



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

FACULTAD DE CIENCIAS

**ANÁLISIS BIOINFORMÁTICO DE
MICRORNAS PARA ENCONTRAR LAS
POSIBLES CAUSAS MOLECULARES QUE DAN
ORIGEN A LOS ISOMIRS**

T E S I S

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

P R E S E N T A:

OSCAR OMAR MORENO PROO

**DIRECTOR DE TESIS:
DR. CARLOS FABIAN FLORES JASSO**

Ciudad Universitaria, CD.Mx. 2019





Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

Hoja de datos del jurado

1. Datos del alumno

Moreno

Proo

Oscar Omar

5525231328

Universidad Nacional Autónoma de México

Facultad de Ciencias

Biología

311031885

2. Datos del tutor

Dr.

Carlos Fabián

Flores

Jasso

3. Datos del sinodal 1

Dr.

Luis Alfonso

Vaca

Domínguez

4. Datos del sinodal 2

Dr.

Arturo Carlos II

Becerra

Bracho

5. Datos del sinodal 3

Dra.

Claudia Andrea

Segal

Kischinevzky

6. Datos del sinodal 4

Dr.

Felipe

Alcántara Sánchez

7. Datos de la tesis

Análisis bioinformático de miRNAs para encontrar las posibles causas moleculares que dan origen a los isomiRNAs

60 p.

2019

Agradecimientos

Este trabajo se realizó en el laboratorio del Consorcio de Metabolismo de RNA y Vesículas Extracelulares de Instituto Nacional de Medicina Genómica (INMEGEN) México.

Agradezco al Dr. Fabián Flores y a la Dra. Eréndira Avendaño por abrirme las puertas de su laboratorio y de darme los conocimientos, observaciones y las herramientas necesarias para desarrollarme en el camino de la buena ciencia.

A mis sinodales por dedicar parte de su tiempo a la revisión de este trabajo:

Dra. Claudia Segal Kischinevzky.

Dr. Arturo Carlos II Becerra Bracho.

Dr. Luis Vaca Domínguez.

Dr. Felipe Alcántara Sánchez.

A la Dra. Mariana Flores, parte fundamental de mi desarrollo, gracias por tanta paciencia, dedicación, apoyo y conocimientos compartidos.

A los miembros y ex - miembros del CRVE Diana, Fili, Viri, Vic, Perlita y Mario. Que aunque algunos de nuestros caminos sean diferentes, siempre serán parte importante de mi formación y de mi vida

A Joshua e Israel, por ayudarme a comprender mucho de lo que ahora sé.

Dedicatoria

A Norberta y a Leonor, que aunque ya no estén presentes, siempre estaré agradecido por ese fragmento en la vida en el que pudimos coincidir y en el que siempre pude disfrutar de su cariño.

A mis padres Oscar y Bertha, que siempre han estado ahí para apoyarme en cualquier situación. Sé que siempre contaré con su apoyo y su amor incondicional y de igual manera estaré para ustedes siempre.

A Marisol y Lourdes, que siempre han sido participes de mi desarrollo y siempre han buscado la manera de apoyarme. A mi familia, que siempre ha sido muy unida en los buenos y malos momentos.

A mis hermanos: Hector, Fernanda, Daniela y Ximena. Que aunque hemos crecido juntos, siempre tenemos algo nuevo que contarnos.

A Brenda, por todo el amor, tiempo y el apoyo que hemos compartido a lo largo de nuestro camino. Todo es mejor si estamos juntos.

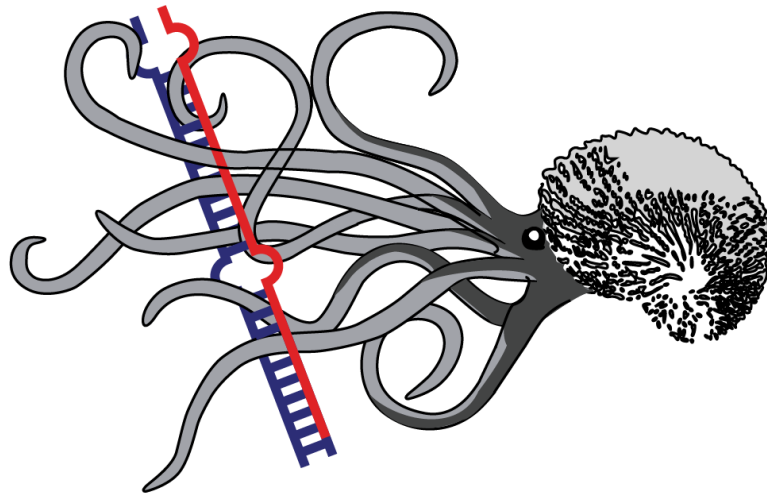
A mis amigos de toda la vida Josué, Luis y Elsy. Gracias por siempre brindarme su amistad y compartir hasta la fecha muchas experiencias.

A mis amigos de la universidad, César, Daniel, Zyanya y Ricardo. Por enseñarme que aunque exista una amplia variedad de puntos de vista, la amistad siempre estará por encima de cualquier adversidad.

A mis compañeros y amigos de la Facultad de Ciencias de la UNAM.

“...mientras este planeta ha ido girando según la ley constante de la gravitación; a partir de un comienzo tan simple, se desarrollaron y están evolucionando infinitas formas, cada vez más bellas y maravillosas.”

— Charles Darwin



Índice

Abreviaturas.....	1
Índice de figuras.....	2
Resumen	3
Introducción	5
¿Qué son y cómo funcionan los microRNAs?.....	5
Biogénesis de microRNAs	6
El complejo RISC.....	9
Los microRNAs y sus RNAs blanco	10
Biogénesis alternativas de microRNAs	12
La base de datos de microRNAs: miRBase	13
Antecedentes	15
Tipos de isomiRs	16
Generación de isomiRs	17
El análisis bioinformático de los microRNAs y su influencia en el estudio de los isomiRs	19
Objetivo general	24
Objetivos específicos	24
Métodos	25
Selección de las bibliotecas públicas.....	25
Pre-procesamiento de las bibliotecas.....	26
Control de calidad del llamado de base	26
Corte de adaptadores y eliminación de secuencias de baja calidad.....	26
Anotación de microRNAs.....	27
Normalización de las lecturas totales	27
Análisis estadístico.....	27
Ordenamiento de secuencias para calcular el isomiR-index	29
Resultados	30
La motivación de la implementación del concepto de <i>MAS</i> , y el propósito de la herramienta <i>isomiR-index</i>	30

El isomiR-index es útil para comparar el contenido de isomiRs entre diferentes muestras	33
La abundancia de microRNAs dentro de una muestra es independiente de su porcentaje de MAS.....	36
Las proporciones del MAS e isomiRs son diferentes entre tejidos de diferentes especies	38
La proporción global de isomiRs se incrementa en células carentes de Dicer, pero no en Drosha o XPO 5	39
Conclusiones	47
Referencias.....	48

Abreviaturas

Ago	Argonauta
ENA	European Nucleotide Archive
GEO	Gene Expression Omnibus
isomiR	Secuencia variante de microRNA
lncRNA	RNA largo no codificante
MAS	Secuencia más abundante
mRNA	RNA mensajero
nt	Nucleótidos
ORF	Marco abierto de lectura
PARN	Ribonucleasa específica de poli A
pre-miRNA	Precursor de microRNA
pri-miRNA	microRNA primario
RNA	Ácido ribonucleico
RNA pol II	RNA polimerasa II
RPM	Lecturas por millón
RPmiR	Lecturas por microRNA
RISC	Complejo de silenciamiento inducido por RNA
RT-PCR	Reacción en cadena de la polimerasa con retro transcriptasa
sRNA-seq	Secuenciación masiva de RNAs pequeños
TUTasa	Uridil transferasa terminal
UTR	Región no traducible
XPO5	Exportina 5

