas9</i>	1 µL
H ₂ O <i>RNAse free</i> csp 20 µL	-

Luego de la incubación, las reacciones de digestión se trataron con Proteinasa K (800 U/mL, NEB) y ARNasa A (20mg/mL, NEB). Primero, se agregó 1 µL de Proteinasa K para 20 µL de reacción de digestión y se incubó a 37°C durante 30 minutos. Se realizó la inactivación de la proteínasa K mediante incubación a 95 °C grados durante 10 minutos. Se dejó enfriar las muestras en hielo durante 10 minutos y se agregó 2 µL de ARNasa y se incubó durante 8 minutos a 50 °C.

Resultados y discusión

Identificación de sitios de interacción microARN-ARNm en el transcriptoma PRAD-TCGA basado en la complementariedad de bases y las correlaciones de abundancia (OE 1)

La idea central detrás del diseño del flujo de trabajo es evaluar la actividad y los mecanismos de interacción de los microARNs sobre sus ARNm blanco. Para lograr estos objetivos en este trabajo analizamos los datos de secuenciación de ARNm y de pequeños ARN de tejido prostático humano no perturbados obtenidos del proyecto TCGA-PRAD. Dicha evaluación se basó en el cálculo de correlaciones de Spearman entre ambos sets de datos, donde se esperan *a priori* correlaciones negativas debido a la actividad represora de los microARNs (Bartel, 2018).

Los datos de interacciones microARN-ARNm fueron obtenidos mediante la implementación de expresiones regulares generadas en el lenguaje de programación *Python*. Nuestra estrategia busca específicamente la interacción mediada por la semilla y la región 3'suplementaria y permite un puente entre ambas de entre 1 y 15 nucleótidos de largo. Optamos por este abordaje porque ninguno de los algoritmos disponibles considera la interacción de la forma que fue descrita mediante la cristalografía del complejo proteico con AGO2 (Sheu-Gruttadauria et al., 2019).

Dada la especificidad tisular de la expresión de microARNs, decidimos realizar este estudio dentro de un solo tipo de tejido. Elegimos la cohorte PRAD porque nuestro laboratorio estudia ARNs vinculados a la biología del cáncer de próstata (PC) y, por tanto, posee un marco teórico y herramientas experimentales que permiten desafiar las hipótesis generadas en los estudios *in silico* con aproximaciones experimentales. Por otro lado, PRAD-TCGA tiene un gran número de muestras (546 pacientes) y ha sido muy estudiada en el contexto de los microARNs (Abeshouse et al., 2015; Lin et al., 2021; Wei et al., 2020; Yang et al., 2019; Ye et al., 2018). Finalmente, el hecho del que el PC sea una

neoplasia muy heterogénea desde el punto de vista molecular, y que, por ende, no existan subtipos mayoritarios discriminados, permite aplicar análisis globales poco sesgados por subtipos moleculares. Es así como investigamos la expresión de los microARNs conservados (221 microARNs) y todos los transcritos codificadores de proteínas humanos anotados en Ensemble (20440), estudiados en las 546 muestras de la cohorte PRAD-TCGA (20531) (resumen del flujo de trabajo en la Figura 4).

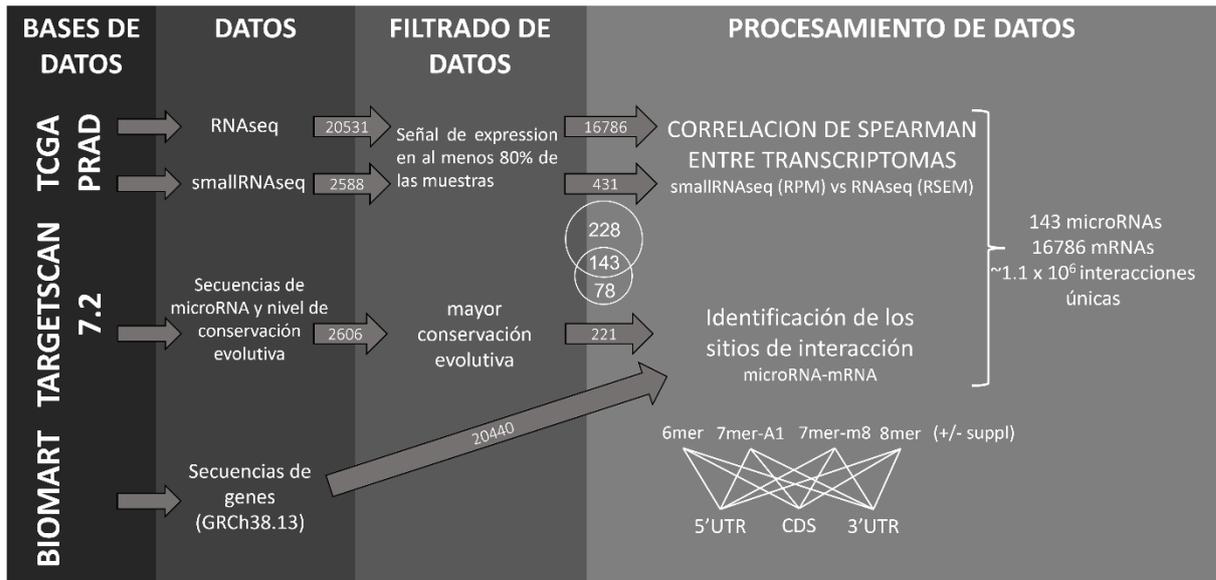


Figura 4. Flujo de trabajo y características de los datos analizados

Para aumentar la relación señal/ruido, filtramos los transcritos de baja expresión, definidos arbitrariamente como aquellos no detectados (0 conteos) en $\geq 20\%$ de las muestras, lo que redujo la lista de microARN y ARNm a 143 y 16786 respectivamente. Proporcionalmente, este filtrado produce un enriquecimiento en transcritos de mayor expresión, que afecta en mayor proporción a los microARNs, tal vez porque tengan más heterogeneidad de expresión entre los tejidos individuales y menos profundidad de secuenciación por muestra (Figura 5, Tabla Suplementaria 1). La distribución bimodal de la expresión de ARNm y su paso a una unimodal luego de remover los transcritos de baja expresión ha sido ya descripta (de Torrenté et al., 2020; Nagaraj et al., 2011).

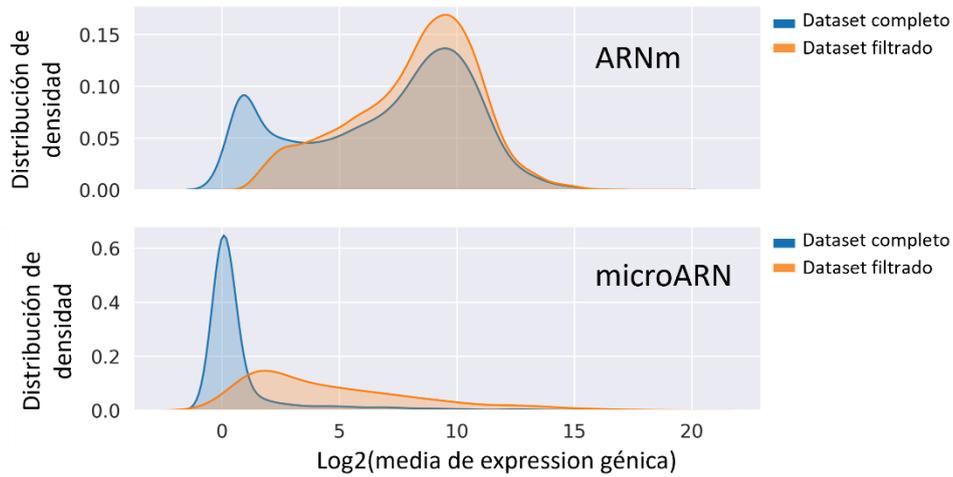


Figura 5. Distribución de densidad. Distribución de densidad *kernel* estimada, representa los datos usando una curva de densidad de probabilidad continua (análogo a histograma) para la media de expresión en el conjunto de datos de microARN y ARNm provenientes del TCGA-PRAD antes (*dataset* completo) y después (*dataset* filtrado) del filtrado por detección en el 80 % de las muestras.

Con el objetivo de comparar el poder represivo de los diferentes tipos de interacciones microARN-ARNms, estudiamos las semillas 6mer, 7mer-A1, 7mer-m8 y 8mer, con o sin la región suplementaria (en lo sucesivo denominados 6mer+suppl, 7mer-A1+suppl, 7mer-m8+suppl y 8mer+suppl, o en términos generales como “región suplementaria”) (Figura 6).

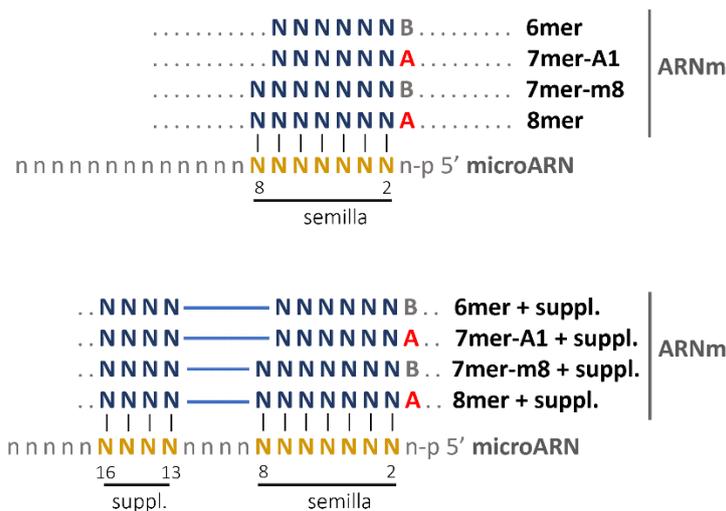


Figura 6. Descripción gráfica de las interacciones microARN-ARNm analizadas en este trabajo. Tipos de interacción analizados en este trabajo. En amarillo se muestran los nucleótidos del microARN que interaccionan con los nucleótidos del ARNm (azules) mediante interacciones de Watson-Crick (barras verticales). Dependiendo del sitio, el nucleótido 8 de la semilla puede o no

estar interaccionando. La letra A en las secuencias corresponde a Adenina y la letra B a cualquier nucleótido menos Adenina (IUPAC).

A continuación, buscamos todos los sitios de ARNm que tienen una complementariedad perfecta con los ocho motivos semilla estudiados (4 semillas con o sin interacción suplementaria), considerando que representan sitios posibles de unión de microARNs. Este análisis se realizó de forma independiente para cada una de las tres regiones de los ARNm (5'UTR, CDS y 3'UTR). Identificamos 3065731 sitios de interacción microARN-ARNm putativos (285237 en 5'UTR, 1226568 en CDS, 1553926 en 3'UTR) entre los 143 microARNs y los 16786 ARNm. Cuando eliminamos las interacciones que involucran más de un sitio en el ARNm (análisis independiente por microARN y por región del ARNm) obtuvimos un total de 1116787 sitios (193203 en 5'UTR, 443323 en CDS, 480261 en 3'UTR), correspondientes a un promedio de interacciones posibles. Mediante este abordaje generamos un conjunto de datos que nos permitirá estudiar las interacciones de manera general sin utilizar microARNs o ARNms específicos. Esta es una de las diferencias de este trabajo con respecto a los trabajos publicados que utilizan una o pocas interacciones para formular hipótesis generales sobre los mecanismos de interacción microARN-ARNm (Baek et al., 2008; Grimson et al., 2007; Hafner et al., 2010; Helwak et al., 2013; Selbach et al., 2008; Sheu-Gruttadauria et al., 2019).

Validación de mecanismos represivos de los microARNs utilizando la actividad de inferida a partir de las correlaciones de expresión de microARN-ARNm en TCGA-PRAD.

Con el objetivo de validar si nuestra aproximación es capaz de detectar la represión mediada por microARNs, evaluamos si permite identificar las reglas de represión de microARN actualmente conocidas. Para esto utilizamos las correlaciones de Spearman transformadas como Z-scores, una normalización que describe la relación de cada correlación con la media de correlaciones del grupo permitiendo la comparación directa y la acumulación de datos entre los distintos microARNs.

Los resultados de nuestro análisis muestran que los niveles de represión aumentan con la longitud de la semilla, es decir, el número de bases involucradas en la interacción (6mer < 7mer-m8 < 7mer-A1 < 8mer) (Figura 7A) (La Tabla 16 muestra diferencias estadísticamente significativas p-valor<0.05), lo que confirma una regla importante de la interacción microARNs-ARNm (Friedman et al., 2009; Grimson et al., 2007; Hafner et al., 2010). Estas diferencias son estadísticamente más significativas para la comparación entre los motivos de la 3'UTR (p-valor promedio = E-25) seguida de la comparación intra-CDS (p-valor promedio = E-16). En el caso de 5'UTR solo existen diferencias significativas entre 6mer/7mer-A1 y 7mer-A1/7mer-m8 (p-valor 1.14 E-02 y 2.93 E-03 respectivamente). Como era de esperar, aunque las diferencias están parcialmente influenciadas por el número de interacciones en cada conjunto de datos, está claro que la represión del ARNm diana, deducida de las correlaciones, aumenta con la longitud del apareamiento, Figura 7A) (Grimson et al., 2007; Nielsen et al., 2007). Otra regla fundamental de la represión mediada por microARNs se confirma en la observación de que la región del transcripto más reprimida es la 3'UTR, seguida de la CDS y, por último, casi sin evidencia de represión, la 5'UTR. Sin embargo, en contraste con los resultados anteriores de la literatura, nuestros resultados indican que la interacción semilla 7mer-A1 resultó ser significativamente más represiva que la 7mer-m8 en las tres regiones del ARNm estudiadas y también en comparación a la semilla de 6mer (mientras que el 7mer-m8 no muestra diferencia con la semilla 6mer) (Friedman et al., 2009; Grimson et al., 2007).

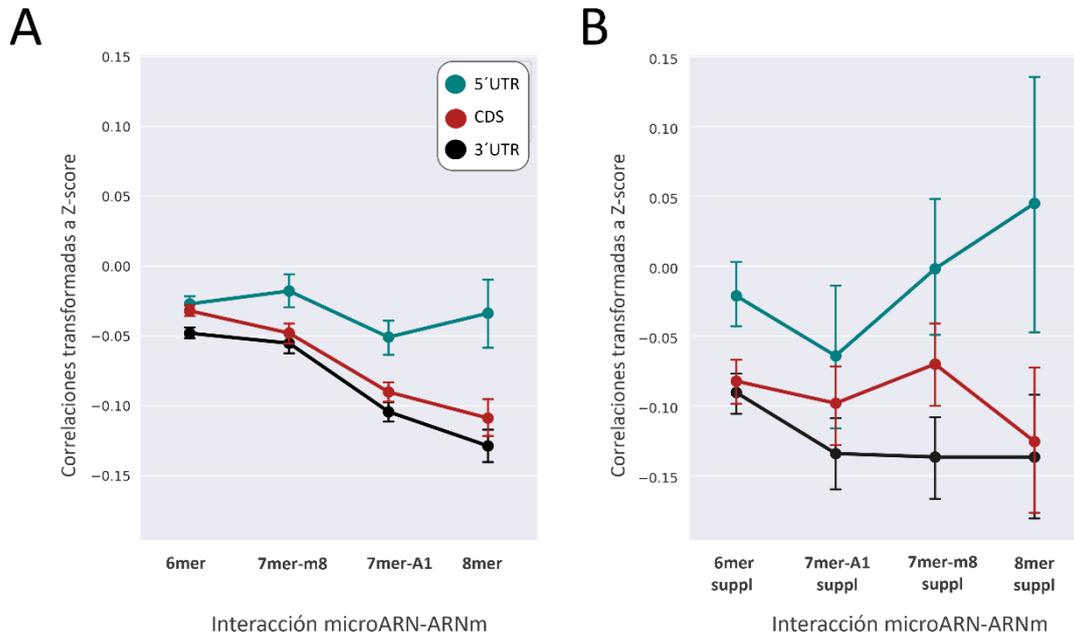


Figura 7. Represión ejercida por las interacciones microARN-ARNm con diferentes sitios de apareamiento. Promedio de las correlaciones entre los 143 normalizadas como Z-score microARN y sus ARNm dianas. Se representan para cada una de las regiones (CDS, 5'UTR y 3'UTR) en curvas separadas. Las barras verticales representan el intervalo de confianza (95%).

La Figura 7B muestra el resultado del mismo análisis realizado para los sitios de interacción que involucran las semillas más la región suplementaria. Nuevamente, las comparaciones muestran una tendencia de mayor represión para semillas más largas, siendo más notoria en las regiones CDS y 3'UTR. Sin embargo, los valores estadísticos resultado de la comparación entre éstas están por encima de p-valor 0.05 (considerado no significativo), lo que puede atribuirse en parte a la menor cantidad de datos de interacciones con semillas suplementarias (Figura 8).