Índice de figuras

Figura 1	Biogénesis clásica de los microRNAs	8
Figura 2	Unión de RISC al mRNA	10
Figura 3	Silenciamiento de mRNAs mediado por microRNAs	12
Figura 4	Biogénesis alternativa de los microRNAs	14
Figura 5	Tipos de isomiRs	17
Figura 6	Cortes imprecisos de Dicer y Drosha	19
Figura 7	Análisis del efecto de inflamación en isomiRs de miR-223-3p	22
Figura 8	Flujo de trabajo para el análisis de las bibliotecas descargadas	28
Figura 9	Ejemplo del ordenamiento de las bases de datos por isomiR-index	29
Figura 10	Captura de pantalla de miRBase sobre Hsa-let-7a-2	31
Figura 11	Ejemplo de la selección del MAS y la fracción de isomiRs	32
Figura 12	Cálculo del porcentaje del MAS y la fracción de isomiRs	32
Figura 13	Porcentaje de MAS por microRNA en muestras de tejido renal	35
Figura 14	Correlación entre el porcentaje de MAS por microRNA en muestras de tejido renal	36
Figura 15	Comparación entre la correlación del porcentaje de MAS y la correlación de abundancia	37
Figura 16	Correlación de tejido neural de humano <i>vs</i> tejido hepático de ratón	39
Figura 17	Porcentaje de MAS por microRNA de HCT116 WT y Dicer KO	41
Figura 18	Correlación entre muestras de HCT116 WT <i>vs</i> Dicer KO	42
Figura 19	Correlación entre HCT116 WT <i>vs</i> Drosha KO y XPO 5 KO	43

Resumen

Los microRNAs son RNAs no codificantes de aproximadamente 22 nucleótidos de largo (en su forma madura) que regulan la expresión génica a nivel postranscripcional. Con el auge de la secuenciación masiva de microRNAs se identificaron isoformas que recibieron el nombre de isomiRs. Los isomiRs contienen variaciones con respecto a la secuencia que se tiene de referencia en miRBase para cada microRNA.

Actualmente se tiene registro de que algunos isomiRs pueden ser más abundantes que las muestras de referencia de miRBase y que también es posible que los resultados obtenidos de estudios que solamente se enfocan en el reporte de la secuencia de referencia de miRBase del microRNA estudiado, estén malinterpretando los resultados que se obtienen al realizar, por ejemplo, ensayos de RT-PCR en tiempo real. Esto es debido a que las sondas son diseñadas para la identificación de la secuencia de referencia de miRBase y es posible que esta sonda no pueda detectar a los isomiRs más abundantes

El estudio actual de isomiRs considera a cada secuencia como una entidad independiente y únicamente se encarga de reportar la abundancia aislada de cada secuencia. Aunque existen diversas herramientas informáticas que están diseñadas para identificar y cuantificar isomiRs, no existe ninguna que permita determinar el panorama completo de isomiRs de cada microRNA cuantitativamente y que además permita comparar los resultados obtenidos con los de otras muestras. Por esto, es necesario el desarrollo de una manera de poder estudiar a los microRNAs en conjunto y simultáneamente a cada uno de sus isomiRs.

En este estudio planteamos una nueva herramienta, isomiR-index, para analizar cuantitativamente el panorama completo de isomiRs de distintas muestras. Acuñamos el término MAS para denominar a la secuencia más abundante de cada microRNA de cada base de datos y analizar el contenido porcentual de isomiRs con respecto a la MAS.

Nuestros resultados muestran que al comparar el contenido porcentual de MAS en bases de datos de tejido renal humano, existe correlación de $r = 0.9$. Mientras que tejidos distintos de humano y ratón tienen una correlación de $r = 0.08$.

Al comparar muestras provenientes de células HCT116 knock-out de distintas proteínas que participan en la biogénesis de los microRNAs como Drosha, XPO5 y Dicer encontramos que la ausencia de la RNasa Dicer tiene un impacto en la

producción de isomiRs; la ausencia de XPO 5, al no ser esta una RNAsa, no tiene un impacto apreciable en la producción de isomiRs al comparar con la línea parental. Un resultado inesperado es que la línea carente de Drosha tampoco tiene un impacto aparente en la producción de isomiRs.

Dada la utilidad del isomiR-index como herramienta para el estudio y clasificación de cambios en las proporciones de isomiRs mostradas en este trabajo, es posible que en estudios futuros el contenido de isomiRs pueda arrojar hallazgos importantes en el campo de investigación sobre la función y los mecanismos de acción de los microRNAs.

Introducción

En 1993 se hizo un descubrimiento cuya importancia pasó desapercibida por la mayoría de los grupos de investigación dedicados a la comprensión de la regulación de la expresión génica. Dicho descubrimiento mostraba que, en el nemátodo *C. elegans*, la traducción del mRNA de un gen conocido como *lin-14*, estaba regulada por la presencia de otro RNA, *lin-4*, cuya secuencia era parcialmente complementaria a su región 3' no traducible (Lee *et al.* 1993). Lo más peculiar del hallazgo, razón que podría explicar por qué pasó tan desapercibido, fue que el RNA regulador era muy pequeño, 21 nt para ser precisos y no se tenía conocimiento de tal tipo de regulación en otros organismos; lo que hizo parecer al hallazgo una mera curiosidad. Ocho años después, gracias al esfuerzo de tres grupos de investigación, se descubrieron una gran cantidad de RNAs, de tamaño similar al RNA *lin-4*, que regulan la expresión de mRNA de manera similar en *C. elegans*, *A. thaliana*, *D. melanogaster* y *H. sapiens* (Lagos-Quintana *et al.* 2001; Lau *et al.* 2001; Lee *et al.* 2001). El descubrimiento fue reportado de manera simultánea por los tres grupos de trabajo en la revista *Science* en el año 2001 y a estos RNAs pequeños se les conoce desde entonces como microRNAs.

A más de 25 años del descubrimiento del primer microRNA, actualmente sabemos que la forma en que los microRNAs regulan la expresión génica está altamente conservada; que alteraciones en sus secuencias, o en las proteínas responsables de su biogénesis, dan lugar a defectos en el metabolismo; y que su ausencia, o sobre-expresión, correlaciona, o es causante, de muchas enfermedades. La avalancha de publicaciones que han revelado los mecanismos y función de los microRNAs, ha ido de la mano de la generación de nuevas tecnologías que han permitido estudiarlos individualmente o en su conjunto, bajo diversas condiciones experimentales. Esto ha dado como resultado conocimiento que es empleado en la clínica actualmente, ya que la detección de la presencia de uno, o múltiples microRNAs puedan correlacionarse con la aparición de diversas enfermedades en humanos (Janssen *et al.* 2013; Lares *et al.* 2010; van Rooij *et al.* 2012).

¿Qué son y cómo funcionan los microRNAs?

Los microRNAs son RNAs cuyo tamaño oscila entre los 21–23 nucleótidos (nt) en su forma madura que regulan la expresión génica en varios niveles. Dependiendo del tipo de especie del que se trate, los microRNAs pueden funcionar en el citoplasma, inhibiendo la traducción de mRNAs o, por el contrario, pueden actuar en el núcleo de la célula, regulando la transcripción (Leung A. 2015; Roberts T. 2014).

En animales, el mecanismo más estudiado por el que los microRNAs actúan es hibridándose parcialmente a los mRNAs en sus regiones 3'-UTR, dando como resultado la atenuación de la traducción por mecanismos no del todo entendidos (Lewis *et al.* 2005). En plantas, el mecanismo de función de los microRNAs principal es hibridándose perfectamente a sus mRNAs blancos e impidiendo su traducción por medio de un corte endonucleolítico (Schwab *et al.* 2005).

Actualmente se sabe que existen microRNAs en las especies en las que se han detectado. En humanos, por ejemplo, la cantidad de microRNAs encontrados sobrepasa las 2000 secuencias distintas; algunos de ellos están conservados, aunque algunos otros han sido postulados como microRNAs “especie-específica” ya que sus secuencias, o la de sus moléculas precursoras, no se han encontrado en otros primates. El que existan tantos microRNAs distintos presenta un reto importante para el estudio de su función en animales: dado que estos se hibridan parcialmente a sus mRNAs blanco, ante la vasta cantidad de RNAs con secuencias distintas, teóricamente un microRNA podría regular a muchos mRNAs. Se estima, por ejemplo, que cada microRNA puede regular a cientos de blancos, y que cada uno de estos, en teoría podría ser regulado por más de un microRNA; lo cual ha llevado al cuestionamiento sobre qué es lo que convierte a un mRNA en un “blanco verdadero” a un nivel de regulación muy complejo.

Biogénesis de microRNAs

La biogénesis de los microRNAs comienza en el núcleo con la transcripción del RNA por medio de la RNA polimerasa tipo II (RNA pol II). El RNA recién transcrito que da lugar al microRNA conocido como microRNA *primario* (pri-microRNA). El pri-miRNA puede tener más de 1,000 nt, tiene un cap en el extremo 5' y cola de poli-A en el extremo 3', tal y como si fuera un mRNA (Lee *et al.* 2002). Cerca del 40% de las secuencias de microRNAs caen dentro de intrones o exones; por ello, su transcripción usualmente es regulada junto con la de los genes que los albergan. Los microRNAs se encuentran presentes en casi todos los organismos eucariontes y se derivan de distintos RNAs más largos, como intrones o de RNAs largos no codificantes (lncRNAs) (Cai *et al.* 2004; Lee *et al.* 2007). También provienen de elementos como exones o incluso pueden ser transcritos como clústeres, los cuales incluyen varias moléculas precursoras para distintos microRNAs (O'Brien *et al.* 2018). Sin embargo la biogénesis y el mecanismo de acción puede variar entre los grupos evolutivos más grandes; por ejemplo, se conoce que la producción de microRNAs maduros en plantas involucra el procesamiento de por lo menos tres moléculas precursoras del microRNA

(Kurihara *et al.* 2004), mientras que en animales únicamente existen dos moléculas precursoras.

Una característica primordial de los pri-microRNAs es que la secuencia que contenga al microRNA forma una estructura de tallo y asa. De tal manera que esta estructura sea reconocida posteriormente por un complejo multi-protéico, compuesto principalmente por las proteínas DGCR8 y Drosha, que toman como sustrato al pri-microRNA para procesarlo a un precursor de menor tamaño. El procesamiento de los pri-microRNAs en el núcleo se conoce con gran detalle. Por ejemplo, DGCR8 tiene múltiples dominios que ayudan en el reconocimiento del pri-miRNA como sustrato del complejo microprocesador. El dominio C-terminal de DGCR8 proporciona la estabilidad y la unión con Drosha (Nguyen *et al.* 2015), el dominio de unión a RNA de doble hebra interactúa con el pri-miRNA para incrementar la actividad enzimática del microprocesador (Quick-Cleveland *et al.* 2014) y el dominio Rhed asegura la precisión del procesamiento y mejora la eficiencia (Roth *et al.* 2013). Drosha es una RNasa de tipo III que funciona como un centro procesador compuesto por dos dominios de RNasa III: RIIIDa y RIIIDb (Kwon *et al.* 2016). El dominio RIIIDa corta el extremo 3' del tallo del pri-miRNA mientras que RIIIDb corta el extremo 5' del mismo. El resultado de este procesamiento es un RNA de aproximadamente 60–75 nt denominado precursor del microRNA (pre-miRNA), mismo que incluye dos nt sobresalientes en el extremo 3', lo cual es una característica del corte de una RNasa tipo III (Han *et al.* 2004) (**Figura 1**).

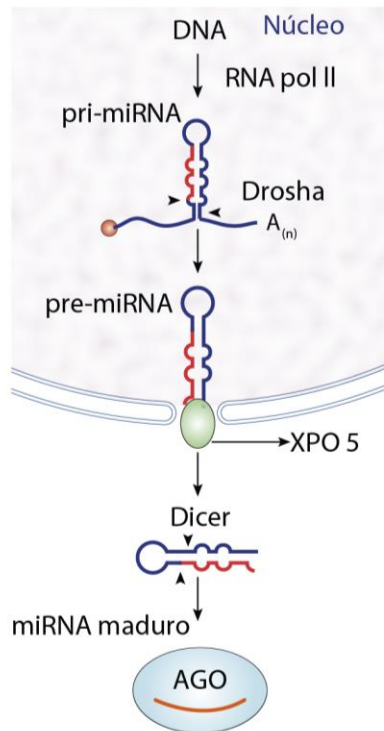


Figura 1. Biogénesis clásica de los microRNAs en mamíferos. Se ilustran las enzimas más importantes para el proceso de biogénesis clásica de los microRNAs. Los triángulos denotan el sitio de corte de las RNAsas tipo III.

El pre-miRNA se exporta del núcleo al citoplasma por medio de un mecanismo de transporte activo mediado por la exportina 5 (XPO5) y dependiente de la hidrólisis de GTP mediado por la GTPasa Ran (Yi *et al.* 2003). Una vez en el citoplasma, el pre-miRNA se convierte en sustrato de Dicer (Macrae *et al.* 2006).

Dicer es otra RNAsa tipo III que actúa en asociación con otra proteína citoplasmática de unión a RNA, TRBP, para reconocer al pre-miRNA y cortarlo. El dominio *Platform PAZ Connector* (PPC) de Dicer contiene los sitios de reconocimiento para el extremo 5' y 3' del microRNA. El reconocimiento de los extremos del pre-miRNA hace posible que su centro de procesamiento corte el tallo del pre-miRNA a partir de los nt sobresalientes del extremo 3' y desde el fosfato del extremo 5' (Park *et al.* 2011). Se sabe que el primer paso del reconocimiento de Dicer ocurre en el extremo 3' del pre-miRNA y que los nt que se encuentran sobresalientes influyen en la selección y precisión del corte de Dicer (Flores-Jasso *et al.* 2009).

El producto del corte de Dicer origina un dúplex de RNA de aproximadamente 22 nt. En el último paso de la biogénesis, una de las dos hebras del dúplex resultante, la hebra guía, es anclada dentro de alguna proteína de la familia Argonauta (Ago) para funcionar como microRNA maduro, mientras que la otra hebra, la hebra pasajera, se disocia y posteriormente es degradada (**Figura 1**) (Hutvágner *et al.* 2001).

El complejo RISC

Es común encontrar en la literatura que los microRNAs son muy importantes para el metabolismo de una célula, tejido u organismo. Sin embargo, por sí solos no tienen una función celular, ya que son rápidamente degradados por RNAsas celulares (Aryani *et al.* 2015). Por ello, es importante puntualizar que los microRNAs participan en la regulación de la expresión génica y primordialmente funcionan como *guías* de las proteínas Ago a las que se encuentran unidos. De tal manera que el silenciamiento de mRNAs previo a su traducción es iniciado por la presencia de alguna de las proteínas Ago, unida al extremo 3'-UTR del mRNA. Estudios pioneros del campo reportaron que la presencia de proteína efectora del silenciamiento Ago2 es suficiente para lograr un silenciamiento, independientemente de su unión al microRNA (Pillai *et al.* 2004).