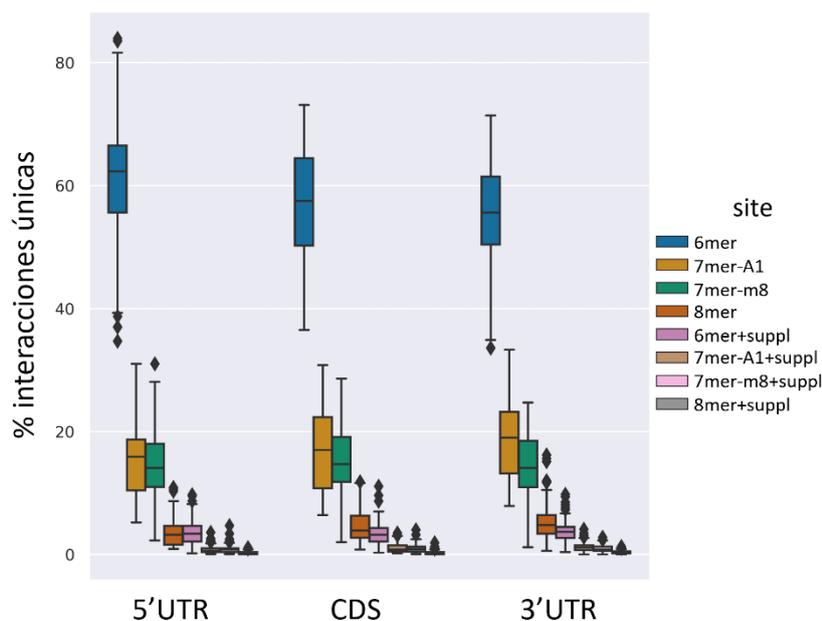


Figura 8. Porcentaje de interacciones únicas microARN-ARNm identificadas en PRAD-TCGA, discriminadas por tipo de interacción y por las regiones del ARNm. Datos calculados para los 143 microARN del análisis.

La comparación de semillas únicas y semillas más regiones suplementarias para todos los sitios muestra que las interacciones suplementarias contribuyen a la represión, particularmente en las regiones 3'UTR y CDS (Figura 7A-B) aunque solo se encuentran diferencias estadísticamente significativas en 6mer (CDS y 3'UTR) y 7mer-m8 (3'UTR) (FDR < 0.05) (Tabla 16). Este hallazgo respalda la contribución a la represión de la región suplementaria que ha sido propuesto utilizando aproximaciones experimentales *in vitro* (Grimson et al., 2007; Helwak et al., 2013; Moore et al., 2015; Sheu-Gruttadauria et al., 2019; Xiao & Macrae, 2020). A su vez, las regiones del ARNm siguen el orden de represión esperado cuando comparamos dentro de cada sitio, siendo la región con mejor respuesta la 3'UTR seguida de la CDS y por último la 5'UTR (Figura 7A-B), en concordancia con los valores de p de comparaciones entre los sitios en las tres diferentes regiones (Tabla 17). De este modo, los resultados del análisis de las interacciones de regiones del ARNm con semillas más región suplementaria confirma reglas establecidas para la actividad de los microARNs y valida nuestro abordaje. En otras palabras, este hallazgo corrobora que, en tejido no perturbado, la represión de microARN opera principalmente en 3'UTR y CDS (Grimson et al., 2007; Hafner et al., 2010; Helwak et al., 2013) y la represión se beneficia con el incremento de la región de complementariedad que involucra no solo a la semilla sino también a la región suplementaria.

Tabla 16 Resultados estadísticos de la comparación entre sitios en las distintas regiones del ARNm

Group1	Group2	p-value		
		5'UTR	CDS	3'UTR
6mer	7mer-m8	n.s.	6.09E-03	n.s.
6mer	7mer-A1	1.14E-02	2.36E-42	4.82E-45
6mer	8mer	n.s.	1.74E-25	9.23E-35
7mer-m8	8mer	n.s.	2.24E-13	4.41E-23
7mer-A1	7mer-m8	2.93E-03	3.94E-14	9.48E-21
7mer-A1	8mer	n.s.	n.s.	1.24E-02
6mer	6mer suppl	n.s.	2.84E-08	9.90E-07
6mer	7mer-m8 suppl	n.s.	n.s.	3.66E-08
6mer	7mer-A1 suppl	n.s.	1.71E-04	2.48E-09
6mer	8mer suppl	n.s.	1.23E-02	3.72E-03
7mer-m8	7mer-m8 suppl	n.s.	n.s.	1.74E-06
7mer-m8	8mer suppl	n.s.	n.s.	1.52E-02
6mer suppl	7mer-m8	n.s.	2.84E-03	6.92E-04
6mer suppl	8mer	n.s.	n.s.	1.74E-03
7mer-A1 suppl	7mer-m8	n.s.	2.20E-02	2.43E-07

Tabla 17 Resultados estadísticos de la comparación entre las regiones para cada interacción

Group1	Group2	p-value			
		6mer	7mer-A1	7mer-m8	8mer
3UTR	5UTR	5.98E-09	2.69E-14	4.34E-07	7.30E-12
3UTR	CDS	2.62E-08	1.37E-02	n.s.	n.s.
5UTR	CDS	n.s.	6.25E-08	6.82E-05	2.06E-07
Group1	Group2	p-value			
		6mer suppl	7mer-A1 suppl	7mer-m8 suppl	8mer suppl
3UTR	5UTR	2.81E-06	3.62E-02	1.14E-05	1.29E-03
3UTR	CDS	n.s.	n.s.	5.11E-03	n.s.
5UTR	CDS	7.30E-05	n.s.	n.s.	4.49E-03

Características de la interacción microARN-ARNm que involucra la región suplementaria

Basándonos en nuestros resultados anteriores, principalmente en lo referido a la contribución de la interacción de la región suplementaria a las semillas, nos propusimos investigar si este abordaje permite profundizar en el estudio de las interacciones suplementarias para así conjeturar nuevas reglas de la interacción microARN-ARNm. Para los siguientes análisis nos enfocamos en las interacciones de las semillas más la región suplementaria de 6mer+suppl (3'UTR y CDS) y 7mer+suppl. (3'UTR), ya que mostraron un aumento estadísticamente significativo en la represión con relación a las semillas correspondientes sin interacción suplementaria (Tabla 16). Dentro de estas interacciones, seleccionamos los microARNs con comportamiento individual igual al global. Es decir, en este caso, utilizamos el mismo método de estudio de las interacciones que en el análisis principal, pero para cada microARN con su propio conjunto de datos. Para quedar dentro del subconjunto con el que continuaremos trabajando, estos microARNs deben de cumplir una de las siguientes condiciones: ejercer una represión mayor significativa cuando interaccionan utilizando regiones suplementarias (6mer+suppl (3'UTR y CDS) o 7mer+suppl. (3'UTR)). En otras palabras, replicar el fenómeno observado en el análisis global. Realizamos esta selección para enriquecer el subconjunto de datos con microARNs de región suplementaria funcional (de ahora en más referidos como “microARNs represivos”) ya que el objetivo es profundizar en las características de esta interacción que, *a priori*, no sabemos cuán extendida está en la represión mediada por microARNs. En total identificamos 39 microARN represivos en uno o más de los tres conjuntos de datos estudiados y 18 microARN con el comportamiento opuesto, es decir, donde la interacción suplementaria disminuiría la actividad represiva. Si cada conjunto de sitios comparados se analiza por separado, el número de microARN a favor o en contra de una actividad represiva son: 19 frente a 11 (6mer 3'UTR, prueba binomial p-valor = 0.2), 16 frente a 4 (7mer-m8 3'UTR, prueba binomial p-valor = 0.01) y 21 frente a 8 (6mer CDS, prueba binomial p-valor = 0.02). Los 39 microARN represivos no tienen una diferencia significativa en los niveles de expresión o el número de ARNm diana con el total de 143 microARN analizados (p valor ≤ 0.05 , datos no mostrados). Esto sugiere

que la selección de estos 39 microARNs no está sesgada por nivel de expresión o capacidad de interacción con ARNm diana.

Utilizando estudios cristalográficos se demostró recientemente que la semilla y la región suplementaria pueden estar separadas mediante un bucle no estructurado de entre 1 y 15 nucleótidos dentro de la proteína AGO2 humana (Sheu-Gruttadauria et al., 2019). Incluso se observó que aún si existe complementariedad entre las bases centrales del microARN (8-12) y el ARNm, el apareamiento de bases es evitado. Este resultado es consistente con los resultados de la técnica *Photoactivatable ribonucleoside enhanced cross-linking and immunoprecipitation* (PAR-CLIP), donde también se demostró que los nucleótidos 8-13 del microARN interaccionan preferentemente con AGO2 en vez de con el ARNm (Hafner et al., 2010). Ambos estudios demuestran que el apareamiento de microARN y ARNm en el contexto de RISC difiere con respecto a las reglas de interacción ARN-ARN en solución, y que existen ciertas restricciones impuestas por la proteína AGO2. En el trabajo cristalográfico de AGO2 humana (Sheu-Gruttadauria et al., 2019), también mostraron que la interacción 8mer+suppl con puentes de hasta 10 nt de largo es más represiva que el sitio 8mer. Los resultados también sugieren que las interacciones que involucran regiones suplementarias con alto contenido en GC pueden establecer bucles/puentes de hasta 15 nucleótidos. En nuestro trabajo evaluamos la contribución de este bucle a la represión. La descripción del largo de este puente, tanto en el microARN como en el ARNm, se realiza con la variable *off-set* o compensación, que consiste en la diferencia de largo (nt) entre el puente del ARNm y el del microARN. Entonces, dado que el puente entre la semilla y la región suplementaria del microARN es de 5 nt para el 6mer+suppl (8-12 nt) y 4 nt para el 7mer-m8+suppl (9-12 nt), el mismo largo en el puente del ARNm determina que no haya formación de bucles en ninguna de las dos moléculas (cero *off-set*) (Figura 9A). En otras palabras, un *offset* de 0 ocurre cuando ambas regiones puente tienen el mismo largo y, por lo tanto, no van a generar bucles en ninguna de las dos moléculas de ARN, entendiendo por bucle/puente los nucleótidos no apareados con la otra molécula de ARN en cuestión. Si el largo del puente del ARNm es mayor que el del puente del microARN se formará un bucle en el ARNm (valores de *off-set* positivos), mientras que en el caso contrario se formará en el en el microARN (valores de *off-set* negativos Figura 9A). Nos propusimos ahondar en la contribución del puente a la

represión utilizando el *off-set* como referencia. En el análisis comparamos a modo de control los 143 microARNs del análisis versus los 39 microARNs represores. Como se puede observar, en el control no hay cambios significativos en la represión con respecto al *off-set* (Figura 9B, paneles izquierdos). Por otro lado, los 39 microARNs represivos muestran una represión heterogénea en función del *off-set* (Figura 9B, paneles de la derecha). En los tres conjuntos de datos, se observa un aumento gradual en la represión entre las medidas cuando los *off-set* se acercan al cero *off-set*, lo que indica que la longitud del bucle entre la semilla y la región suplementaria es relevante para la represión y que 0-offset es un óptimo. Este patrón se hace más evidente en el conjunto integrado de los datos (Figura 10). A su vez, este aumento de represión consecutivo alcanza su máximo en +1, hallazgo consistente con resultados experimentales previos (McGeary et al., 2022; Sheu-Gruttadauria et al., 2019). En el resto de *off-sets* positivos (bucle en el ARNm) no se observa una tendencia clara a nivel individual ni en el conjunto integrado de los datos. Los patrones observados son similares para los tres conjuntos de datos, aunque las diferencias son menos evidentes en los conjuntos de 6mer CDS y 7mer-m8 3'UTR, probablemente debido al menor número de interacciones que los componen.

Las interacciones suplementarias de los 39 microARN represivos muestran una represión heterogénea según el *off-set* (Figura 9B, paneles de la derecha). Para las interacciones 6mer que ocurren en el 3'UTR, se observa una ganancia gradual de represión con bucles en el microARN más cortos (*off-set* negativos), alcanzando su máxima represión en el *off-set* 0 (no bucle). Por el contrario, la longitud del bucle en el ARNm (*off-set* +1 a +10) no es directamente proporcional a la represión de la interacción 6mer en la 3'UTR, aunque el *off-set* +1 es una de las longitudes más represivas. Se presenta un patrón similar de represión para los otros dos conjuntos de datos (6mer CDS y 7mer-m8 3'UTR), si bien las diferencias son menos evidentes, probablemente debido a la menor cantidad de interacciones que los componen. Para acumular la información proveniente de los tres conjuntos de datos, ajustamos sus *Z-score* usando la normalización de mínimo-máximo (para normalizar sus diferentes rangos de *Z-score*), y luego combinamos los tres valores. Este análisis integrado indica que la longitud del bucle de microARN es inversamente proporcional a la represión, mientras que un *off-set* +1 proporciona la máxima represión (Figura 10). Consistentemente, la represión óptima +1 se ha mostrado en trabajos previos

(McGeary et al., 2022; Sheu-Gruttadauria et al., 2019). Sin embargo, nuestros resultados sugieren que otros *off-set* pueden producir una represión similar.

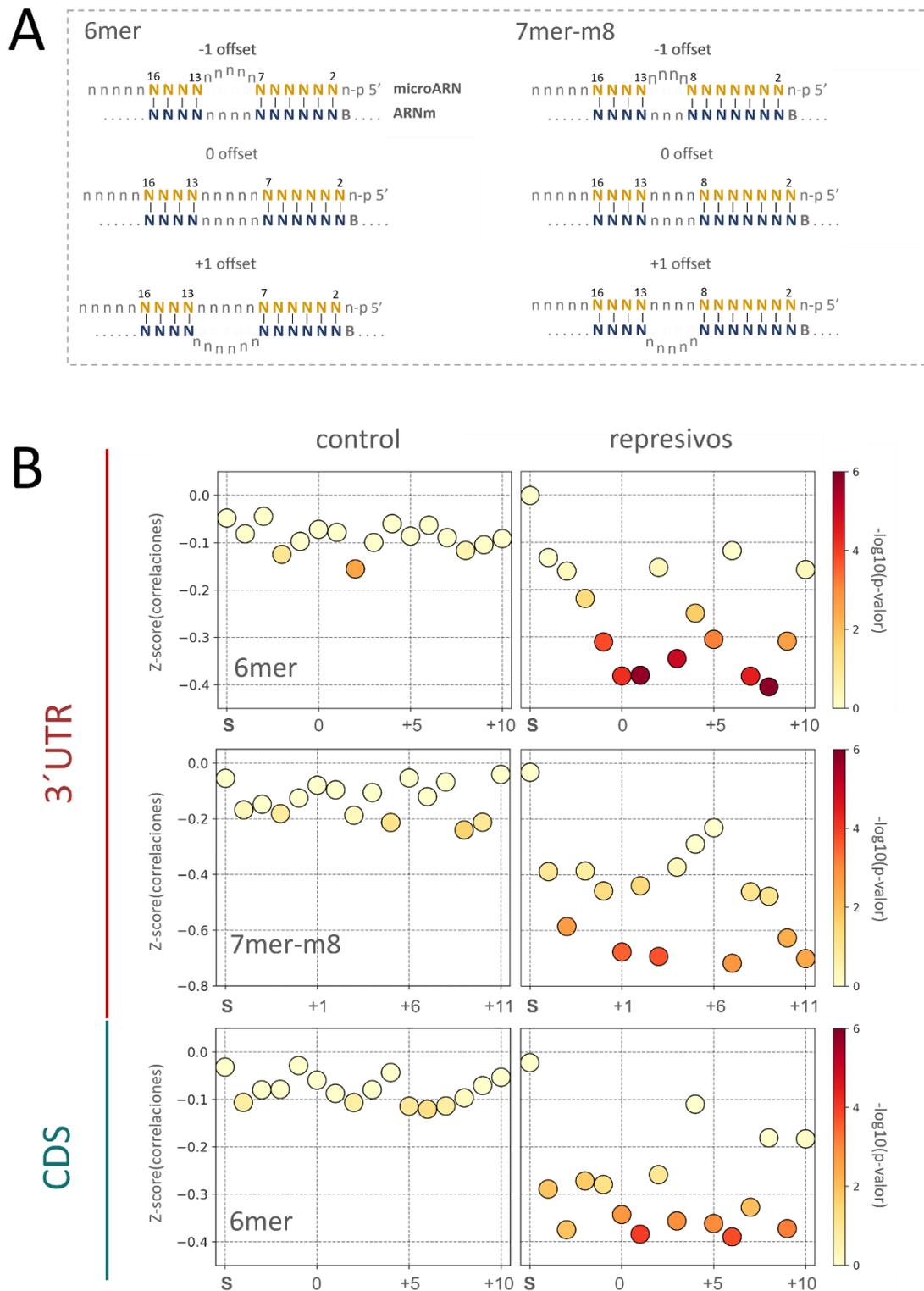


Figura 9. Contribución de la longitud del puente semilla-región suplementaria a la represión de *microARN* A) Representación gráfica del concepto de compensación u *off-set*. La región denominada “puente” se define como los nucleótidos desde la semilla hasta la región

suplementaria, en este caso considerando una semilla de microARN de la posición 2-7 (6mer) o 2-8 (7mer-m8) y la región suplementaria de la posición 13-16, dejando una región de 5 nt (6mer) o 4 nt (7mer-m8) entre ellos. Si el ARNm blanco tiene el mismo valor en el puente implica que no hay formación de bucles en ninguna de las dos moléculas lo que denominamos *off-set* cero. Valores más altos implican la formación de bucles en el ARNm (*off-sets* positivos) y los valores más bajos implican la formación de bucles en el microARN (*off-sets* negativos). **B)** Análisis de la contribución del *off-set* a el fenómeno represivo mediado por microARNs sobre el ARNm usando los 39 microARN con interacciones suplementarias represivas y los 143 microARN totales del estudio como control. Se indican las regiones del gen y el tipo de semillas analizadas. La comparación estadística (Test de T con corrección Bonferroni) fue realizada dentro de cada recuadro comparando las interacciones solo de semilla (S) con cada uno de los *off-sets* calculados.

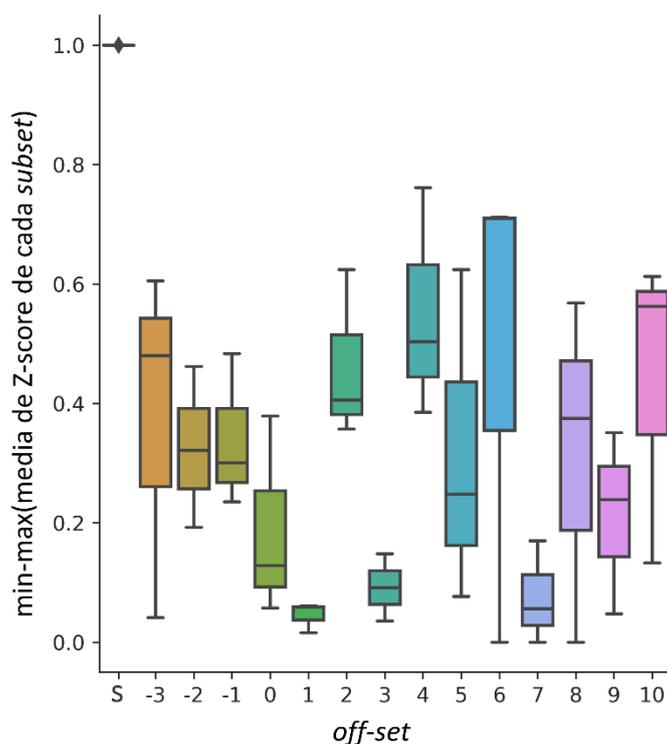


Figura 10. Niveles de represión de los diferentes off-sets del puente semilla-región suplementaria en la interacción microARN-ARNm. Datos combinados de los tres tipos de interacción estudiados. Se normalizó dentro de cada conjunto de datos por mínimo-máximo para evidenciar el perfil. Cada valor de *off-set* cuenta con 3 datos; la media de *Z-score* (min-máx) de cada conjunto de datos.

Basados nuevamente en los descubrimientos de la cristalografía de AGO2 donde muestran que el contenido de GC de la región suplementaria afecta la capacidad represiva (Sheu-Gruttadauria et al., 2019), comparamos el contenido de GC de los 39 microARNs represivos seleccionados *versus* el total de microARNs de nuestro análisis.

Observamos un claro desvío del contenido de GC entre las posiciones 13-16 de los microARNs; la comparación muestra consistentemente un contenido de GC significativamente menor en los microARNs represivos para los tres sitios analizados (Figura 11A-B). Cabe destacar a su vez, que las posiciones 4, 5, 6 y 19 de los sitios 6mer tienen un contenido de GC desviado del total. Asimismo, el análisis de la composición de bases muestra una preferencia por la base A en las posiciones 13-15 y por la base T en la posición 16 correspondientes a la región suplementaria del microARN (Figura 11C-D), preferencia que no se observa en el conjunto completo. También existe una tendencia global por una A en la posición 2 y 5. El logotipo representado en la Figura 11D resume lo antes mencionado y es resultado de la comparación de la secuencia completa de los 39 microARNs represivos con los 143 microARNs del conjunto completo utilizando el “Modo Discriminativo del Descubrimiento del Motivo” del *MEME Suite* y está respaldado por 16 microARNs (valor $E = 4.7e-15$) (Bailey et al., 2015). Este logo contribuye a la idea de bajo contenido de GC en la región suplementaria. Aunque el enriquecimiento de la base T y A en la posición 1 y 2 de todos los microARN observados en la Figura 11C se ha descrito previamente en mamíferos y otras especies (Friedman et al., 2009; B. Wang, 2013), el logotipo enriquecido sugiere una sobrerrepresentación de esta base en las posiciones 1 y 5.

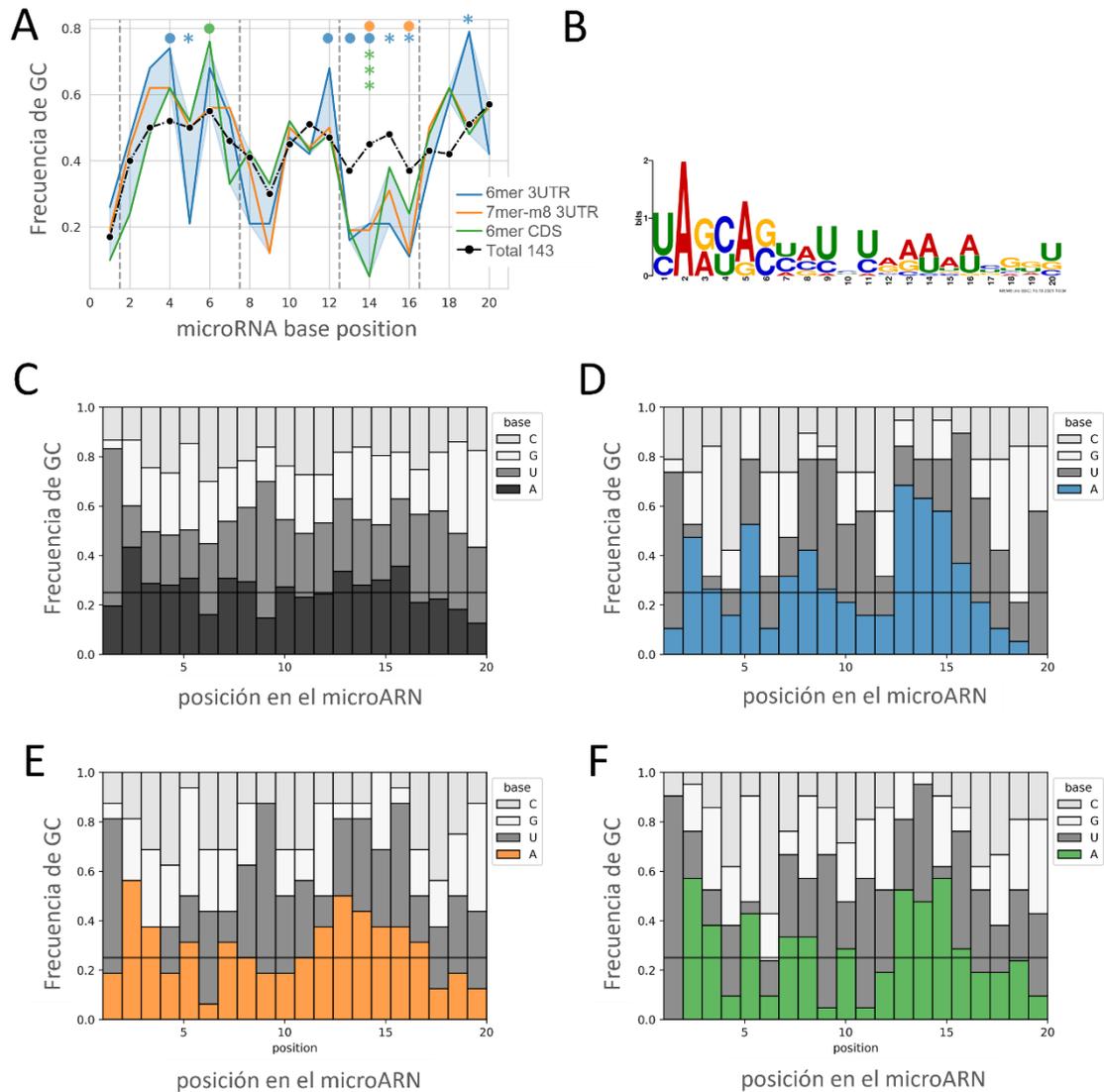


Figura 11 Composición de nucleótidos de los microARN con interacciones suplementarias represivas. **A)** Distribución del contenido de GC a lo largo de las posiciones de los microARN con las tres interacciones represivas indicadas (en color) frente al conjunto completo de microARN analizados (en negro). Se realizó la prueba de Fisher para comparar cada uno de los sitios representados con los 143 microARN totales ($\bullet < 0,1$, $* < 0,05$, $*** < 0,001$). **B)** Motivo de secuencia de microARN identificado por *MEME suite* (Modo Discriminativo) usando los 39 frente a los 143 microARN. El motivo está representado por 16 de los 39 microARN (valor $E = 4.7e-15$). **C-F)** Composición de bases en cada posición del microARN. Los tipos de sitio están representados por colores como se indica en A (C: negro/todos los microARN, D: azul/6mer-3'UTR, E: naranja/7mer-m8 3'UTR y F: verde/6mer CDS). La línea horizontal representa la frecuencia de $\frac{1}{4}$ esperada para una distribución aleatoria de nucleótidos.

Diseño de Programa ejecutable para búsqueda de interacciones microARN-ARNm

Finalmente, con el objetivo de hacer accesible la búsqueda de sitios de interacción microARN-ARNm que involucren la región suplementaria, desarrollamos dos *scripts* en el lenguaje de programación *Python* y generamos los archivos ejecutables en *Windows* (.exe) disponibles en [github](#). En la búsqueda de extender la herramienta a usuarios sin conocimiento de programación desarrollamos para cada *script* una interfaz gráfica (Figura 12). Para esto se utilizó la librería [PySimpleGUI](#). Los archivos ejecutables los generamos utilizando [PyInstaller](#). *PyInstaller* analiza módulos y librerías que la secuencia de comandos necesita para ejecutarse. Luego recopila copias de todos esos archivos, incluido el intérprete activo de *Python* y los coloca con su secuencia de comandos en una sola carpeta u, opcionalmente, en un solo archivo ejecutable como lo hicimos en este trabajo. Seleccionamos el sistema operativo *Windows* por su extensa utilización.

Según la pregunta del usuario se deberá de utilizar uno de los dos programas ejecutables. Si el usuario desea buscar todos los microARNs que regulan a un ARNm particular debe usar el ejecutable [searchByMessengerRNA.exe](#), mientras que si el usuario desea buscar todos los ARNm que regula un microARN debe utilizar [searchByMicroARN.exe](#). Ambos programas reportarán no solo las interacciones semilla canónicas sino también aquellas interacciones que involucren a la semilla más la región suplementaria. La ventaja de este algoritmo es que busca específicamente las regiones de interacción mapeadas experimentalmente para el complejo con AGO2, en lugar de utilizar algoritmos de hibridación ARN-ARN que no tienen en cuenta este aspecto biológicamente fundamental (Agarwal et al., 2015). Según el programa elegido, se desplegará una u otra interfaz gráfica. En el caso de *searchByMicroARN.exe*, generamos una ventana en la cual se escribe o se selecciona de la lista el nombre del microARN de interés. El usuario también debe elegir de qué región del ARNm desea obtener los sitios de interacción. Por lo tanto, la ventana cuenta con una segunda sección donde se selecciona la región de interés (3'UTR, CDS o 5'UTR) (Figura 12A). En el caso de *searchByMessengerRNA* optamos por una interfaz más sencilla donde se selecciona en una misma ventana toda la información necesaria, el nombre del gen, su transcripto, tipo de gen y región de interés (Figura 12B). La salida de ambos programas es un archivo (.csv) el cual incluye toda la información

pertinente al sitio de interacción como: nombre del gen, ID del transcripto, tipo de gen, región del ARNm donde se encontró el sitio, interacción (semilla o suplementaria), nombre del microARN, tipo de interacción (tipo de semilla con o sin suplementaria), dónde comienza el sitio y dónde termina, secuencia nucleotídica del sitio, secuencia nucleotídica del puente, largo del puente. El nombre del archivo de salida va a tener a su vez la información seleccionada por el usuario dependiendo del programa que se haya seleccionado.

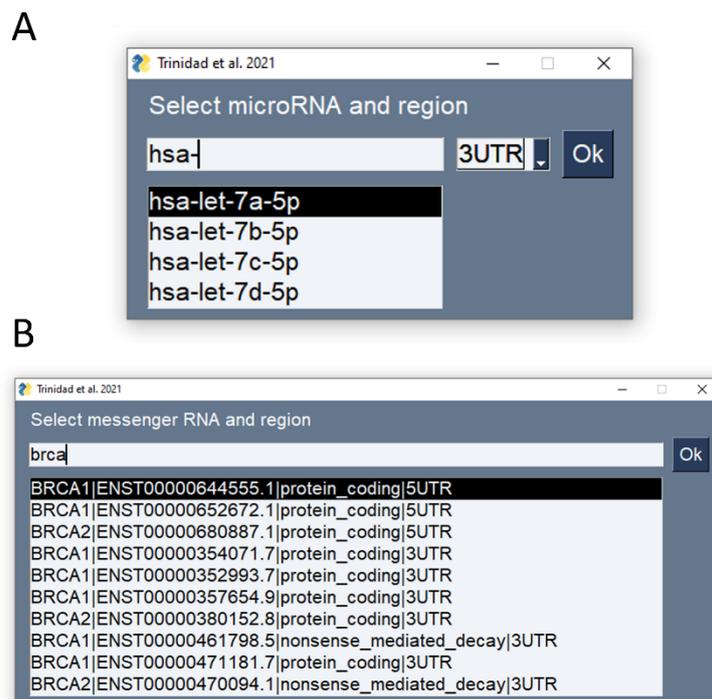


Figura 12. Interfaces gráficas de los archivos ejecutables para la búsqueda de interacciones microARN-ARNm.

TargetScan, que es el recurso más robusto y comprensivo hasta ahora desarrollado para predecir sitios efectivos de interacción microARN-ARNm, pondera la contribución de la interacción microARN-ARNm en la región suplementaria. Sin embargo, la asignación de puntaje la hace dentro de un marco teórico que incluye restricciones en el largo del puente y para cada interacción proporciona un único score combinado para la semilla y la región suplementaria que se agrega con los 13 criterios adicionales que utiliza para la predicción (Agarwal et al., 2015; McGeary et al., 2019). Nuestro *pipeline* proporciona una predicción menos compleja, pero también menos sesgada por hipótesis previas ya que

solo se basa en la complementariedad de bases y la distancia y longitud del apareamiento en la región semilla. Por otro lado, despliega las interacciones microARN-ARNm proporcionando así información libre de prejuicios que queda accesible para el uso. Por último, nuestro abordaje permite evaluar la región suplementaria incluyendo variaciones en el largo del puente. A diferencia de muchos algoritmos este programa puede ser usado para cualquier especie, lo que amplía el espectro de posibilidades de aplicación.

Avances experimentales en metodologías de inmunoprecipitación de AGO2 para la validación futura de hipótesis surgidas de los análisis *in silico* de interacción microARN-ARNm (OE 2)

En este trabajo de tesis, nos propusimos implementar técnicas de biología molecular para la validación futura de las hipótesis surgidas tanto del análisis de datos transcriptómicos como mediante la aproximación bioinformática del objetivo anterior. El objetivo es obtener un sistema para inmunoprecipitar el complejo RISC, que permita circunscribir los estudios de interacción microARN-ARNm a este complejo. Esta aproximación constituye un paso inicial fundamental para experimentos donde se evalúen interacciones de microARNs y ARNms específicas (endógenos o exógenos) por diversos ensayos subsiguientes (qRT-PCR, RNA-Seq, etc.). A su vez, puede estar también precedida por aproximaciones de entrecruzamiento, PAR u otras, o asociadas a ensayos funcionales (genes reporteros). Utilizamos tres estrategias para inmunoprecipitar AGO2: un anticuerpo anti-AGO2 humana, un etiquetado endógeno de AGO2 utilizando el sistema CRISPR/Cas9 y un etiquetado exógeno a través de generar y clonar un plásmido con AGO2 etiquetada para transfectar a células.

Como primera estrategia planteamos la utilización de un anticuerpo monoclonal anti-AGO2. Este abordaje tiene como ventaja evitar la sobreexpresión de la proteína y ahorrar el agregado de un epítipo, disminuyendo la probabilidad de resultados artefactuales causados por niveles no fisiológicos o cambios en la actividad de la proteína etiquetada respectivamente. Sin embargo, tiene la desventaja de que su eficacia depende de la

calidad del anticuerpo en cuanto a su especificidad y afinidad. Los epítomos que se fusionan a las proteínas son muy útiles porque confieren un epítome inexistente en la célula y se han desarrollado anticuerpos especialmente seleccionados para conferir alta especificidad. Por eso, la fusión de epítomos inmunogénicos es una estrategia ampliamente utilizada y es por eso por lo que quisimos implementarla. Con este objetivo, probamos editar la proteína AGO2 endógena, así como generar una proteína de fusión para expresar forzosamente en la célula. La edición genómica de AGO2 confiere la ventaja de mantener el contexto genómico, y así los controles de regulación de la expresión génica generando menos perturbación que la incorporación de plásmidos. Sin embargo, la edición endógena y la selección de las células editadas implica un riesgo alto de fracaso y un tiempo considerable de generación. Es importante mencionar que, hasta el momento, no se ha publicado ningún etiquetado endógeno de la proteína AGO2, por lo tanto, es pertinente la generación de un protocolo experimental funcional a distintos modelos biológicos. Como una alternativa a las dificultades del etiquetado endógeno, nos planteamos la expresión de una proteína etiquetada exógena, para lo que se construyó un plásmido con AGO2 fusionada a una etiqueta para transfectar. En contrapartida, las proteínas plasmídicas suelen sobre expresarse a niveles considerablemente mayores conduciendo a interacciones no fisiológicas, o a toxicidad celular. La intención final de esta búsqueda de estrategias es lograr un protocolo que perturbe lo menos posible el sistema biológico para disminuir resultados artefactuales que puedan producirse.