Al complejo de microRNA guía que permanece en asociación con Ago, se le conoce como complejo RISC (del inglés, *RNA Induced Silencing Complex*). Las proteínas Argonauta están constituidas por cuatro dominios estructurales que cumplen ciertas funciones en la asociación con los microRNAs. Los dominios MID y PIWI forman un sitio de anclaje en donde los aminoácidos pueden interactuar con el esqueleto de azúcar y fosfato del microRNA, de tal manera que el extremo 5' de la hebra guía del microRNA queda anclado en este sitio (Höck *et al.* 2008). En animales el posicionamiento de la hebra guía en el sitio de unión del dominio MID y el dominio PIWI provoca un acomodo de forma helicoidal en los nucleótidos que componen la región *semilla* (posiciones 2-8 del microRNA guía), haciendo posible la interacción con el mRNA blanco (Swarts *et al.* 2014). El dominio PIWI cobra mayor importancia en Argonautas que tienen función catalítica, como Ago2 de humano, ya que este está constituido además por el motivo DEDX, en donde X usualmente se refiere a ácido aspártico o histidina –motivo que le otorga capacidad a algunos Argonautas de actuar como si fuera una RNAsa-H (Swarts *et al.* 2014).

El dominio PAZ se une al extremo 3' del microRNA por medio de interacciones con el esqueleto de azúcar y fosfato de las últimas bases del microRNA; aunque esta unión no es esencial para la interacción con el mRNA blanco (Wang *et al.* 2008), le otorga estabilidad al microRNA protegiéndolo de enzimas que puedan degradarlo (Hur *et al.* 2013). Este último punto es muy importante para explicar la conservación de los microRNA en animales, especialmente en mamíferos. A pesar de que su tamaño oscila entre 21-23 nt, la región conservada comprende típicamente las posiciones 2-8 o 2-9 de los microRNAs guía; las posiciones 10-22 no están conservados, ni entre familias de microRNAs dentro de una misma especie, ni entre especies. La explicación mecánica de este fenómeno vino de

estudios bioquímicos donde se pudo medir la afinidad que tienen las diferentes regiones de un microRNA con sus mRNA blancos. Por medio de experimentos de *binding* se descubrió que las posiciones 13 en adelante del microRNA guía tienen poca o nula participación tanto en la unión, como en la catálisis, del corte de mRNA mediado por Ago2. Por lo que el modelo actual de la función del complejo RISC propone a los dos extremos del microRNA, 5' y 3', firmemente anclados al MID y PAZ, respectivamente (Tomari *et al.* 2004). De esta manera la posición 1 del microRNA guía, al estar anclada por medio de su fosfato al dominio MID, no participa en la unión con el mRNA blanco; la región que muestra conservación entre familias y especies, que comprende las posiciones 2-8, se le conoce comúnmente como región *semilla* (Figura 2).

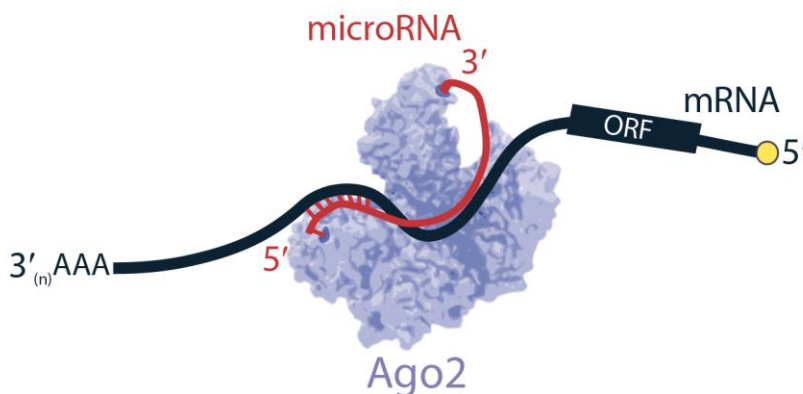


Figura 2. Unión del complejo RISC al mRNA. Se ilustra la manera en la que el complejo RISC reconoce al mRNA blanco por medio de la región semilla (nt 2-8) del microRNA guía.

Los microRNAs y sus RNAs blanco

Posterior a la formación del complejo RISC, el microRNA está listo para hibridarse a un RNA. A pesar de que es comúnmente encontrado en la literatura que los microRNAs silencian la traducción de mRNAs, sus RNAs blanco no solamente comprenden mensajeros. Básicamente, RISC puede hibridarse a cualquier RNA que se encuentre accesible y cuya secuencia sea complementaria a la región semilla (Agarwal *et al.* 2015). Sin embargo, solamente los microRNAs que estén unidos a las regiones 3'-UTR tienen función sobre el silenciamiento de los mRNAs a los que se encuentra unidos (Agarwal *et al.* 2015).

Probablemente uno de los aspectos más fascinantes del RISC es la cantidad de nucleótidos con los que la hebra guía hibrida con los RNAs blanco (Bartel, 2018). A pesar de contar con 6-7 nt para su unión, la región semilla en el microRNA guía es altamente específica y tiene una gran afinidad por el RNA al que se encuentra

unido (1–10 pmolar) (Wee *et al.* 2012). Esto tiene varias implicaciones bioquímicas que ayudan a comprender mejor la función biológica de los microRNAs. Probablemente la más importante de mencionar es que la proteína Argonata preordena la conformación del microRNA de hebra sencilla, a una forma hélice alfa, lo cual predispone a la secuencia en las posiciones 2–8 a unirse a algún RNA complementario, dando lugar de esta manera a la región semilla (Wee *et al.* 2012; Schirle *et al.* 2014).

De la explicación mecánica anterior, es fácil imaginar que un RISC tiene la capacidad de unirse prácticamente a cualquier RNA complementario. Lo cual lleva a preguntar, entonces ¿qué hace a un RNA ser un blanco de un microRNA? La respuesta a esta pregunta no es sencilla. Inicialmente se pensaba que los únicos RNA blanco de los microRNAs eran mRNA dispuestos a ser traducidos. Sin embargo, con el paso de los años y después de muchos años de investigación, se ha encontrado que hay varios tipos de RNAs que también podrían ser blancos de microRNAs, como lncRNAs, tRNAs, rRNA. Muchos de estos otros tipos de RNAs, diferentes a mRNAs codificantes, inclusive se han propuesto como inhibidores o *esponjas* de microRNAs, de tal manera que su alta abundancia celular compite con los mRNA blanco, y estos últimos se traducen. A esta explicación se le conoce como la “hipótesis ceRNA” (del inglés, *Competitive Endogenous RNAs*) (Chitwood *et al.* 2007; Ebert *et al.* 2007; Seitz, 2009).

El modelo actual sobre la función reguladora del silenciamiento de mRNAs mediado por microRNAs se muestra en la **Figura 3**. Un microRNA, cargado en un RISC, puede hibridarse a muchos mensajeros (**Figura 3A**); cada mRNA tiene diversos sitios de unión a diversos microRNAs (**Figura 3B**), cada uno con secuencias y abundancias diferentes (Ulhman S, *et al.* 2012). La estimación actual es que aproximadamente 60–70% de los mRNA producto de la transcripción de los genes del genoma humano están bajo regulación mediada por microRNAs (Catalanotto *et al.* 2016; Friedman *et al.* 2009).