Anticuerpo Anti AGO2

En primer lugar, se ensayó anti-AGO2 con el objetivo de utilizarlo posteriormente en ensayos de inmunoprecipitación ([clon 11A9](#), hecho en Rata) (Rüdel et al., 2008). Dicho anticuerpo ha sido utilizado en la mayoría de los trabajos de inmunoprecipitación de AGO2, seguidos de secuenciación masiva para la identificación de ARNm diana de microARN (Hamilton et al., 2016; Kishore et al., 2011). Sin embargo, más recientemente se ha visto que interacciona con el complejo SWI/SNF (Van Eijl et al., 2017), lo que agrega un paso de complejidad a la hora de interpretar los resultados. Estudiamos la especificidad del anticuerpo por AGO2 en nuestras condiciones (lisis, condiciones de bloqueo, etc.) mediante la técnica de *Western blot*. Para esto utilizamos la línea celular de cáncer de próstata Du145 y como control de carga utilizamos un anticuerpo anti-GAPDH. En lo que respecta al protocolo de *Western blot* tuvimos que probar distintas condiciones (optimización de tiempo transferencia, tipo -leche o BSA-, concentración y tiempo de bloqueo, incubación del anticuerpo con y sin bloqueo y su concentración, así como el revelado con varios anticuerpos secundarios. Como se puede apreciar en la Figura 13 conseguimos las condiciones necesarias para evaluar mediante *Western blot* la especificidad del anticuerpo anti-AGO2. Comprobamos que las condiciones más rigurosas para incubar este anticuerpo primario disponibles en la bibliografía son las que evitan en mayor medida el reconocimiento de bandas adicionales a AGO2 en el gel. Sin embargo, de cualquier manera, utilizando estas condiciones se obtuvieron bandas inespecíficas (Figura 13, ver carril 800 nm). Si bien, de acuerdo con esta observación, otros autores han reportado previamente la reacción cruzada de este anticuerpo con proteínas del complejo SWI/SNF (180 kDa) (Van Eijl et al., 2017), las bandas inespecíficas que nosotros evidenciamos (50-60 kDa) no coinciden con ese peso esperado. La caracterización más fina de la especificidad de este anticuerpo requeriría de estudios adicionales, como la inmunoprecipitación seguida de identificación de las proteínas unidas al anticuerpo por espectrometría de masas.

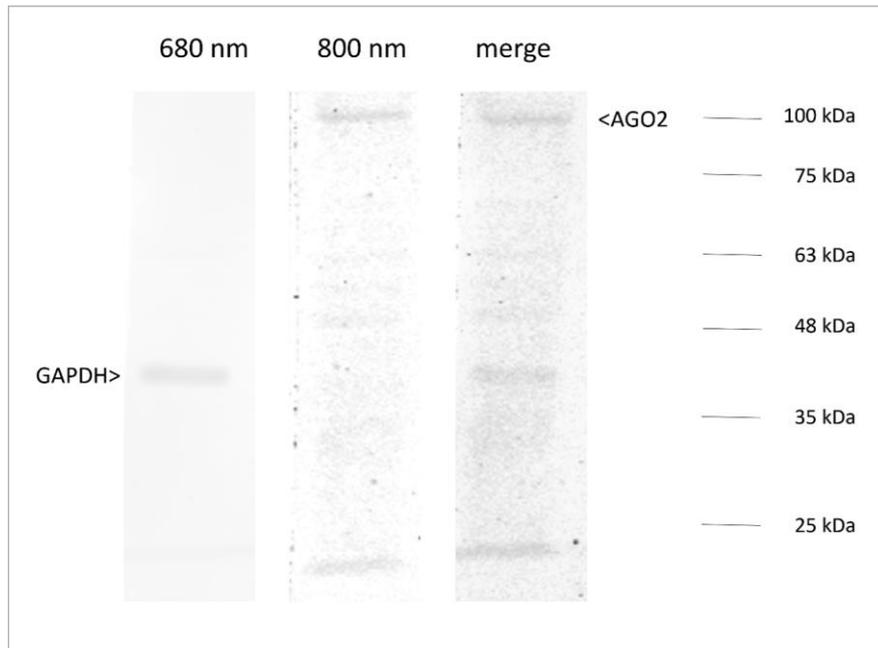


Figura 13. Western blot de extractos proteicos de la línea celular DU145 utilizando los anticuerpos anti-AGO2 y anti-GAPDH revelados a 800 nm y 680 nm respectivamente. Las membranas fueron bloqueadas con PBS 5% leche, incubadas con ambos primarios: anti-AGO2 (rata) en buffer de bloqueo y posteriormente anti-GAPDH en PBS+Tween 0.05% (conejo). Continuamos con los lavados, y luego la membrana fue incubada con los anticuerpos secundarios (anti-rata 800 nm, anti-conejo 780 nm), lavada y revelada.

Plásmido de expresión de AGO2 fusionada a un epítipo inmunogénico

El segundo abordaje propuesto en este trabajo para la inmunoprecipitación de AGO2 es la expresión forzada de esta proteína fusionada a algún epítipo inmunogénicamente fuerte. Para esto utilizamos el plásmido comercial [pFRT/FLAG/HA-DEST EIF2C2](#) que contiene la secuencia de ADNc de AGO2 clonada (sin intrones) y ha sido utilizado ya con este propósito (Landthaler et al., 2008). No utilizaremos este plásmido para la transfección en células porque solo funciona en líneas celulares que tenga previamente clonado el sistema Flp-In™, y nosotros no lo tenemos en el laboratorio. Por lo tanto, aprovechamos este plásmido para obtener la secuencia de AGO2 y clonarla en un plásmido que dirija la expresión de la proteína en nuestras líneas celulares de interés. Utilizando técnicas clásicas de biología molecular, clonamos esta CDS en el vector p3XFLAG-CMV-14 bajo la regulación del promotor T7, para luego usarla para transfectar líneas celulares de cáncer de próstata (Figura 14).

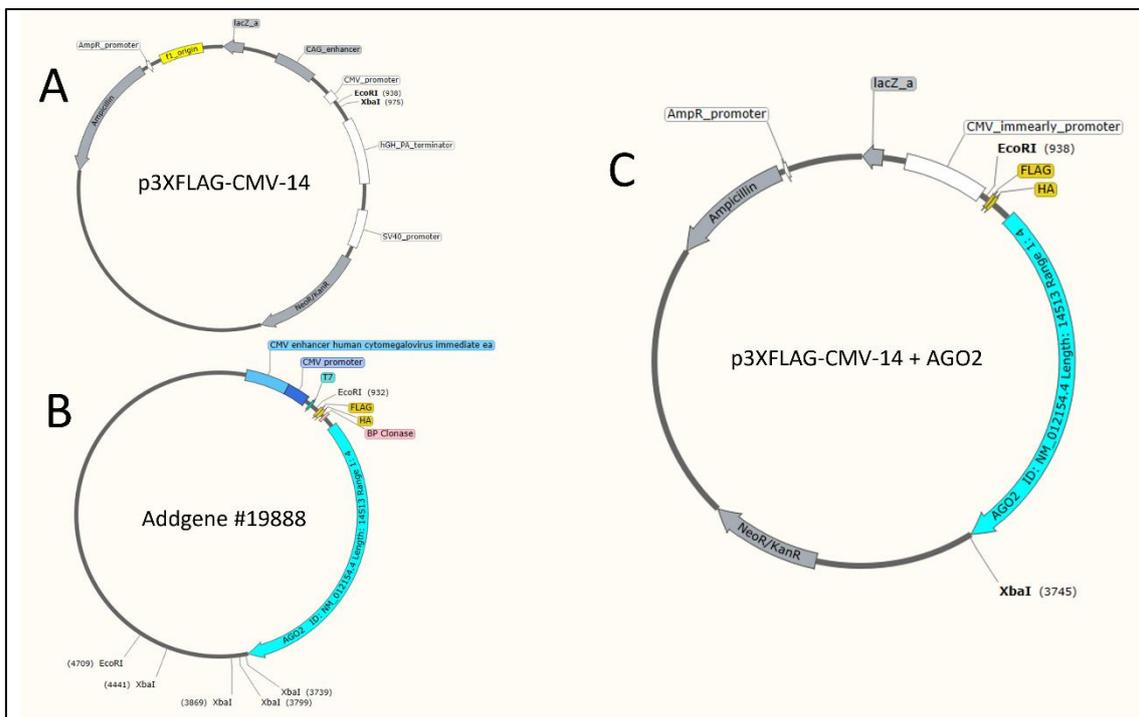


Figura 14. Representación esquemática de los plásmidos utilizados para la expresión de AGO2. En la imagen se indican las secuencias importantes para el desarrollo del protocolo en los tres plásmidos utilizados. **A)** Plásmido receptor de la secuencia de AGO2. **B)** Plásmido donante de la secuencia de AGO2 adquirido de Addgene. **C)** Plásmido generado para transfectar las líneas celulares. Imágenes generadas en el programa SnapGene (versión 6.0.2).

El primer paso fue obtener del repositorio [Addgene](#) el vector comercial pFRT/FLAG/HA-DEST EIF2C2 con la secuencia de *AGO2*. Recibimos bacterias transformadas crecidas en forma de *stab*, para partir de las cuales inoculamos medio de cultivo LB con el antibiótico correspondiente, en este caso ampicilina, y realizamos la extracción del plásmido mediante el protocolo de Miniprep (Ausubel et al., 2011). El rendimiento fue satisfactorio (300 µg/µL, concordante con el protocolo) y permitió continuar con los pasos de digestión para liberar el fragmento de ADN que contiene la secuencia de *AGO2* con las enzimas de restricción *EcoRI* y *XbaI* (resultado no mostrado). El resultado de esta digestión fue analizado por electroforesis en gel de agarosa visualizado por tinción con bromuro de etidio y el fragmento liberado con la secuencia de *AGO2* fue purificado de gel (resultado no mostrado). Como resultado de la purificación obtuvimos 50 µL de una solución con una concentración de ADN de 132.8 µg/µL y una relación de absorbancia 260/280 nm de 1.84, lo que nos permite continuar con el procedimiento. Por otro lado, para amplificar el plásmido receptor p3XFLAG-CMV-14 transformamos bacterias quimio competentes y luego lo purificamos de cultivos en LB-ampicilina por el método de Miniprep (Ausubel et al., 2011). Para prepararlo para el clonado, el plásmido fue digerido por doble digestión utilizando las mismas enzimas con que se escindió el plásmido donante del cDNA de *AGO2* (*EcoRI* y *XbaI*). El plásmido linealizado se aisló de un gel de agarosa sometido a electroforesis, obteniéndose una solución con una concentración de ADN de 332 ng/µL y una relación de absorbancia 260/280nm de 1.7. Para disminuir la probabilidad de recircularización del plásmido y, por ende, aumentar la cantidad de bacterias transformadas con el plásmido conteniendo el inserto, desfosforilamos el plásmido receptor linealizado por incubación con fosfatasa alcalina. Luego, procedimos a ligar el plásmido receptor y el inserto previamente purificado, realizamos la transformación de bacterias mediante electroporación y el aislamiento de colonias resistentes a la ampicilina. Posteriormente, se usó PCR de colonias para determinar la presencia de plásmidos recombinantes con inserto en las colonias. Para esta PCR utilizamos dos cebadores, uno dirigido a una secuencia dentro del inserto y otro en el plásmido receptor (M13, AGO2AmpCorto) que amplifican un producto de 1088 pb. De esta manera, la amplificación ocurre sólo si se introdujo en el plásmido el inserto en el sitio deseado. De

las 12 bacterias transformantes monitoreadas, 11 eran resistentes a ampicilina debido a la re-circularización del plásmido, y solo una contenía el plásmido con el inserto de interés (carril 12, Figura 15). Este rendimiento puede estar subestimado debido a inhibición de la PCR de colonia, que no fue controlada en nuestro procedimiento por la amplificación de un gen endógeno bacteriano y/o el análisis de diluciones de la muestra. Subsecuentemente, a partir de la colonia positiva y con el fin de obtener suficiente plásmido para las posteriores transfecciones, usamos el protocolo para Maxiprep (Ausubel et al., 2011). Obtuvimos una concentración de 740 ng/ μ L de plásmido a una relación 260/280nm de 1.65. Para confirmar la presencia del inserto en el plásmido generado realizamos un análisis por digestión con las mismas enzimas con que clonamos (*EcoRI* y *XbaI*) y por otro analizamos parte de su secuencia. La doble digestión confirmó la presencia del inserto del tamaño esperado en el plásmido receptor (Figura 16). A su vez, la secuenciación permitió la lectura de las secuencias perfectas del vector, el inserto en la zona del clonado con las enzimas de restricción, la etiqueta inmunogénica FLAG y la secuencia perfecta de al menos la mitad del cDNA de *AGO2* (Figura 17).

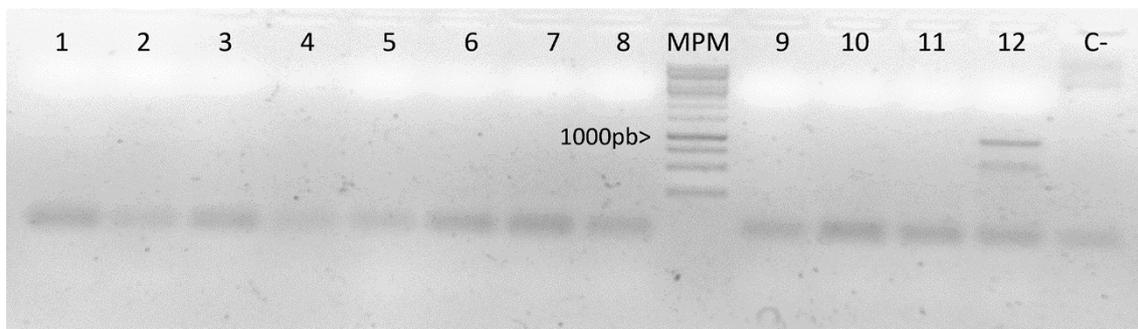


Figura 15. Identificación de colonias bacterianas transformadas con el plásmido recombinante p3XFLAG-CMV-14- AGO2 por PCR de colonias. Electroforesis en gel de agarosa 1% de la PCR de 12 colonias indicadas. El control negativo (C-) es la PCR utilizando como molde el plásmido p3XFLAG-CMV-14 sin el inserto. MPM: *GeneRulerTM 1 Kb DNA Ladder*.

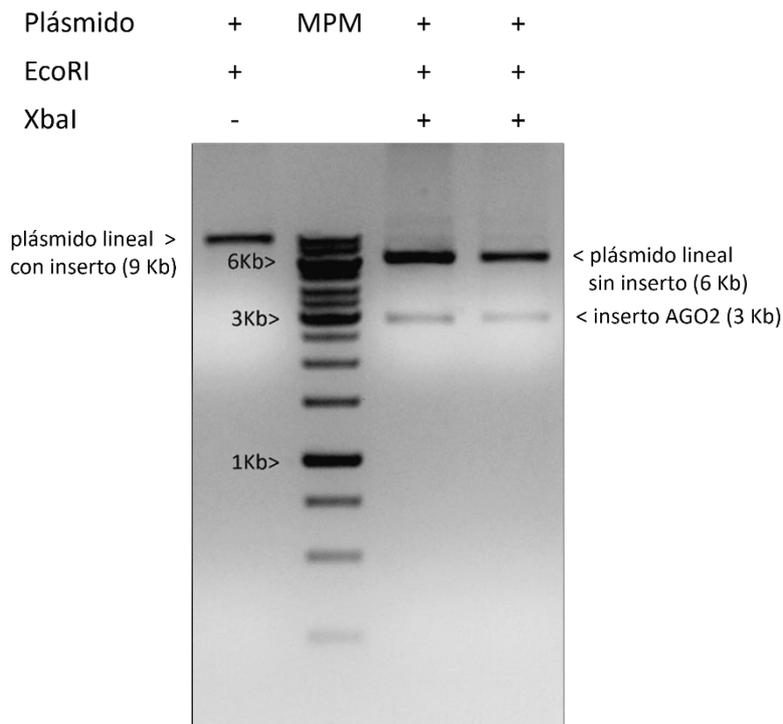


Figura 16. Confirmación de la presencia de plásmido recombinante *p3XFLAG-CMV-14- AGO2* en la colonia seleccionada. Electroforesis en gel de agarosa 1% del plásmido *p3XFLAG-CMV-14-AGO2* digerido con *EcoRI* y *XbaI*. MPM: GeneRuler™ 1 Kb DNA Ladder.

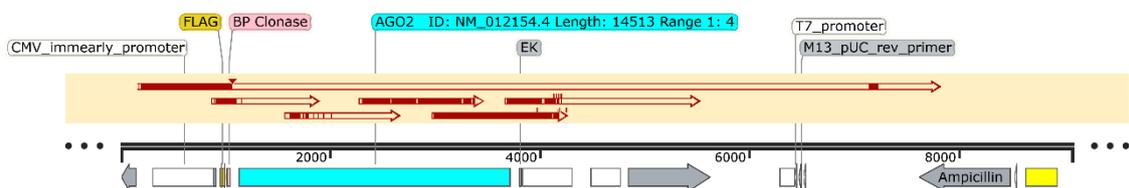


Figura 17. Análisis de la secuencia del plásmido *p3XFLAG-CMV-14- AGO2*. Las flechas indican la secuencia leída del plásmido. El color rojo indica que se logró alinear la lectura con la secuencia de referencia, mientras que el color blanco indica que la lectura no corresponde perfectamente con la secuencia de referencia porque la calidad de la secuenciación disminuye en esa región. De izquierda a derecha los *primers* tienen el siguiente orden: M13, Seq1, Seq2, Seq3, Seq4, Seq5. Captura de pantalla del programa *SnapGene 6.0.2* realizada sobre la secuencia del plásmido construido alineadas con las secuencias determinadas por la secuenciación de Sanger de las regiones indicadas.

El siguiente paso fue transfectar la línea celular de próstata [RWPE1](#) con el plásmido generado, para lo que utilizamos [Lipofectamine 3000](#) siguiendo las indicaciones del fabricante (con modificación, ver Materiales y Métodos). A modo de control de la transfección utilizamos en paralelo el plásmido [pGFPmax](#), el cual tiene clonada la

secuencia de la proteína fluorescente verde (GFP, de sus siglas en inglés) bajo el control de un promotor CMV. Esto nos permite evaluar y calcular la eficiencia de transfección. Este control no lo realizamos con p3XFLAG-CMV-14 directamente porque no tiene un marcador fluorescente que podamos evaluar por microscopía. La eficiencia de transfección estimada por observación microscópica fue de aproximadamente 20 % del cultivo. Comenzamos con la etapa de selección utilizando el antibiótico neomicina a una concentración de 300 µg/mL (primeros 7 días post transfección). En los días 8 y 9 utilizamos menor concentración de antibiótico (200 µg/mL) debido a la baja confluencia de las células causada por la muerte de aquellas que no incorporaron el plásmido. Por último, en los días del 10 al 15, el antibiótico se redujo a 100 µg/mL del fármaco. En ese momento se observaron colonias individuales distanciadas, varias de las cuales fueron aisladas utilizando [TrypLE™ Express Enzyme](#) y O-rings para generar líneas celulares transfectadas clonales. Tres de las líneas aisladas fueron luego expandidas y nombradas como 1, 3 y 4. Para confirmar que la resistencia farmacológica adquirida se debió a la incorporación estable del plásmido transfectado en el genoma, y no a una adaptación al fármaco, se analizó su presencia mediante qPCR. En primera instancia para el PCR de colonias se usó un oligonucleótido específico para la región del FLAG de la proteína, pero no fue posible obtener un producto de amplificación único a partir de ADN genómico de la línea RWPE-1 (resultado no mostrado). Por este motivo se usó nuevos cebadores ahora complementarios a la secuencia de *AGO2* en dos exones consecutivos. Esto tiene la desventaja de amplificar conjuntamente los transcriptos del gene *AGO2* endógeno y los provenientes del constructo *AGO2* etiquetado plasmídico. Analizando la expresión de *AGO2* por RT-qPCR, observamos que las tres clonas transfectantes aisladas tuvieron un *fold-change* de 4.1 (línea 1), 24.5 (línea 3) y 7.8 (línea 4) con respecto a la relativa a la línea RWPE1 sin transfectar (Figura 18). Restaría confirmar la expresión de la proteína AGO-FLAG por ensayos de *Western blot* utilizando un anticuerpo anti-FLAG.

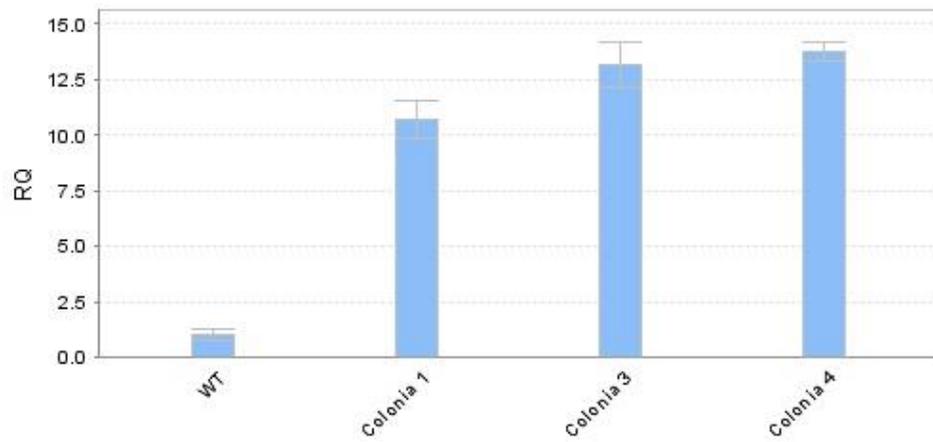


Figura 18. Análisis cuantitativo de la expresión total de AGO2 en las líneas celulares que incorporaron el plásmido CMV-AGO2-FLAG aisladas. RT- qPCR de la línea celular RWPE-1. Cuantificación relativa (RQ, por sus siglas en inglés) del transcripto AGO2 de las colonias aisladas versus el control (WT: RWPE1 sin transfectar). Se utilizó como gen normalizador el transcripto de TBP.

Sistema CRISPR/Cas

La aproximación anterior muestra que la expresión de proteínas recombinantes dirigidas por plásmidos produce un aumento de la cantidad de proteína, porque se suma a la expresión de los dos alelos endógenos, lo que puede provocar interacciones no fisiológicas entre AGO2 y ARNs. Otro inconveniente es que la proteína endógena compite con la etiquetada en los ensayos de inmunoprecipitación bajando la eficiencia del ensayo. Por último, la AGO-etiquetada se expresa en forma constitutiva independientemente de los controles normales de la célula por lo que no puede usarse para estudiar situaciones en las que haya cambios en la regulación transcripcional de AGO2. Es por esto que, en paralelo al abordaje de etiquetado mediado por plásmido, quisimos implementar el sistema CRISPR/Cas como herramienta de edición genómica.

Parte de este trabajo se realizó en el contexto de la tesina de grado titulada “Construcción de un sistema *in vitro* para el etiquetado endógeno de AGO2 humana mediante edición genómica por CRISPR/Cas9 recombinante”, realizada por Lic. Lucia Colantuono para la Licenciatura en Ciencias Biológicas, en la cual actué de co-tutor junto con la Dra. María Ana Duhagon en el rol de tutora.

Síntesis de SaCas9 recombinante

Como se anticipa en los apartados anteriores, en este trabajo decidimos utilizar la proteína Cas9 recombinante para luego transfectarla en el modelo de interés. Esta tecnología tiene la ventaja de que la Cas9 no queda constitutivamente expresada en el modelo (evitando la acumulación de ediciones *off-target*) y además permite extrapolar la técnica de manera sencilla a otros modelos (cambiando la secuencia del sgARN). Para lograr este objetivo transformamos cepas de *E. coli* Rosetta™2 con el plásmido comercial inducible p6xHis_NLS-SaCas9. A partir de la inducción con IPTG de 5.4 L de cultivo se obtuvieron seis *pellets* de 9.0 g de peso húmedo derivados c/u de 900 mL de cultivo, lo que correspondió a un rendimiento de 10 g/L de cultivo.

Las diferentes etapas del proceso de expresión de la proteína recombinante fueron monitoreadas por *Western blot* (Figura 19). Podemos dividir el proceso en los siguientes puntos: las células sin transformar (comúnmente referidas como *wild type*), células

transformadas con el plásmido sin inducir e inducidas. La muestra de proteínas no transformadas (carril 1) y bacterias transformadas sin inducir (carril 2) presentan el mismo perfil proteico, indicando que el plásmido no tiene expresión espuria y no modifica el normal funcionamiento de las bacterias. Esto permite un buen crecimiento bacteriano, evitando estrés traduccional o toxicidad previa a la inducción con IPTG. Es así como no se observa alteración en el perfil proteico en ninguna de las muestras a excepción de la expresión de la SaCas9 luego de la inducción con IPTG (carril 3, carril 4, carril 5). Se obtuvo SaCas9 tanto en la fracción soluble como insoluble de las bacterias inducidas, lo que sugiere un correcto plegamiento de aproximadamente la mitad de la proteína SaCas9 expresada por las bacterias y habilita continuar con los siguientes pasos.

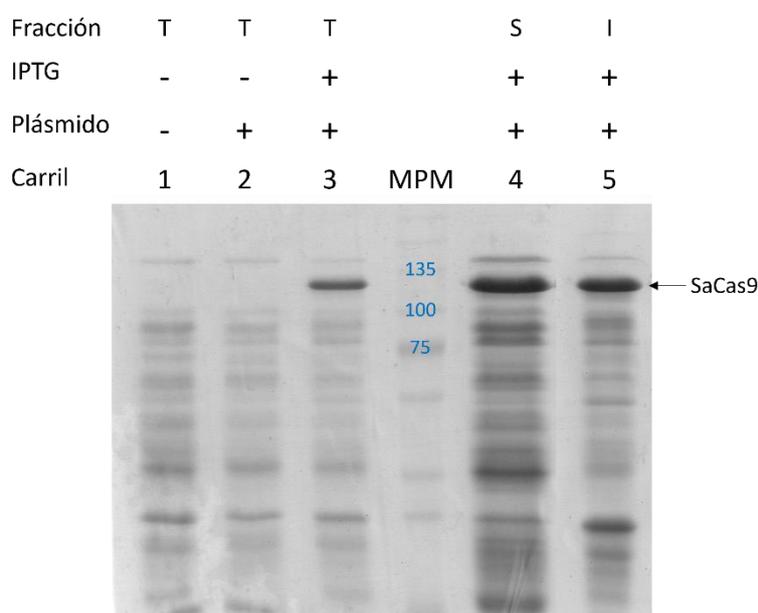


Figura 19. Expresión de SaCas9 en las bacterias transformadas con el plásmido p6xHis_NLS-SaCa. Perfil proteico obtenido a largo del proceso de inducción de la expresión de SaCas9 en bacterias transformadas con el plásmido p6xHis_NLS-SaCas9. Electroforesis SDS-PAGE teñida con azul de Coomassie de extractos de bacterias transformadas con el plásmido de expresión de SaCas9. Fracción: total (T), soluble (S) e insoluble (I). En el marcador de peso molecular (MPM) se indica el tamaño de las tres bandas relevantes para la identificación de SaCas9 (75, 100 y 135 kDa). Se indica con una flecha la banda correspondiente a la SaCas9.

Se procedió con la purificación de SaCas9 a partir de la fracción soluble, mediante cromatografía de afinidad utilizando una columna de níquel (*HisTrap HP* de 1mL, *GE™*). En la Figura 20 se observan las distintas etapas correspondientes al proceso de

purificación que consta de cargar la columna con la fracción soluble para luego ir aumentando gradualmente la concentración de imidazol (competidor) y así eluir las proteínas en orden creciente de afinidad. Este aumento gradual está dividido en tres lavados con concentraciones crecientes (carriles 2-4) de 53 mM, 82 mM y 124 mM de imidazol respectivamente). Estos pasos se los agregamos al protocolo original que indicaba solo un lavado por volumen con 50mM de imidazol previo a la elución. Para el lavado final (según el protocolo “elución”) se utilizó 500 mM de imidazol (carril 5) (Burlle-Caldas et al., 2018). A modo de control positivo, incluimos a la proteína SaCas9 producida en el laboratorio de la Profesora Santuza Teixeira (Universidad Federal de Minas Gerais) en el análisis electroforético (carril 6). En el análisis por SDS-PAGE se puede apreciar que la SaCas9 purificada coincide en peso molecular de la SaCas9 control (~130 kDa). En los primeros pasos con 30 mM de imidazol se colectaron las proteínas que no se unen a la columna (carril 1), fracción denominada comúnmente *flow-throw*. La concentración más alta de SaCas9 fue obtenida en el lavado correspondiente a 82 mM de imidazol (carril 3) y, por lo tanto, fue elegida para continuar con los pasos posteriores; desalado y ensayos de actividad. En esta elución se pueden observar bandas de menor tamaño que podrían corresponder a proteínas contaminantes que eluyen con la SaCas9 y/o degradación parcial de la proteína recombinante. También se observa la presencia de la SaCas9 en los carriles 2 y 4, en concentraciones considerablemente menores comparado con el carril 3. En la elución final con 500 mM de imidazol (carril 5) no se visualizan bandas.

El imidazol fue removido utilizando la columna PD-10 en el proceso de desalado mediante purificación por gel filtración y, posteriormente, se aumentó su concentración de la SaCas9 utilizando centrífuga *Milipore* alcanzando una concentración final de 2.4 µg/µL (en total se obtuvo 24 mg de proteína). La fracción insoluble fue descartada ya que el rendimiento de la fracción soluble fue satisfactorio.

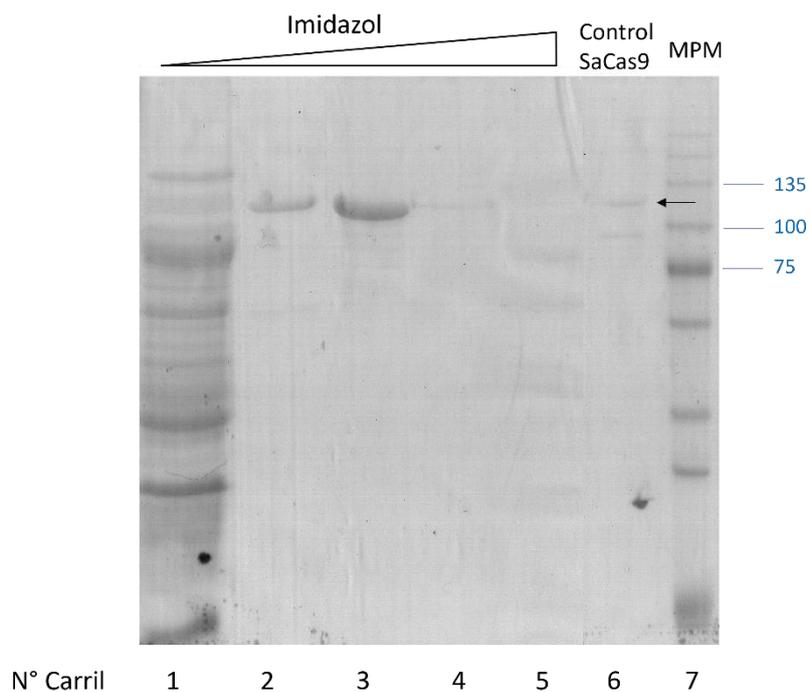


Figura 20. Perfil de elusión de la SaCas9 purificada por columna de níquel His-Trap. La flecha negra indica el peso molecular esperado para la SaCas9 según el control. Electroforesis SDS-PAGE de fracciones de cromatografía de afinidad por níquel. Carril 1- *flow-throw* (30 mM imidazol) Carril 2-5: 53 mM, en 82 mM, 124 mM y 500mM de imidazol respectivamente, Carril 6: SaCas9 control, Carril 7: marcador de peso molecular (MPM) se indica el tamaño de las tres bandas relevantes para la identificación de SaCas9 (75, 100 y 135 kDa).

Si bien el grado de pureza de la proteína eluida con 82mM imidazol (carril 3) no fue cuantificado, a modo cualitativo se observa una única banda claramente mayoritaria en los geles teñidos con azul de Coomassie, lo cual consideramos aceptable para proseguir estudiando su actividad nucleasa (utilizado así en Burde-Caldas et al., 2018).