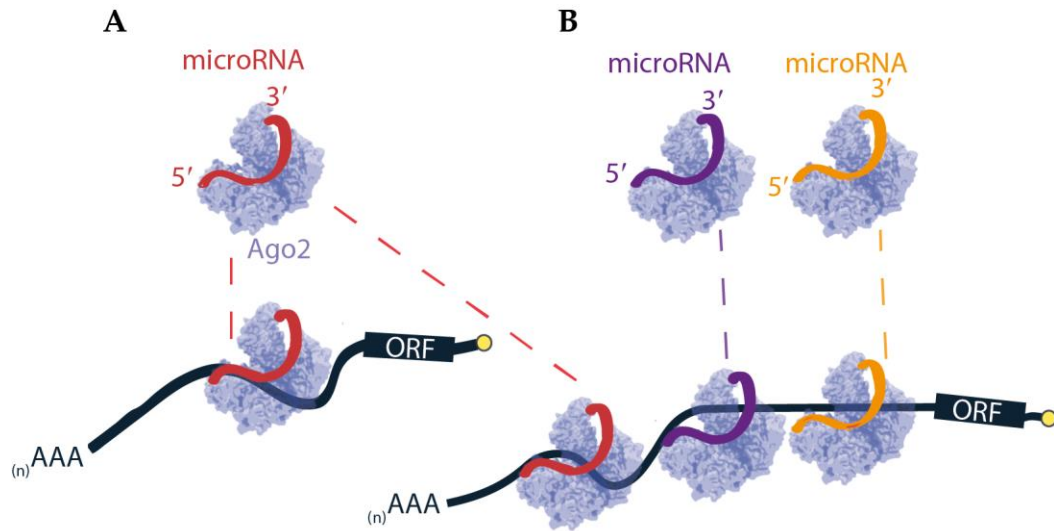


Figura 3. Silenciamiento de mRNAs mediado por microRNAs. **A)** un microRNA regula varios mRNAs. **B)** un mRNA está regulado por varios microRNAs.

Por la intrincada red de regulación mediada por los microRNAs, es de esperarse que dentro de los blancos de estos haya mRNAs correspondientes a genes importantes en el desarrollo y metabolismo. Por lo que la desregulación de microRNAs está asociada a distintas patologías, y por ello se ha propuesto que algunos podrían funcionar como posibles oncogenes. Por ejemplo, miR-17 se ha propuesto como bio-marcador para cáncer, además de ser blanco potencial para el desarrollo de terapias basadas en RNA (Janssen *et al.* 2013; Ventura *et al.* 2008).

Biogénesis alternativas de microRNAs

Además de la biogénesis canónica explicada anteriormente, también existen variantes de esta que dan lugar a microRNAs maduros funcionales. Por ejemplo, los miRtrons son microRNAs que se derivan de fragmentos de RNA productos del splicesosoma y provenientes de intrones. Estos fragmentos de tallo y asa no son cortados por el complejo micro-procesador sino por la maquinaria de *splicing* y son exportados directamente hacia el citoplasma por la XPO5. Una vez en el citoplasma, continúan con la vía canónica y se vuelven sustratos directos de Dicer para convertirse en microRNAs maduros (**Figura 4A**) (Havens *et al.* 2012).

Otro ejemplo es el de miR-451, que tiene un proceso de maduración independiente de Dicer. Se ha reportado que después de ser recortado por el microprocesador, mir-451 es exportado al citoplasma, posiblemente por la XPO 5, una vez en el citoplasma mir-451 se asocia directamente con Ago2, la única Ago que tiene actividad catalítica en humano. Esta asociación resulta en el corte del asa que compone al mir-451. Posteriormente la ribonucleasa específica de poli-a

(PARN) se encarga de recortar los nucleótidos salientes del miR-451 en dirección 3'-5', hasta dejarlo con el tamaño típico de 22 nt de los miRNAs maduros (**Figura 4B**) (Ha M, *et al.* 2014).

Existe otra situación en la que la vía de maduración requiere de modificaciones extra para producir microRNAs funcionales. Por ejemplo, el microRNA *let-7* depende de una uridil transferasa terminal (TUTasa) para poder ser procesado por Dicer. El procesamiento del pri-miRNA de *let-7* deja como resultado un pre-miRNA que únicamente tiene un nucleótido sobresaliente en su extremo 3', esto dificulta que Dicer pueda reconocerlo como sustrato, y por lo tanto necesita de la adición de un uracilo en el extremo 3' por medio de alguna TUTasa, para continuar con el proceso de maduración (**Figura 4C**) (Ha M, *et al.* 2014).

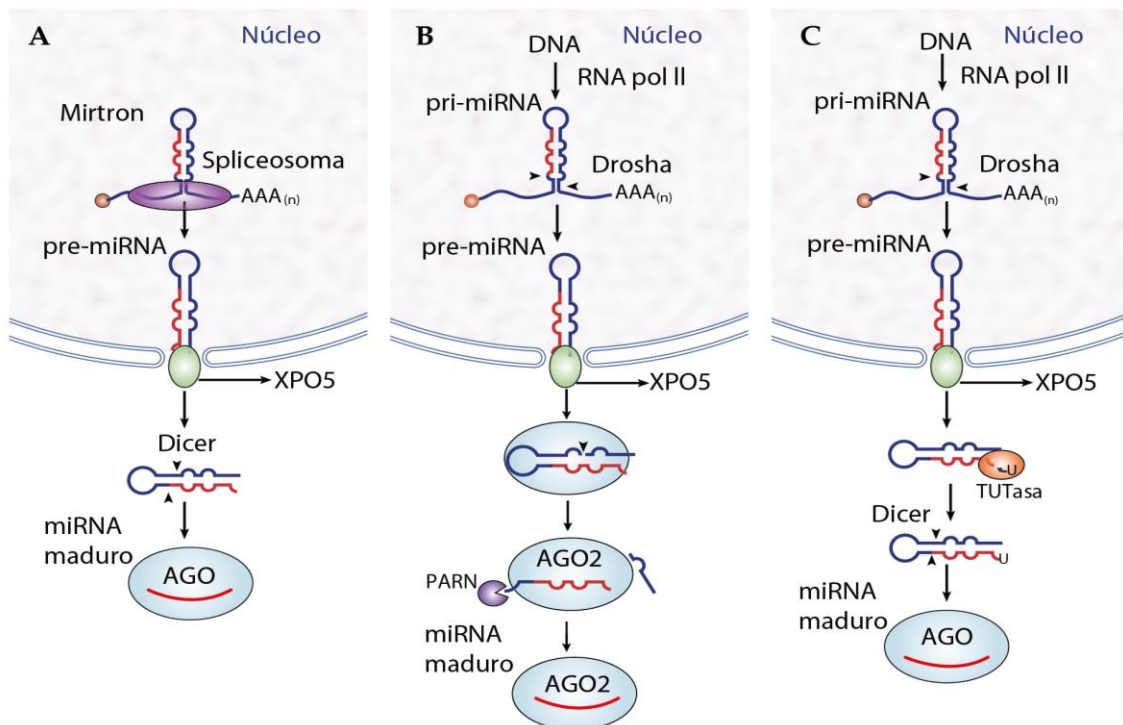


Figura 4. Biogénesis alternativa de los microRNAs. **A)** Biogénesis independiente del Microprocesador. **B)** Biogénesis independiente de Dicer. **C)** Biogénesis dependiente de TUTasas.

La base de datos de microRNAs: miRBase

Para el registro y anotación de todos los microRNAs que se han descrito hasta la fecha, Sam Griffiths-Jones estableció miRBase en el año 2004; desde entonces alberga a todos los microRNAs de todas las especies de organismos en las que se han detectado microRNAs. Esta es la base de datos de microRNAs más grande que existe y cuenta con las secuencias de los microRNAs maduros reportadas como “canónicas”, además de las secuencias que pertenecen a sus formas de pre-

miRNAs. En los primeros años de miRBase, la nomenclatura fue un tanto confusa debido a que se anotaban nuevos microRNAs sin tomar en cuenta que algunas secuencias ya existían en miRBase. Con el transcurso de versiones posteriores de miRBase, esto se ha reducido y ahora la anotación de microRNAs, contiene tanto secuencias originalmente denominadas como guías, hasta secuencias denominadas con un asterisco (hebra pasajera). Por ejemplo, la estructura de tallo y asa del pre-miRNA produce dos moléculas de microRNAs maduros de sus brazos (o hebras) opuestos. Para su nomenclatura, los microRNAs maduros provenientes del brazo que incluye el extremo 5' se le añade la notación "5p", mientras que para el brazo que incluye el extremo 3' se le añade la notación "3p"; de tal forma que la notación queda de la siguiente manera: *hsa-mir-128* es el pre-miRNA, su microRNA maduro proveniente del brazo 5' del pre-miRNA es *hsa-miR-128-5p*, y la secuencia madura que proviene del brazo 3' es *hsa-miR-128-3p* (Griffith-Jones, *et al.* 2006). (Kozomara *et al.* 2019).