Evaluación de la actividad de la SaCas9

Con el propósito de comprobar si la proteína SaCas9 purificada era capaz de generar un DSB en los genes *mCherry* y *AGO2* se la incubó en conjunto con un ADN blanco amplificado por PCR y un sgARN para cada gen. El ADN diana para el corte de Cas9 en la región 3' del gen *AGO2* se sintetizó por PCR desde ADN, utilizando las condiciones indicadas en Materiales y Métodos y dio lugar a un producto único de amplificación específica del tamaño esperado (resultado no mostrado). La reacción de PCR se escaló para obtener una cantidad y una pureza de amplicón adecuada para el análisis del corte

mediado por SaCas9 mediante electroforesis (resultados no mostrados). De la misma manera, el ADN diana para el corte del ADN correspondiente al inicio del gen *mCherry* fue generado por PCR a partir de ADN plasmídico del vector que contiene este gen, lo que amplificó un producto único del tamaño esperado, escalándose la reacción para producir cantidades adecuadas para el análisis electroforético del corte endonucleotídico por Cas9 (resultados no mostrados). La síntesis de los sgARN se realizó según los protocolos comerciales obteniéndose 652 ng/ μ L (*mCherry*) y 1272 ng/ μ L de ARN de tamaños esperados por reacción según análisis electroforético y espectrofotométrico (resultados no mostrados). Una vez obtenidos así los componentes necesarios para el corte endonucleotídico sitio específico mediado por Cas9 en presencia de los sgARN diseñados para los dos ADN diana ensayados, se evaluó *in vitro* el corte por la proteína SaCas9 producida en este trabajo. Para esto se incubaron los 3 componentes sintetizados (Cas9, los sgARN y los ADN diana) en las condiciones indicadas y los productos de las reacciones fueron analizados por electroforesis en gel de agarosa 2%. De esta forma, conociendo el tamaño del ADN blanco y la ubicación del sitio de corte, se puede verificar la presencia y tamaño de los fragmentos esperados comprobando así la funcionalidad de la SaCas9. En la Figura 21 se muestra el resultado del ensayo de actividad para el gen *mCherry*. Para las reacciones de digestión (SaCas9+ADN blanco+sgARN) se utilizó la mitad de concentración de los componentes en relación con la requerida por el protocolo (Burle-Caldas et al., 2018), obteniendo así un volumen final de 10 μ L en cada una. Los dos primeros carriles corresponden a los controles y los carriles 3-4 a digestiones. Como resultado, obtuvimos una digestión exitosa del ADN diana, aunque parcial dado que no logramos digerir la totalidad del ADN. Evidencia de esto es el carril 4, donde se pueden observar las 2 bandas del tamaño esperado para la digestión del ADN molde (200 y 160 pb, señalado con flechas negras) al igual que la banda correspondiente al ADN molde (360 pb). Estos productos no se observan en ausencia de sgARN (carril 1), lo que evidencia el estricto requerimiento de estos para la actividad nucleotídica de la SaCas9. Si comparamos los carriles 3 y 4 podemos concluir que el uso de proteinasa K y ARNasa A mejora la visualización de los resultados. La proteinasa K contribuye desarmando los complejos Cas9+ADN (Sternberg et al., 2014) y la ARNasa A degrada el sgARN que, en este caso, dificulta la visualización de los fragmentos digeridos por presentar una

migración similar. Ambos pasos fueron agregados en este trabajo al protocolo utilizado como referencia (Burle-Caldas et al., 2018).

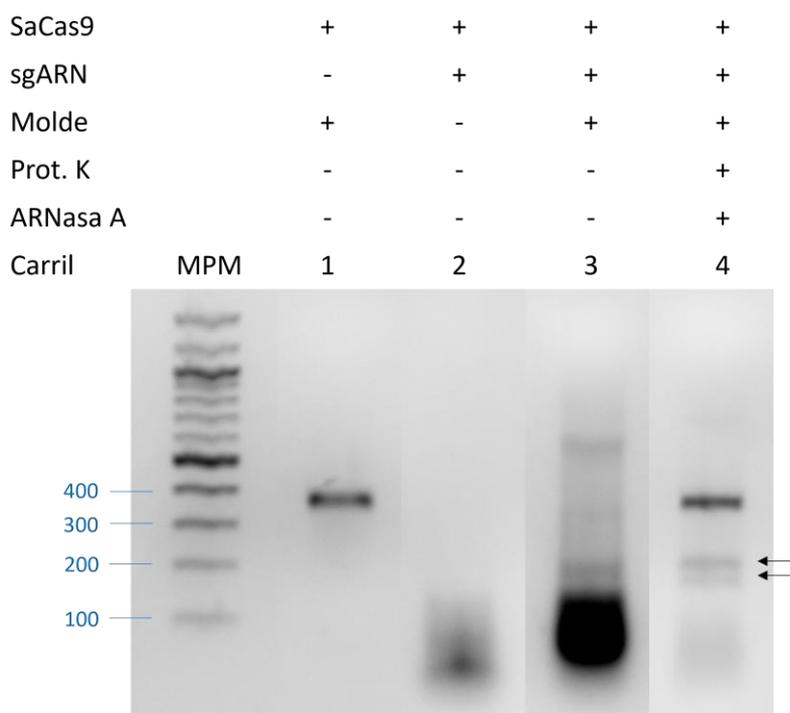


Figura 21. Ensayo de digestión del ADN molde del gen mCherry por SaCas9 recombinante. Gel de agarosa 2% teñido con bromuro de etidio. Carril 1: DNA molde, Carril 2: sgRNA+SaCas9, Carril 3: digestión sin ARNasa, Carril 4: digestión con ARNasa. En el marcador de peso molecular (MPM) se indica el tamaño de las tres bandas relevantes (100, 200, 300 y 400 pb). Se indican los fragmentos derivados de la digestión con flechas negras.

Los resultados aquí mostrados no permiten discernir si la limitada eficiencia del corte de SaCas9 para el gen *mCherry* se debe al diseño de los sgARNs, a condiciones de la reacción (pureza de los reactivos, estequiometría, tamaño del ADN molde u otros), o a la actividad de la SaCas9 recombinante producida. Sin embargo, hemos realizado dos controles adicionales que permitieron desestimar estos dos últimos factores. En condiciones de reacción idénticas el sgARN para el gen *GP72* de *T. cruzi*, tomado de Burle-Caldas et al. 2018, dio lugar a un corte 100% del ADN diana (resultados de la tesina de grado de la Lic. Selene Piriz). Por otro lado, una SaCas9 recombinante comercial, adquirida a NEB no logró incrementar la eficiencia de este corte (resultados no mostrados). Estos controles sugieren que el problema estaría en el sgARN o en factores de la reacción específicos para

la interacción entre el ADN y el sgARN específicos diseñados para *mCherry* en nuestro trabajo.

A pesar de que los reactivos y procedimientos utilizados en este trabajo mostraron ser efectivos para la digestión del gen control *mCherry*, dirigido por la Cas9, no conseguimos el corte buscado en el gen que codifica para AGO2 con los sgARNs que diseñamos (Figura 22). Como se puede apreciar en la Figura 20, no se observan las bandas esperadas por el corte de la SaCas9 (928 y 827 pb). Sin embargo, se puede ver que la intensidad de la banda molde es menor al control. También, parece haber ADN en el pocillo indicando que la proteinasa K incluida en la incubación, no logró desarmar los complejos resultando en la acumulación de estos en el comienzo del gel. Estas dos observaciones no llevan a plantearnos la hipótesis de que el complejo SaCas9+sgARN se une al molde, pero sin capacidad de DSB, fenómeno que ha sido descrito previamente (Sternberg et al., 2014). Consideramos que probablemente el sgARN sea la principal determinante del fracaso porque su diseño presentó algunas dificultades. Por ejemplo, el alto contenido GC de la región N-terminal de AGO2 implicó el diseño de un sgARN que superó ampliamente el máximo porcentaje GC recomendado (40-60%) resultando en 80% (Concordet & Haeussler, 2018). Como nuestro objetivo es generar un etiquetado en la región N-terminal de la proteína, estamos restringidos a utilizar sgARNs que estén cerca del codón de inicio. Según los análisis en CRISPOR, existen dos posibles sgARNs que cumple dicha condición; ambas secuencias guías poseen un contenido GC de 80%, una puntuación de especificidad de aproximadamente 100/100 y se predice un valor de eficiencia de 73/100 para el sgARN1 y de 76/100 para el sgARN2. Adicionalmente, la secuencia guía del sgARN1 presenta veinte posibles sitios *off-targets* en genoma humano, mientras que el sgARN2 presenta doce. Decidimos optar por el sgARN1 ya que no presenta grandes diferencias con el sgARN2 y guía un corte doble hebra 6 nucleótidos más cerca del codón de inicio de AGO2 aumentando la probabilidad de HDR (Liang et al., 2017). Creemos que la modificación al protocolo más pertinente es el uso de otro sgARN. En este contexto, existe la posibilidad de utilizar un solo sgARN distante del sitio de edición, esto implicaría una mejor sgARN pero una menor eficiencia de edición. Alternativamente, existe la posibilidad de utilizar dos sgARN en simultáneo resultando en la escisión de una región del gen y direccionar la recombinación utilizando los lugares de corte como homología

(Vazquez et al., 2018). Esto requerirá sin embargo un esfuerzo adicional para construir un ADN molde de homología de mayor tamaño, además de testar 2 nuevos sgARNs.

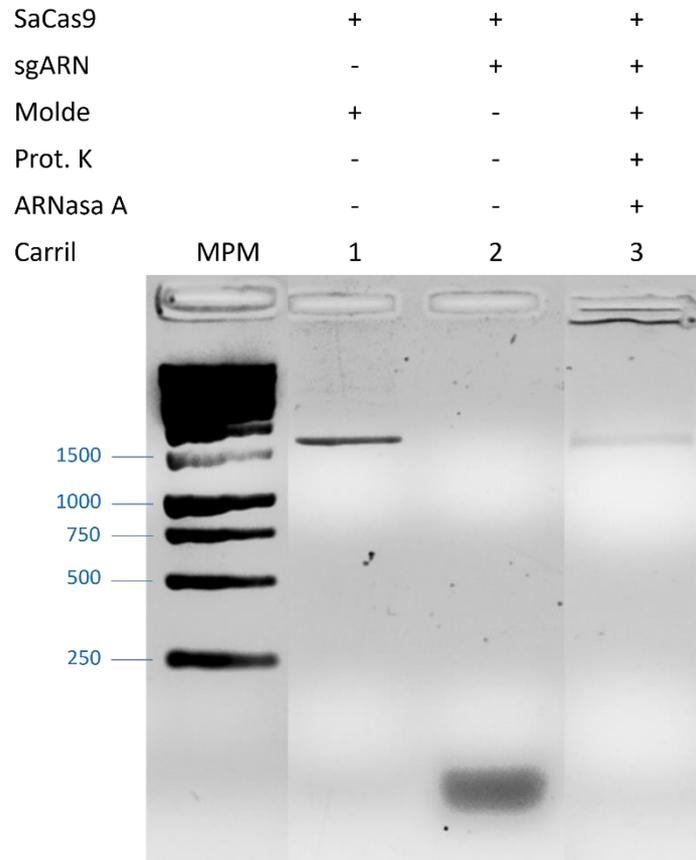


Figura 22. Ensayo de digestión del ADN molde de AGO2 por SaCas9 recombinante. Gel de agarosa 2% teñido con bromuro de etidio. Carril 1: DNA molde, Carril 2: digestión sin ARNasa, Carril 3: digestión con ARNasa. En el marcador de peso molecular (MPM) se indica el tamaño de las bandas relevantes (250, 500, 750, 1000 y 1500 pb).

Conclusiones y Perspectivas

Abordaje bioinformático

La interacción entre un microARN y su ARNm diana puede deducirse de la regulación de los niveles de proteína y/o transcripto provocada por la modulación (sobre expresión/inhibición) de la abundancia de un microARN y han sido aplicados ampliamente para identificar entre otras cosas los mecanismos de interacción de microARNs-ARNm (Baek et al., 2008; Grimson et al., 2007; Hafner et al., 2010; Helwak et al., 2013; Selbach et al., 2008). Estos enfoques experimentales tienen como desventaja estar limitados por el uso de pocos y seleccionados microARN en condiciones controladas de sobreexpresión o silenciamiento en células en cultivo que muchas veces escapan de la realidad fisiológica de los organismos. Incluso, algunos de estos tampoco involucran el contexto celular, siendo realizados en ensayos completamente *in vitro* que incluyen pocos componentes de la interacción (generalmente secuencias cortas representando al ARNm, el microARN no siempre completo y pocas veces AGO2) (McGeary et al., 2019; Sheu-Gruttadauria et al., 2019). Nuestro abordaje no presenta esas limitaciones permitiendo contribuir desde otra perspectiva a comprender la interacción entre estos ARNs. En este trabajo logramos corroborar que el análisis de las correlaciones de microARN-ARNm en grandes conjuntos de datos es capaz de validar algunas reglas ya establecidas de apareamiento de microARN-ARNm. Con esta herramienta profundizamos en el estudio de una característica de la interacción menos establecida: la contribución de la región suplementaria a la represión mediada por las semillas de microARNs. Nuestros resultados muestran que la región suplementaria incrementa los niveles de represión en términos generales, aunque alcanza diferencias estadísticamente significativas solo en los análisis de 6mer y 7mer_{m8}, probablemente por el tamaño de muestreo. Surge como hallazgo interesante, que la región suplementaria aumenta la actividad de la semilla de 6mer, haciendo comparable su represión con la de un sitio 7mer-A1. Sin embargo la semilla 6mer es clasificada como *marginal site* (Bartel, 2009), un sitio de prácticamente nula actividad, no suelen ser considerada a la hora de seleccionar ARNm diana. Por otro lado, en contraste a informes anteriores (Bartel, 2009; Grimson et al., 2007; Nielsen et al., 2007), encontramos que la semilla 7mer-A1 es

significativamente más represiva que la 7mer-m8 en las tres regiones analizadas. Para comprender si este hallazgo inesperado se debe a la naturaleza de los modelos analizados (*in vitro* frente a *in vivo*), o a los distintos parámetros establecidos en la búsqueda, se necesita más investigación. Probablemente debamos en primera medida evaluar los sitios utilizados en *TargetScan* con su ponderación de contexto, así como los niveles de conservación evolutiva de la región suplementaria y, además, incorporar el puente semilla-suplementaria de hasta 15 nucleótidos que hasta el momento no ha sido considerado en su algoritmo.

Los resultados de esta sección se encuentran en formato de manuscrito titulado "[Transcriptome-wide analysis of microARN-ARNm correlations in unperturbed tissue transcriptomes identifies microARN targeting determinants](#)" que ha sido sometido a revisión y está accesible en Biorxiv (<https://www.biorxiv.org/content/10.1101/2021.12.22.473932v1>).

Abordaje experimental

El objetivo de esta sección fue poner a punto metodologías para la validación experimental de los resultados obtenidos de la interacción microARN-ARNm. Estas aproximaciones se describieron en orden de dificultad de implementación, comenzando con el uso del anticuerpo anti-AGO2 reportado en la literatura para inmunoprecipitación, pasando por la construcción de un plásmido para expresar AGO2-FLAG en células transfectadas y finalizando con el uso de la herramienta de edición genómica CRISPR/Cas9 para el etiquetado endógeno de la proteína AGO2 en líneas celulares. El anticuerpo anti-AGO2 mostró especificidad baja puesto que reconoce más de una proteína en las condiciones encontradas para el reconocimiento mayoritario de AGO2 por *Western blot*. Por lo tanto, optamos por intentar otras metodologías como la generación de una proteína AGO2 fusionada a un epítipo artificial que permita la inmunoprecipitación más específica y eficiente. En ambas metodologías logramos avances significativos. Se generaron clonas sobreexpresantes para AGO2 en las líneas celulares DU145. Se produjeron los componentes del sistema de SaCas9 recombinantes, lo que generó una metodología económica y versátil para incorporar como estrategia de

análisis de diferentes preguntas en todo tipo de organismos en nuestro laboratorio. Se ensayaron 2 diseños de sgRNAs que mostraron la actividad del sistema y la eficiencia adecuada de la edición *in vitro* del reportero *mCherry*. El sgARN para *AGO2* no resultó efectivo, y debido al alto contenido de GC y las limitaciones de la región se concluye que sería conveniente implementar una aproximación de dos sgARNs. Queda como perspectiva probar los componentes dentro de las células, para testar la eficiencia de edición *in vivo*.

Bibliografía

- Abeshouse, A., Ahn, J., Akbani, R., Ally, A., Amin, S., Andry, C. D., Annala, M., Aprikian, A., Armenia, J., Arora, A., Auman, J. T., Balasundaram, M., Balu, S., Barbieri, C. E., Bauer, T., Benz, C. C., Bergeron, A., Beroukhim, R., Berrios, M., ... Zmuda, E. (2015). The Molecular Taxonomy of Primary Prostate Cancer. *Cell*, 163(4), 1011–1025. <https://doi.org/10.1016/j.cell.2015.10.025>
- Agarwal, V., Bell, G. W., Nam, J. W., & Bartel, D. P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. *ELife*, 4(AUGUST2015), 1–38. <https://doi.org/10.7554/eLife.05005>
- Ahmed, M., Nguyen, H., Lai, T., & Kim, D. R. (2018). miRCancerdb: a database for correlation analysis between microRNA and gene expression in cancer. *BMC Research Notes*, 11(1), 103. <https://doi.org/10.1186/S13104-018-3160-9>
- Amoasii, L., Hildyard, J. C. W., Li, H., Sanchez-Ortiz, E., Mireault, A., Caballero, D., Harron, R., Stathopoulou, T.-R., Massey, C., Shelton, J. M., Bassel-Duby, R., Piercy, R. J., & Olson, E. N. (2018). Gene editing restores dystrophin expression in a canine model of Duchenne muscular dystrophy. *Science*, 1549, eaau1549. <https://doi.org/10.1126/science.aau1549>
- Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., & Struhl, K. (2011). Current Protocols in Molecular Biology. *The Quarterly Review of Biology*, 66(2), 199–200. <https://doi.org/10.1086/417163>
- Baek, D., Villén, J., Shin, C., Camargo, F. D., Gygi, S. P., & Bartel, D. P. (2008). The impact of microRNAs on protein output. *Nature*, 455(7209), 64–71. <https://doi.org/10.1038/nature07242>
- Bailey, T. L., Johnson, J., Grant, C. E., & Noble, W. S. (2015). The MEME Suite. *Nucleic Acids Research*, 43(W1), W39–W49. <https://doi.org/10.1093/NAR/GKV416>
- Bartel, D. P. (2009). MicroRNAs: Target Recognition and Regulatory Functions. *Cell*, 136(2), 215–233. <https://doi.org/10.1016/j.cell.2009.01.002>
- Bartel, D. P. (2018). *Metazoan MicroRNAs*. <https://doi.org/10.1016/j.cell.2018.03.006>
- Berezikov, E. (2011). Evolution of microRNA diversity and regulation in animals. *Nature Publishing Group*, 12(12), 846–860. <https://doi.org/10.1038/nrg3079>
- Bortesi, L., & Fischer, R. (2015). The CRISPR/Cas9 system for plant genome editing and beyond. *Biotechnology Advances*, 33(1), 41–52. <https://doi.org/10.1016/J.BIOTECHADV.2014.12.006>
- Bråte, J., Neumann, R. S., Fromm, B., Haraldsen, A. A. B., Tarver, J. E., Suga, H., Donoghue, P. C. J., Peterson, K. J., Ruiz-Trillo, I., Grini, P. E., & Shalchian-Tabrizi, K. (2018). Unicellular Origin of the Animal MicroRNA Machinery. *Current Biology*, 28(20), 3288–3295.e5. <https://doi.org/10.1016/J.CUB.2018.08.018>
- Broughton, J. P., Lovci, M. T., Huang, J. L., Yeo, G. W., & Pasquinelli, A. E. (2016). Pairing beyond the Seed Supports MicroRNA Targeting Specificity. *Molecular Cell*, 64(2), 320–333. <https://doi.org/10.1016/j.molcel.2016.09.004>
- Burle-Caldas, G. A., Soares-Simões, M., Lemos-Pechnicki, L., DaRocha, W. D., & Teixeira, S. M. R. (2018). Assessment of two CRISPR-Cas9 genome editing protocols for rapid generation of Trypanosoma cruzi gene knockout mutants. *International Journal for Parasitology*, 48(8), 591–596. <https://doi.org/10.1016/j.ijpara.2018.02.002>
- Ceccaldi, R., Rondinelli, B., & Andrea, A. D. D. (2015). Repair Pathway Choices and Consequences at the Double-Strand Break. *Trends in Cell Biology*, xx(x), 1–13. <https://doi.org/10.1016/j.tcb.2015.07.009>
- Chen, Y., & Wang, X. (2020). miRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Research*, 48(D1), D127–D131. <https://doi.org/10.1093/NAR/GKZ757>
- Concordet, J. P., & Haeussler, M. (2018). CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing

- experiments and screens. *Nucleic Acids Research*, 46(W1), W242–W245. <https://doi.org/10.1093/NAR/GKY354>
- Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., & Zhang, F. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems HHS Public Access. *Science*, 339(6121), 819–823. <https://doi.org/10.1126/science.1231143>
- de Torrenté, L., Zimmerman, S., Suzuki, M., Christopheit, M., Greally, J. M., & Mar, J. C. (2020). The shape of gene expression distributions matter: how incorporating distribution shape improves the interpretation of cancer transcriptomic data. *BMC Bioinformatics*, 21(21), 1–18. <https://doi.org/10.1186/S12859-020-03892-W/FIGURES/7>
- Doudna, J. A., & Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. *Science*, 346(6213). <https://doi.org/10.1126/science.1258096>
- Eulalio, A., Huntzinger, E., & Izaurralde, E. (2008). Getting to the Root of miRNA-Mediated Gene Silencing. *Cell*, 132(1), 9–14. <https://doi.org/10.1016/J.CELL.2007.12.024>
- Fajrial, A. K., He, Q. Q., Wirusanti, N. I., Slansky, J. E., & Ding, X. (2020). A review of emerging physical transfection methods for CRISPR/Cas9-mediated gene editing. *Theranostics*, 10(12), 5532. <https://doi.org/10.7150/THNO.43465>
- Fort, R. S., Mathó, C., Oliveira-Rizzo, C., Garat, B., Sotelo-Silveira, J. R., & Duhagon, M. A. (2018). An integrated view of the role of miR-130b/301b miRNA cluster in prostate cancer. *Experimental Hematology and Oncology*, 7(1). <https://doi.org/10.1186/s40164-018-0102-0>
- Friedman, R. C., Farh, K. K. H., Burge, C. B., & Bartel, D. P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. *Genome Research*, 19(1), 92–105. <https://doi.org/10.1101/gr.082701.108>
- Goldman, M. J., Craft, B., Hastie, M., Repečka, K., McDade, F., Kamath, A., Banerjee, A., Luo, Y., Rogers, D., Brooks, A. N., Zhu, J., & Haussler, D. (2020). Visualizing and interpreting cancer genomics data via the Xena platform. *Nature Biotechnology*, 38(6), 675–678. <https://doi.org/10.1038/s41587-020-0546-8>
- Grimson, A., Farh, K. K. H., Johnston, W. K., Garrett-Engele, P., Lim, L. P., & Bartel, D. P. (2007). MicroRNA Targeting Specificity in Mammals: Determinants beyond Seed Pairing. *Molecular Cell*, 27(1), 91–105. <https://doi.org/10.1016/j.molcel.2007.06.017>
- Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jungkamp, A. C., Munschauer, M., Ulrich, A., Wardle, G. S., Dewell, S., Zavolan, M., & Tuschl, T. (2010). Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP. *Cell*, 141(1), 129–141. <https://doi.org/10.1016/j.cell.2010.03.009>
- Hamilton, M. P., Rajapakshe, K., Hartig, S. M., Reva, B., McLellan, M. D., Kandoth, C., Ding, L., Zack, T. I., Gunaratne, P. H., Wheeler, D. A., Coarfa, C., & McGuire, S. E. (2013). Identification of a pan-cancer oncogenic microRNA superfamily anchored by a central core seed motif. *Nature Communications*, 4, 1–13. <https://doi.org/10.1038/ncomms3730>
- Hamilton, M. P., Rajapakshe, K. I., Bader, D. A., Cerne, J. Z., Smith, E. A., Coarfa, C., Hartig, S. M., & McGuire, S. E. (2016). The Landscape of microRNA Targeting in Prostate Cancer Defined by AGO-PAR-CLIP. *NEO*, 18(6), 356–370. <https://doi.org/10.1016/j.neo.2016.04.008>
- Helwak, A., Kudla, G., Dudnakova, T., & Tollervey, D. (2013). Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell*, 153(3), 654–665. <https://doi.org/10.1016/j.cell.2013.03.043>
- Huang, H. Y., Lin, Y. C. D., Li, J., Huang, K. Y., Shrestha, S., Hong, H. C., Tang, Y., Chen, Y. G., Jin, C. N., Yu, Y., Xu, J. T., Li, Y. M., Cai, X. X., Zhou, Z. Y., Chen, X. H., Pei, Y. Y., Hu, L., Su, J. J., Cui, S. D., ... Huang, H. D. (2020). miRTarBase 2020: updates to the experimentally validated microRNA–target interaction database. *Nucleic Acids Research*, 48(D1), D148–D154. <https://doi.org/10.1093/NAR/GKZ896>
- Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., & Doudna, J. (2013). RNA-programmed genome editing in human cells. *ELife*, 2013(2). <https://doi.org/10.7554/ELIFE.00471>

- Jonas, S., & Izaurralde, E. (2015). Towards a molecular understanding of microRNA-mediated gene silencing. *Nature Reviews Genetics*, *16*(7), 421–433. <https://doi.org/10.1038/nrg3965>
- Karagkouni, D., Paraskevopoulou, M. D., Chatzopoulos, S., Vlachos, I. S., Tastsoglou, S., Kanellos, I., Papadimitriou, D., Kavakiotis, I., Maniou, S., Skoufos, G., Vergoulis, T., Dalamagas, T., & Hatzigeorgiou, A. G. (2017). DIANA-TarBase v8: a decade-long collection of experimentally supported miRNA-gene interactions. *Nucleic Acids Research*, *46*, 239–245. <https://doi.org/10.1093/nar/gkx1141>
- Kishore, S., Jaskiewicz, L., Burger, L., Hausser, J., Khorshid, M., & Zavolan, M. (2011). A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. *Nature Methods*, *8*(7), 559–564. <https://doi.org/10.1038/nmeth.1608>
- Landthaler, M., Gaidatzis, D., Rothballer, A., Chen, P. Y., Soll, S. J., Dinic, L., Ojo, T., Hafner, M., Zavolan, M., & Tuschl, T. (2008). Molecular characterization of human Argonaute-containing ribonucleoprotein complexes and their bound target mRNAs. *Rna*, *14*(12), 2580–2596. <https://doi.org/10.1261/rna.1351608>
- Lee, R. C., Feinbaum, R. L., & Ambros, V. (1993). The *C. elegans* Heterochronic Gene *lin-4* Encodes Small RNAs with Antisense Complementarity to *lin-14*. *Cell*, *75*, 843–854.
- Liang, X., Potter, J., Kumar, S., Ravinder, N., & Chesnut, J. D. (2017). Enhanced CRISPR/Cas9-mediated precise genome editing by improved design and delivery of gRNA, Cas9 nuclease, and donor DNA. *Journal of Biotechnology*, *241*, 136–146. <https://doi.org/10.1016/J.JBIOTECH.2016.11.011>
- Lin, B. B., Lei, H. Q., Xiong, H. Y., Fu, X., Shi, F., Yang, X. W., Yang, Y. F., Liao, G. L., Feng, Y. P., Jiang, D. G., & Pang, J. (2021). MicroRNA-regulated transcriptome analysis identifies four major subtypes with prognostic and therapeutic implications in prostate cancer. *Computational and Structural Biotechnology Journal*, *19*, 4941–4953. <https://doi.org/10.1016/J.CSB.2021.08.046>
- Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods*, *25*(4), 402–408. <https://doi.org/10.1006/meth.2001.1262>
- Martinez-Sanchez, A., & Murphy, C. L. (2013). MicroRNA Target Identification—Experimental Approaches. *Biology 2013, Vol. 2, Pages 189-205*, *2*(1), 189–205. <https://doi.org/10.3390/BIOLOGY2010189>
- McGeary, S. E., Bisaria, N., & Bartel, D. P. (2021). Pairing to the microRNA 3' region occurs through two alternative binding modes, with affinity shaped by nucleotide identity as well as pairing position. *BioRxiv*, 2021.04.13.439700. <https://doi.org/10.1101/2021.04.13.439700>
- McGeary, S. E., Bisaria, N., Pham, T. M., Wang, P. Y., & Bartel, D. P. (2022). MicroRNA 3'-compensatory pairing occurs through two binding modes, with affinity shaped by nucleotide identity and position. *ELife*, *11*. <https://doi.org/10.7554/elife.69803>
- McGeary, S. E., Lin, K. S., Shi, C. Y., Pham, T. M., Bisaria, N., Kelley, G. M., & Bartel, D. P. (2019). The biochemical basis of microRNA targeting efficacy. *Science*, *366*(6472), 1–20. <https://doi.org/10.1126/science.aav1741>
- McKinney Wes. (2010). Data Structures for Statistical Computing in Python. *Proceedings of the 9th Python in Science Conference (SCIPY 2010)*, *41*(3), 51–56.
- Moore, M. J., Scheel, T. K. H., Luna, J. M., Park, C. Y., Fak, J. J., Nishiuchi, E., Rice, C. M., & Darnell, R. B. (2015). miRNA–target chimeras reveal miRNA 3'-end pairing as a major determinant of Argonaute target specificity. *Nature Communications*, *6*(May), 8864. <https://doi.org/10.1038/ncomms9864>
- Moran, Y., Agron, M., Praher, D., & Technau, U. (2017). The evolutionary origin of plant and animal microRNAs. *Nature Ecology and Evolution*, *1*(3), 1–8. <https://doi.org/10.1038/s41559-016-0027>
- Mortazavi, S. S., Bahmanpour, Z., Daneshmandpour, Y., Roudbari, F., Sheervalilou, R., Kazeminasab, S., & Emamalizadeh, B. (2021). An updated overview and classification of bioinformatics tools for MicroRNA analysis, which one to choose? *Computers in Biology and Medicine*, *134*(January), 104544. <https://doi.org/10.1016/j.combiomed.2021.104544>

- Nagaraj, N., Wisniewski, J. R., Geiger, T., Cox, J., Kircher, M., Kelso, J., Pääbo, S., & Mann, M. (2011). Deep proteome and transcriptome mapping of a human cancer cell line. *Molecular Systems Biology*, 7(548), 1–8. <https://doi.org/10.1038/msb.2011.81>
- Nielsen, C. B., Shomron, N., Sandberg, R., Hornstein, E., Kitzman, J., & Burge, C. B. (2007). Determinants of targeting by endogenous and exogenous microRNAs and siRNAs. *RNA (New York, N.Y.)*, 13(11), 1894–1910. <https://doi.org/10.1261/RNA.768207>
- Prado, M. M., Frampton, A. E., Giovannetti, E., Stebbing, J., Castellano, L., Krell, J., Mato, M., Frampton, A. E., Giovannetti, E., & Stebbing, J. (2016). Expert Review of Molecular Diagnostics Investigating miRNA-mRNA regulatory networks using crosslinking immunoprecipitation methods for biomarker and target discovery in cancer. *Expert Review of Molecular Diagnostics*, 00(00), 1–8. <https://doi.org/10.1080/14737159.2016.1239532>
- Reinhart, B. J., Slack, F. J., & Basson, M. (2000). *The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans*. 901–906.
- Rüdel, S., Flatley, A., Weinmann, L., Kremmer, E., & Meister, G. (2008). A multifunctional human Argonaute2-specific monoclonal antibody. *RNA (New York, N.Y.)*, 14(6), 1244–1253. <https://doi.org/10.1261/rna.973808>
- Schnall-Levin, M., Zhao, Y., Perrimon, N., & Berger, B. (2010). Conserved microRNA targeting in Drosophila is as widespread in coding regions as in 3'UTRs. *Proceedings of the National Academy of Sciences of the United States of America*, 107(36), 15751–15756. <https://doi.org/10.1073/pnas.1006172107>
- Selbach, M., Schwanhäusser, B., Thierfelder, N., Fang, Z., Khanin, R., & Rajewsky, N. (2008). Widespread changes in protein synthesis induced by microRNAs. *Nature*, 455(7209), 58–63. <https://doi.org/10.1038/nature07228>
- Shabalina, S. A., & Koonin, E. V. (2008). *Origins and evolution of eukaryotic RNA interference*. August. <https://doi.org/10.1016/j.tree.2008.06.005>
- Sheu-Gruttadauria, J., Pawlica, P., Klum, S. M., Wang, S., Yario, T. A., Schirle Oakdale, N. T., Steitz, J. A., & MacRae, I. J. (2019). Structural Basis for Target-Directed MicroRNA Degradation. *Molecular Cell*, 75(6), 1243-1255.e7. <https://doi.org/10.1016/j.molcel.2019.06.019>
- Sheu-Gruttadauria, J., Xiao, Y., Gebert, L. F., & MacRae, I. J. (2019). Beyond the seed: structural basis for supplementary microRNA targeting by human Argonaute2. *The EMBO Journal*, e101153. <https://doi.org/10.15252/embj.2018101153>
- Shin, C., Nam, J.-W., Farh, K. K.-H., Chiang, H. R., Shkumatava, A., & Bartel, D. P. (2010). Expanding the MicroRNA Targeting Code: Functional Sites with Centered Pairing. *Molecular Cell*, 38(6), 789–802. <https://doi.org/10.1016/j.molcel.2010.06.005>
- Soares Medeiros, L. C., South, L., Peng, D., Bustamante, J. M., Wang, W., Bunkofsky, M., Perumal, N., Sanchez-Valdez, F., & Tarleton, R. L. (2017). Rapid, Selection-Free, High-Efficiency Genome Editing in Protozoan Parasites Using CRISPR-Cas9 Ribonucleoproteins. *MBio*, 8(6), e01788-17. <https://doi.org/10.1128/mBio.01788-17>
- Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C., & Doudna, J. A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*, 507(7490), 62–67. <https://doi.org/10.1038/nature13011>
- Sticht, C., De La Torre, C., Parveen, A., & Gretz, N. (2018). miRWalk: An online resource for prediction of microRNA binding sites. *PLOS ONE*, 13(10), e0206239. <https://doi.org/10.1371/JOURNAL.PONE.0206239>
- Szczyrba, J., Jung, V., Beitzinger, M., Nolte, E., Wach, S., Hart, M., Sapich, S., Wiesehöfer, M., Taubert, H., Wennemuth, G., Eichner, N., Stempf, T., Wullich, B., Meister, G., & Grässer, F. A. (2017). Analysis of Argonaute Complex Bound mRNAs in DU145 Prostate Carcinoma Supplementary. *Prostate Cancer*, 2017. <https://doi.org/10.1155/2017/4893921>