La numeración de los microRNAs se basa en un nombre secuencial que incluye la identidad de la especie en la que fueron descritos. Por ejemplo, *hsa-miR-10* se refiere al microRNA 10 descrito en la especie *H. sapiens*. La numeración también puede hacer evidente la relación funcional que pueden tener algunos microRNAs por ejemplo: *hsa-miR-10* y *mmu-miR-10* son microRNAs provenientes de genes ortólogos entre humanos y ratones. A los microRNAs parálogos, con diferencias de una o dos posiciones, se les otorgan letras secuenciales como sufijo, por ejemplo: *hsa-miR-10a* y *hsa-miR-10b*. Para las secuencias que provienen de distintas moléculas precursoras pero que dan como resultado una secuencia madura idéntica se les añade un sufijo numérico, por ejemplo: *mmu-miR-281-1* y *mmu-miR-281-2*. Por último, si se habla de la secuencia madura, el afijo "miR" se escribe con la letra mayúscula R, mientras que si se hace referencia a la molécula precursora, el afijo "mir" se escribe con minúscula.

Las familias de microRNAs pueden incluir a aquellos que se encuentren relacionados evolutivamente, siempre y cuando sigan compartiendo las mismas características para elegir a sus mRNA blancos. Por otro lado, aquellos microRNAs que hayan divergido en estas características, no se consideran dentro de una misma familia aunque compartan un linaje evolutivo (Bartel, 2018).

Antecedentes

Desde el descubrimiento de cientos de secuencias de microRNAs en plantas y animales en el año 2001, varios grupos de investigación se comenzaron a dar cuenta de que las secuencias que correspondían a un microRNA, no siempre tenían una secuencia uniforme, sino que tenían variaciones –principalmente en el extremo 3' (Lagos-Quintana *et al.* 2001). Esta observación no tuvo mucha repercusión durante aquel tiempo debido a que el mero hallazgo de los microRNAs fue suficiente como para opacar sus variantes. Fue hasta varios años después, en 2008, que formalmente se describieron variaciones en las secuencias de los microRNAs con respecto a las secuencias que ya se tenían como referencia en sitios como miRBase. Se definió a estas variantes como isómeros de microRNAs, o simplemente “isomiRs”, y se comenzó a reflexionar acerca del origen de los mismos, relacionándolo a variaciones en el proceso de biogénesis de los microRNAs (Morin *et al.* 2008).

El estudio de los isomiRs ha ido cobrando importancia a lo largo de los últimos diez años y principalmente ha estado ligado al metabolismo de los microRNAs, e incluso se han generado herramientas bioinformáticas dedicadas únicamente a su análisis. Las técnicas de secuenciación masiva trajeron consigo la capacidad de caracterizar e identificar a microRNAs y a isomiRs. El hecho de que se encuentran en proporciones constantes en distintas fases del estadio de desarrollo de animales como *D. melanogaster*, sugirieron fuertemente que los isomiRs no son simplemente productos de degradación o errores de las técnicas de secuenciación masiva (Fernández-Valverde *et al.* 2010). Ensayos *in vitro* demostraron la capacidad funcional de los isomiRs, ya que son capaces de ensamblarse en RISC y de unirse a mRNA blancos (Tan G, *et al.* 2014).

La creciente identificación y descripción de nuevos isomiRs para cada microRNA promovió la creación de repositorios de acceso libre que permitieran el almacenamiento y logística de esta nueva información. Un ejemplo es el repositorio IsomiR-bank, el cual alberga más de 308,000 isomiRs provenientes de 4,000 familias de microRNAs de ocho especies diferentes (Zhang *et al.* 2016). Además, se crearon diversas herramientas bioinformáticas que se especializan en el análisis y reporte de isomiRs como las antes mencionadas isomiR-SEA o isomiR-ID (Amsel *et al.* 2017). Gracias a ello, existen reportes en los que se ha identificado que los isomiRs pueden expresarse de manera diferencial entre tejidos, y también que esto correlaciona en diversas condiciones patológicas (Guo *et al.* 2016). Existen estudios bioinformáticos en los que se ha podido distinguir distintos tipos de cáncer a partir del análisis de presencia y ausencia de isomiRs (Telonis *et al.* 2017);

usando esta información se pudieron construir perfiles de isomiRs que distinguen entre varios subtipos de cáncer de mama (Lan *et al.* 2018).

Tipos de isomiRs

Los isomiRs que están constituidos básicamente por adiciones, deleciones o variaciones de secuencia. De acuerdo con la ubicación de la modificación de la secuencia que hace diferente al isomiR con respecto su microRNA anotado en miRBase, los isomiRs se han clasificado en tres tipos: isomiRs 5', isomiRs polimórficos, e isomiRs 3' (Lan *et al.* 2018).

Los isomiRs 5' (**Figura 5A**) son aquellos en los que el extremo 5' del microRNA es diferente con respecto al microRNA anotado en miRBase. Estos cambios se deben a adiciones o eliminaciones de algunos nucleótidos. Puesto que la región semilla del microRNA se encuentra constituida por los nt del 2 al 8, es posible que algunos isomiRs 5' modifiquen dicha región y por lo tanto, esto pueda ocasionar que se pierda la afinidad por su mRNA blanco y que además puedan hibridarse y regular otros mRNA distintos (Chin T, *et al* 2014).

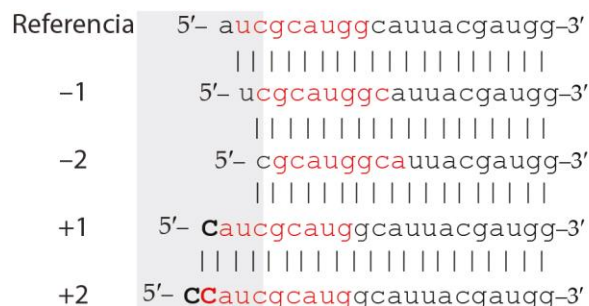
Los isomiRs polimórficos (**Figura 5B**), están constituidos por aquellas variaciones que se encuentran dentro del cuerpo de la secuencia del microRNA. Este tipo de isomiRs se ha descrito como una de las formas más raras e inusuales de encontrar y dependiendo de la cercanía de la modificación con la región semilla, este puede alterar los mRNA blanco contra los que tiene complementariedad (Neilsen *et al.* 2012).

Los isomiRs 3' (**Figura 5C**), dan origen a fenómenos denominados “trimming” o “tailing”. El “trimming” ocurre cuando la secuencia original del microRNA es recortada en su extremo 3' dando como resultado isoformas de secuencias más cortas. Por ejemplo, cortes imprecisos de Dicer podrían ser una causa de que existan isomiRs más cortos en el extremo 3'. El proceso de “tailing” consta de la adición de nucleótidos al extremo 3' y estos puede ser de dos tipos: 1) si los nucleótidos añadidos coinciden con la secuencia del pre-microRNA, se dice que es una modificación “templated” (o con molde), ya que hace referencia a que el isomiR conserva la secuencia original del molde del microRNA, en este caso el pre-microRNA. Los isomiRs que tienen adiciones que no corresponden con la secuencia molde de las moléculas precursoras reciben el nombre de adiciones “non-templated”, haciendo referencia a que las bases que se integran a la molécula del microRNA no coinciden con la secuencia molde de los precursores antes de pasar por los recortes de maduración, y son por lo tanto, resultado de *tailing*.