- Van Eijl, R. A. P. M., Van Den Brand, T., Nguyen, L. N., & Mulder, K. W. (2017). Reactivity of human AGO2 monoclonal antibody 11A9 with the SWI/SNF complex: A case study for rigorously defining antibody selectivity. *Scientific Reports*, 7(1), 1–11. <https://doi.org/10.1038/s41598-017-07539-4>
- Vazquez, N., Sanchez, L., Marks, R., Martinez, E., Fanniel, V., Lopez, A., Salinas, A., Flores, I., Hirschmann, J., Gilkerson, R., Schuenzel, E., Dearth, R., Halaby, R., Innis-Whitehouse, W., & Keniry, M. (2018). A protocol for custom CRISPR Cas9 donor vector construction to truncate genes in mammalian cells using pcDNA3 backbone. *BMC Molecular Biology*, 19(1), 3. <https://doi.org/10.1186/s12867-018-0105-8>
- Wang, B. (2013). Base composition characteristics of mammalian mirnas. *Journal of Nucleic Acids*, 2013. <https://doi.org/10.1155/2013/951570>
- Wang, L., Zhu, J., Deng, F. Y., Wu, L. F., Mo, X. B., Zhu, X. W., Xia, W., Xie, F. F., He, P., Bing, P. F., Qiu, Y. H., Lin, X., Lu, X., Zhang, L., Yi, N. J., Zhang, Y. H., & Lei, S. F. (2018). Correlation analyses revealed global microRNA–mRNA expression associations in human peripheral blood mononuclear cells. *Molecular Genetics and Genomics*, 293(1), 95–105. <https://doi.org/10.1007/s00438-017-1367-4>
- Wang, T. T., Lee, C. Y., Lai, L. C., Tsai, M. H., Lu, T. P., & Chuang, E. Y. (2019). AnamiR: Integrated analysis of MicroRNA and gene expression profiling. *BMC Bioinformatics*, 20(1), 1–7. <https://doi.org/10.1186/S12859-019-2870-X/TABLES/6>
- Wei, J., Yin, Y., Deng, Q., Zhou, J., Wang, Y., Yin, G., Yang, J., & Tang, Y. (2020). Integrative Analysis of MicroRNA and Gene Interactions for Revealing Candidate Signatures in Prostate Cancer. *Frontiers in Genetics*, 11, 176. <https://doi.org/10.3389/FGENE.2020.00176/BIBTEX>
- Xiao, Y., & Macrae, I. J. (2020). Robust differential microRNA targeting driven by supplementary interactions in vitro. *RNA*, 26(2), 162–174. <https://doi.org/10.1261/rna.072264.119>
- Yang, Y., Jia, B., Zhao, X., Wang, Y., & Ye, W. (2019). miR-93-5p may be an important oncogene in prostate cancer by bioinformatics analysis. *Journal of Cellular Biochemistry*, 120(6), 10463–10483. <https://doi.org/10.1002/jcb.28332>
- Ye, Y., Li, S. L., & Wang, S. Y. (2018). Construction and analysis of mRNA, miRNA, lncRNA, and TF regulatory networks reveal the key genes associated with prostate cancer. *PLOS ONE*, 13(8), e0198055. <https://doi.org/10.1371/JOURNAL.PONE.0198055>
- Yue, D., Liu, H., & Huang, Y. (2009). Survey of Computational Algorithms for MicroRNA Target Prediction. *Current Genomics*, 10(7), 478–492. <https://doi.org/10.2174/138920209789208219>
- Zhang, K., Zhang, X., Cai, Z., Zhou, J., Cao, R., Zhao, Y., Chen, Z., Wang, D., Ruan, W., Zhao, Q., Liu, G., Xue, Y., Qin, Y., Zhou, B., Wu, L., Nilsen, T., Zhou, Y., & Fu, X. D. (2018). A novel class of microRNA-recognition elements that function only within open reading frames. *Nature Structural and Molecular Biology*, 25(11), 1019–1027. <https://doi.org/10.1038/s41594-018-0136-3>

Material suplementario

Tabla suplementaria 1

miR family	MiRBase ID	MiRBase Accession
let-7-5p/98-5p	hsa-let-7a-5p	MIMAT0000062
let-7-5p/98-5p	hsa-let-7b-5p	MIMAT0000063
let-7-5p/98-5p	hsa-let-7c-5p	MIMAT0000064
let-7-5p/98-5p	hsa-let-7d-5p	MIMAT0000065
let-7-5p/98-5p	hsa-let-7e-5p	MIMAT0000066
let-7-5p/98-5p	hsa-let-7f-5p	MIMAT0000067
let-7-5p/98-5p	hsa-let-7g-5p	MIMAT0000414
let-7-5p/98-5p	hsa-let-7i-5p	MIMAT0000415
let-7-5p/98-5p	hsa-miR-98-5p	MIMAT0000096
miR-101-3p.1	hsa-miR-101-3p.1	MIMAT0000099
miR-103-3p/107	hsa-miR-103a-3p	MIMAT0000101
miR-103-3p/107	hsa-miR-107	MIMAT0000104
miR-10-5p	hsa-miR-10a-5p	MIMAT0000253
miR-10-5p	hsa-miR-10b-5p	MIMAT0000254
miR-125-5p	hsa-miR-125a-5p	MIMAT0000443
miR-125-5p	hsa-miR-125b-5p	MIMAT0000423
miR-126-3p.1	hsa-miR-126-3p.1	MIMAT0000445
miR-128-3p	hsa-miR-128-3p	MIMAT0000424
miR-129-5p	hsa-miR-129-5p	MIMAT0000242
miR-130-3p/301-3p/454-3p	hsa-miR-130a-3p	MIMAT0000425
miR-130-3p/301-3p/454-3p	hsa-miR-130b-3p	MIMAT0000691
miR-130-3p/301-3p/454-3p	hsa-miR-301a-3p	MIMAT0000688
miR-130-3p/301-3p/454-3p	hsa-miR-454-3p	MIMAT0003885
miR-132-3p/212-3p	hsa-miR-132-3p	MIMAT0000426
miR-132-3p/212-3p	hsa-miR-212-3p	MIMAT0000269
miR-133a-3p.2/133b	hsa-miR-133a-3p.2	MIMAT0000427
miR-133a-3p.2/133b	hsa-miR-133b	MIMAT0000770
miR-135-5p	hsa-miR-135a-5p	MIMAT0000428
miR-135-5p	hsa-miR-135b-5p	MIMAT0000758
miR-139-5p	hsa-miR-139-5p	MIMAT0000250
miR-1-3p/206	hsa-miR-1-3p	MIMAT0000416
miR-140-3p.2	hsa-miR-140-3p.2	MIMAT0004597
miR-140-5p	hsa-miR-140-5p	MIMAT0000431
miR-141-3p/200a-3p	hsa-miR-141-3p	MIMAT0000432
miR-141-3p/200a-3p	hsa-miR-200a-3p	MIMAT0000682
miR-142-3p.1	hsa-miR-142-3p.1	MIMAT0000434
miR-142-5p	hsa-miR-142-5p	MIMAT0000433
miR-143-3p	hsa-miR-143-3p	MIMAT0000435
miR-144-3p	hsa-miR-144-3p	MIMAT0000436

miR-145-5p	hsa-miR-145-5p	MIMAT0000437
miR-146-5p	hsa-miR-146a-5p	MIMAT0000449
miR-146-5p	hsa-miR-146b-5p	MIMAT0002809
miR-148-3p/152-3p	hsa-miR-148a-3p	MIMAT0000243
miR-148-3p/152-3p	hsa-miR-148b-3p	MIMAT0000759
miR-148-3p/152-3p	hsa-miR-152-3p	MIMAT0000438
miR-150-5p	hsa-miR-150-5p	MIMAT0000451
miR-153-3p	hsa-miR-153-3p	MIMAT0000439
miR-155-5p	hsa-miR-155-5p	MIMAT0000646
miR-15-5p/16-5p/195-5p/424-5p/497-5p	hsa-miR-15a-5p	MIMAT0000068
miR-15-5p/16-5p/195-5p/424-5p/497-5p	hsa-miR-15b-5p	MIMAT0000417
miR-15-5p/16-5p/195-5p/424-5p/497-5p	hsa-miR-16-5p	MIMAT0000069
miR-15-5p/16-5p/195-5p/424-5p/497-5p	hsa-miR-195-5p	MIMAT0000461
miR-15-5p/16-5p/195-5p/424-5p/497-5p	hsa-miR-424-5p	MIMAT0001341
miR-15-5p/16-5p/195-5p/424-5p/497-5p	hsa-miR-497-5p	MIMAT0002820
miR-17-5p/20-5p/93-5p/106-5p/519-3p	hsa-miR-106a-5p	MIMAT0000103
miR-17-5p/20-5p/93-5p/106-5p/519-3p	hsa-miR-106b-5p	MIMAT0000680
miR-17-5p/20-5p/93-5p/106-5p/519-3p	hsa-miR-17-5p	MIMAT0000070
miR-17-5p/20-5p/93-5p/106-5p/519-3p	hsa-miR-20a-5p	MIMAT0000075
miR-17-5p/20-5p/93-5p/106-5p/519-3p	hsa-miR-20b-5p	MIMAT0001413
miR-17-5p/20-5p/93-5p/106-5p/519-3p	hsa-miR-93-5p	MIMAT0000093
miR-181-5p	hsa-miR-181a-5p	MIMAT0000256
miR-181-5p	hsa-miR-181b-5p	MIMAT0000257
miR-181-5p	hsa-miR-181c-5p	MIMAT0000258
miR-181-5p	hsa-miR-181d-5p	MIMAT0002821
miR-182-5p	hsa-miR-182-5p	MIMAT0000259
miR-183-5p.1	hsa-miR-183-5p.1	MIMAT0000261
miR-184	hsa-miR-184	MIMAT0000454
miR-18-5p	hsa-miR-18a-5p	MIMAT0000072
miR-18-5p	hsa-miR-18b-5p	MIMAT0001412
miR-187-3p	hsa-miR-187-3p	MIMAT0000262
miR-190-5p	hsa-miR-190a-5p	MIMAT0000458
miR-191-5p	hsa-miR-191-5p	MIMAT0000440
miR-192-5p/215-5p	hsa-miR-192-5p	MIMAT0000222
miR-192-5p/215-5p	hsa-miR-215-5p	MIMAT0000272
miR-193-3p	hsa-miR-193a-3p	MIMAT0000459
miR-193-3p	hsa-miR-193b-3p	MIMAT0002819
miR-193a-5p	hsa-miR-193a-5p	MIMAT0004614
miR-19-3p	hsa-miR-19a-3p	MIMAT0000073
miR-19-3p	hsa-miR-19b-3p	MIMAT0000074
miR-194-5p	hsa-miR-194-5p	MIMAT0000460
miR-196-5p	hsa-miR-196a-5p	MIMAT0000226
miR-196-5p	hsa-miR-196b-5p	MIMAT0001080
miR-199-3p	hsa-miR-199a-3p	MIMAT0000232
miR-199-3p	hsa-miR-199b-3p	MIMAT0004563
miR-199-5p	hsa-miR-199a-5p	MIMAT0000231

miR-199-5p	hsa-miR-199b-5p	MIMAT0000263
miR-200bc-3p/429	hsa-miR-200b-3p	MIMAT0000318
miR-200bc-3p/429	hsa-miR-200c-3p	MIMAT0000617
miR-200bc-3p/429	hsa-miR-429	MIMAT0001536
miR-204-5p/211-5p	hsa-miR-204-5p	MIMAT0000265
miR-205-5p	hsa-miR-205-5p	MIMAT0000266
miR-210-3p	hsa-miR-210-3p	MIMAT0000267
miR-214-5p	hsa-miR-214-5p	MIMAT0004564
miR-21-5p/590-5p	hsa-miR-21-5p	MIMAT0000076
miR-21-5p/590-5p	hsa-miR-590-5p	MIMAT0003258
miR-217	hsa-miR-217	MIMAT0000274
miR-218-5p	hsa-miR-218-5p	MIMAT0000275
miR-219-5p	hsa-miR-219a-5p	MIMAT0000276
miR-221-3p/222-3p	hsa-miR-221-3p	MIMAT0000278
miR-221-3p/222-3p	hsa-miR-222-3p	MIMAT0000279
miR-223-3p	hsa-miR-223-3p	MIMAT0000280
miR-22-3p	hsa-miR-22-3p	MIMAT0000077
miR-23-3p	hsa-miR-23a-3p	MIMAT0000078
miR-23-3p	hsa-miR-23b-3p	MIMAT0000418
miR-23-3p	hsa-miR-23c	MIMAT0018000
miR-24-3p	hsa-miR-24-3p	MIMAT0000080
miR-25-3p/32-5p/92-3p/363-3p/367-3p	hsa-miR-25-3p	MIMAT0000081
miR-25-3p/32-5p/92-3p/363-3p/367-3p	hsa-miR-32-5p	MIMAT0000090
miR-25-3p/32-5p/92-3p/363-3p/367-3p	hsa-miR-363-3p	MIMAT0000707
miR-25-3p/32-5p/92-3p/363-3p/367-3p	hsa-miR-92a-3p	MIMAT0000092
miR-25-3p/32-5p/92-3p/363-3p/367-3p	hsa-miR-92b-3p	MIMAT0003218
miR-26-5p	hsa-miR-26a-5p	MIMAT0000082
miR-26-5p	hsa-miR-26b-5p	MIMAT0000083
miR-27-3p	hsa-miR-27a-3p	MIMAT0000084
miR-27-3p	hsa-miR-27b-3p	MIMAT0000419
miR-29-3p	hsa-miR-29a-3p	MIMAT0000086
miR-29-3p	hsa-miR-29b-3p	MIMAT0000100
miR-29-3p	hsa-miR-29c-3p	MIMAT0000681
miR-30-5p	hsa-miR-30a-5p	MIMAT0000087
miR-30-5p	hsa-miR-30b-5p	MIMAT0000420
miR-30-5p	hsa-miR-30c-5p	MIMAT0000244
miR-30-5p	hsa-miR-30d-5p	MIMAT0000245
miR-30-5p	hsa-miR-30e-5p	MIMAT0000692
miR-31-5p	hsa-miR-31-5p	MIMAT0000089
miR-33-5p	hsa-miR-33a-5p	MIMAT0000091
miR-33-5p	hsa-miR-33b-5p	MIMAT0003301
miR-338-3p	hsa-miR-338-3p	MIMAT0000763
miR-34-5p/449-5p	hsa-miR-34a-5p	MIMAT0000255
miR-34-5p/449-5p	hsa-miR-34c-5p	MIMAT0000686
miR-365-3p	hsa-miR-365a-3p	MIMAT0000710
miR-375	hsa-miR-375	MIMAT0000728

miR-383-5p.1	hsa-miR-383-5p.1	MIMAT0000738
miR-425-5p	hsa-miR-425-5p	MIMAT0003393
miR-451	hsa-miR-451a	MIMAT0001631
miR-455-3p.1	hsa-miR-455-3p.1	MIMAT0004784
miR-455-5p	hsa-miR-455-5p	MIMAT0003150
miR-7-5p	hsa-miR-7-5p	MIMAT0000252
miR-9-5p	hsa-miR-9-5p	MIMAT0000441
miR-96-5p/1271-5p	hsa-miR-1271-5p	MIMAT0005796
miR-96-5p/1271-5p	hsa-miR-96-5p	MIMAT0000095
miR-99-5p/100-5p	hsa-miR-100-5p	MIMAT0000098
miR-99-5p/100-5p	hsa-miR-99a-5p	MIMAT0000097
miR-99-5p/100-5p	hsa-miR-99b-5p	MIMAT0000689

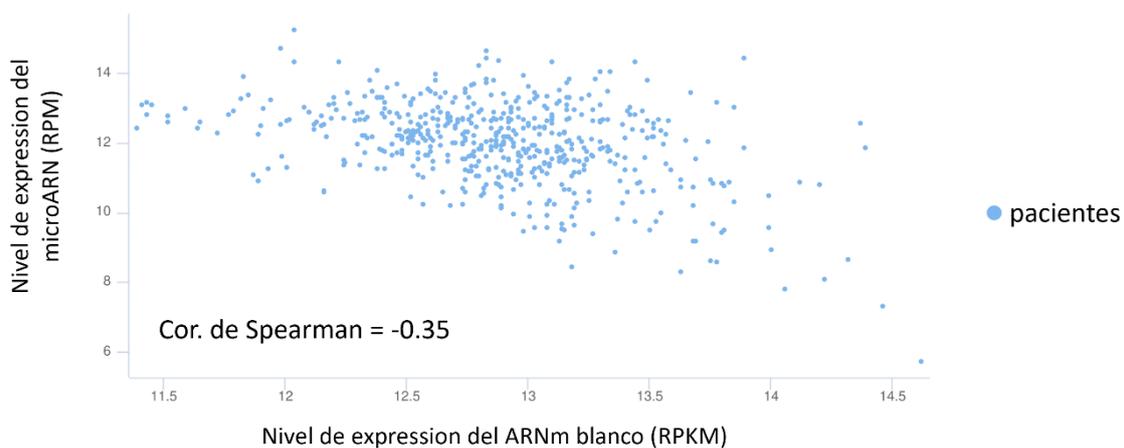


Figura suplementaria 1. Representación del análisis de correlación de expresión entre microARN y un ARNm blanco.