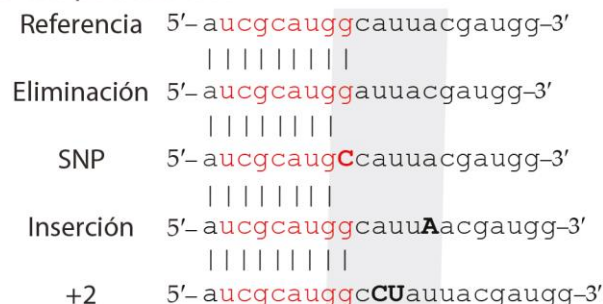
Las adiciones más comunes están representadas por adenilaciones (adiciones de adenina) y uridilaciones (adiciones de uracilo) en el extremo 3' del isomiR. También se conoce que su abundancia es tejido específica y pueden variar dependiendo de los estadios de desarrollo y condiciones patológicas (Wyman S. *et al.* 2011).

Generación de isomiRs

A IsomiRs 5'



B IsomiRs polimórficos



C IsomiRs 3'

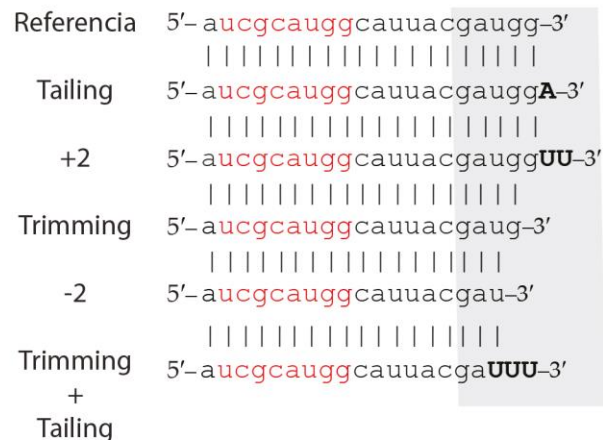


Figura 5. Tipos de isomiRs. **A)** isomiRs 5'. **B)** isomiRs polimórficos. **C)** isomiRs 3'. En rojo se resalta la región seed de cada microRNA, en gris se señala la región de la secuencia del microRNA en la que se ilustran las variaciones (en negritas) con respecto a la secuencia de referencia de miRBase.

Existen diversos mecanismos que pueden generar isomiRs. Estos pueden estar relacionados con algunas enzimas que también están involucradas en el proceso de biogénesis de los microRNAs, o también por proteínas que aparentemente no participan en el proceso canónico de maduración de microRNAs, pero si en la formación de isomiRs. Por ejemplo, existe evidencia que sugiere que cambios en la secuencia interna de algunos microRNAs pueden generar motivos que ocasionen abultamientos en la estructura del pre-microRNA y al momento de encontrarse con Dicer, se produce un corte impreciso y como consecuencia se formen isomiRs (Ma *et al.* 2016).

El mecanismo que utilizan Dicer y Drosha para procesar sus RNAs sustrato se basa en el reconocimiento de la estructura secundaria de las moléculas de tallo y asa (Li *et al.* 2016). A pesar de ser un corte preciso, la topología o estructura secundaria de cada molécula de RNA puede variar ligeramente, lo que ocasiona que ambas RNAsas pueden generar cortes imprecisos y por consiguiente, generar isomiRs en los extremos 5' o 3' (**Figura 6**).

Las adiciones *non-templated* se originan de interacciones del microRNA con algunas enzimas que se encargan de agregar distintos nucleótidos. Esto también es causa de que dichas adiciones puedan contribuir con la estabilidad y vida media del microRNA (Kato T. *et al.* 2009). Un ejemplo específico de este mecanismo es el de *hsa-miR-21* en el cual la adenilación por parte de la PAPD5 ocasiona que el microRNA sea degradado (Boele J. *et al.* 2014). Se ha reportado que PAPD4 provoca la mono-adenilación de miR-122 en tejido hepático en murinos, aumentando así su estabilidad (Burns D. *et al.* 2011).

La uridilación también está relacionada con procesos que pueden alterar la función de los microRNAs. Por ejemplo, la uridilación de miR-26b en humano está mediada por la TUT4, y esto reduce la represión de sus mRNA blancos (Jones M. *et al.* 2009); además de que funciona como una marca para la degradación de varios microRNAs durante la activación de células T del sistema inmunológico (Gutierrez-Vázquez C. *et al.* 2017).

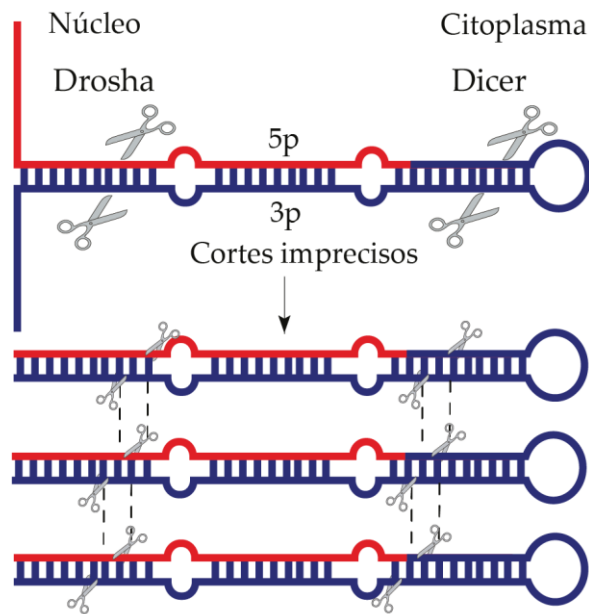


Figura 6. Cortes imprecisos de Dicer y Drosha. Se señalan las regiones en donde se generan los cortes imprecisos en el pri-miRNA y pre-miRNA por Drosha y Dicer, respectivamente.

El análisis bioinformático de los microRNAs y su influencia en el estudio de los isomiRs

Desde que comenzó el estudio molecular de los microRNAs, fue evidente que las secuencias que dan lugar a sus formas maduras de usualmente llevan consigo cambios de tamaño (Lagos-Quintana *et al.* 2001). Las herramientas que se tuvieron disponibles para estudiar los microRNAs, típicamente contemplaban alguna forma de radioisótopos y sondas que los reconocieran por medio de complementariedad. Con el auge de la secuenciación masiva y las grandes cantidades de datos que se podían obtener de las bibliotecas que se secuenciaron, surgieron nuevas necesidades para poder procesar y hacer más fácil la interpretación de los resultados obtenidos. Las herramientas bioinformáticas y bases de datos en línea llegaron a cumplir objetivos simples, y poco a poco se han transformado y especializado de acuerdo a las necesidades de las preguntas de investigación.

El descubrimiento de los microRNAs, en conjunto con las innovaciones tecnológicas (como las técnicas de secuenciación masiva), trajeron consigo una nueva manera de identificar y estudiar la gran diversidad de estos RNAs pequeños y a su vez dieron paso al desarrollo de herramientas bioinformáticas que son capaces de almacenar y organizar a los microRNAs en bases de datos disponibles para todo el mundo e incluso son capaces de predecir los posibles blancos que los microRNAs pudieran tener en el genoma.

Existen distintas herramientas que se encargan de identificar y de caracterizar a los isomiRs. Con el fin de poder ordenar y de hacer más accesibles las herramientas bioinformáticas, han surgido sitios web como tools4miRs, que ayudan a gestionar y comparar las distintas herramientas, ponderando su grado de complejidad de conocimiento informático necesario para operarlas, mostrando también sus ventajas y desventajas (Lukasik *et al.* 2016). Algunos estudios se han enfocado en realizar comparaciones entre las distintas herramientas bioinformáticas que se encargan de identificar isomiRs, por ejemplo, se conoce que una de las herramientas más famosas, isomiR-SEA únicamente tiene una sensibilidad del 25% al 30% para identificar isomiRs en donde la región semilla fue mutada. También se conoce que isomiRID, otra herramienta que soporta la identificación de isomiRs de diferentes grupos de animales, tiene la mayor tasa de reconocimiento de isomiRs, pero solamente representando una mutación a la vez (Amsel *et al.* 2017). Por ello, actualmente sería difícil imaginar el estudio de isomiRs, con sus microRNAs y familias de los mismos, sin el apoyo de herramientas computacionales.

Planteamiento del problema

A pesar de que existen trabajos en los que se ha reportado la importancia biológica del estudio de los isomiRs, un porcentaje muy pequeño de la literatura enfocada en el estudio de los microRNAs describe o analiza el contenido o participación de isomiRs. Esto puede explicarse debido a que la manera actual de evaluar la presencia o ausencia de microRNAs obedece tendencias en la investigación que facilitan su estudio y en algunos casos están dirigidas por tendencias comerciales. El caso más común es, por ejemplo, el uso de sondas para realizar ensayos de qRT-PCR. Estos ensayos con sondas comerciales, a pesar de ser bastante eficaces y poder diferenciar con alta precisión secuencias entre microRNAs muy parecidos, están dirigidos únicamente a las secuencias de referencia de miRBase y esto puede ocasionar la pérdida, o el sesgo de información importante acerca la abundancia de isomiRs.

Como se mencionó anteriormente, existen estudios que demuestran que en ocasiones el cambio en la expresión diferencial está dado en los isomiRs, y que las secuencias detectadas, ya sea por RT-qPCR o microarreglos correspondientes al microRNA de referencia reportado en miRBase permanece constante (Pillman *et al.* 2019). Por análisis hechos en nuestro laboratorio acerca de la abundancia y presencia de microRNAs y sus isomiRs, hemos identificado que algunos de estos últimos pueden ser inclusive más abundantes que las secuencias de referencia de microRNAs en miRBase. Esta observación nos lleva a preguntar si el que haya isomiRs más abundantes que sus microRNAs de referencia es consecuencia de un fenómeno aislado para algunas muestras y algunos isomiRs, o si en realidad es algo más generalizado y más bien depende del tipo de muestra, o estado celular que dio origen a esa muestra biológica.

Desafortunadamente, esta pregunta no puede contestarse hasta el momento con las herramientas computacionales y los despliegues de información que abordan la abundancia de isomiRs. Esto es debido a que en todos los estudios donde se ha puesto de manifiesto la importancia funcional y biológica de los isomiRs, se considera a grupos de estos con respecto a un solo microRNA, y se *escogen* algunos ejemplos que correlacionan con algún fenómeno biológico. Por ejemplo, en la **Figura 7**, tomada de Pillman *et al.* 2019, se muestra cómo el análisis de isomiRs de un solo microRNA cambia con respecto a su secuencia de referencia. Sin embargo, se dice poco o nada de cómo están representados todos los demás microRNAs con respecto a sus isomiRs.

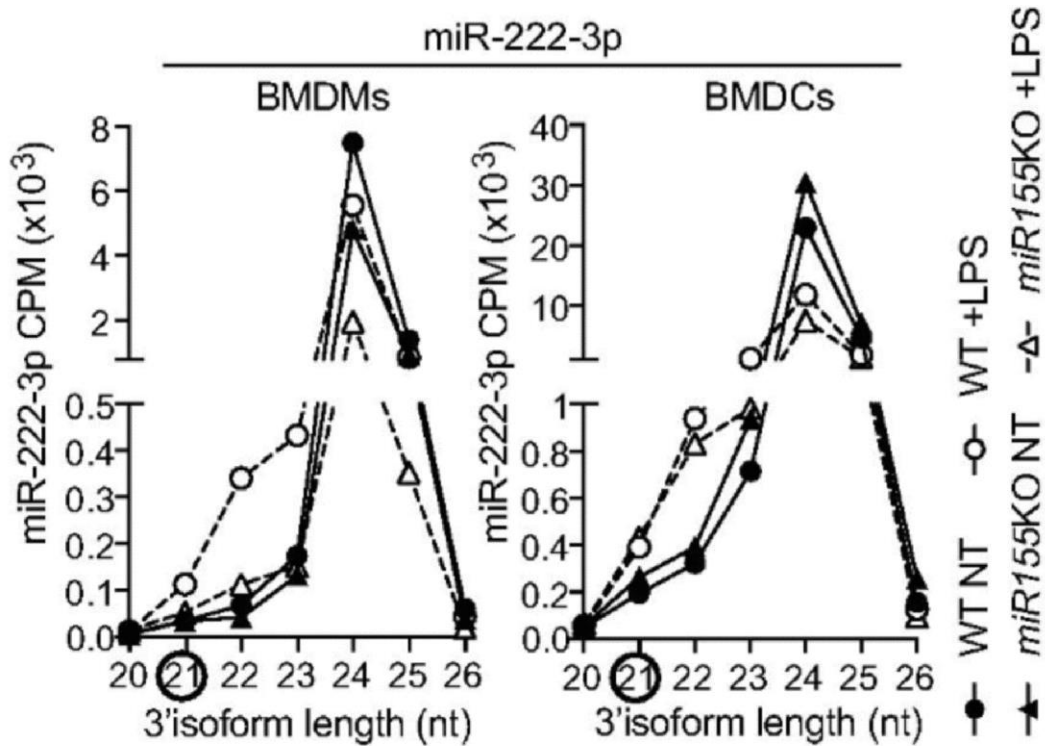


Figura 7. Análisis de abundancia de isomiRs de miR-222-3p en macrófagos murinos (BMDMs) bajo estímulos de factores de inflamación. Tomada de Pillman *et al.* 2019. En un círculo en el eje X se señala el tamaño (en nucleótidos) de la secuencia de referencia de miRBase de miR-222-3p, mientras que los demás números representan el tamaño de los isomiRs.

Esto claramente dificulta el poder entender el alcance e importancia funcional que tiene la abundancia de isomiRs en comparación con sus respectivas secuencias de microRNAs de referencia provenientes de miRBase. Por ello, para poder comenzar a tener un mejor entendimiento de los cambios de proporción entre cada microRNA y sus isomiRs, es necesario contar con representaciones gráficas y análisis entre la relación que existen entre ambos para tener mejor comprensión de su variabilidad, así como su funcionalidad en diferentes contextos biológicos y condiciones experimentales. El estudio actual de isomiRs considera a cada secuencia como una entidad independiente y únicamente se encarga de reportar la abundancia aislada de cada secuencia. Aunque existen diversas herramientas informáticas que están diseñadas para identificar y cuantificar isomiRs, no existe ninguna que permita determinar el panorama completo de isomiRs de cada microRNA cuantitativamente y que además permita comparar los resultados obtenidos con los de otras muestras.

Evidentemente, el estudio preciso de las isoformas de microRNAs no puede hacerse sistemáticamente sin el uso de la bioinformática de datos provenientes de secuenciación masiva. Esta tecnología es hasta el momento la herramienta más poderosa que existe para hacer análisis puntuales de secuencias de RNAs pequeños y además permite observar el panorama general de un conjunto diverso de secuencias, lo cual facilita su clasificación para posteriormente poder dar una interpretación científica. Otras herramientas, ampliamente usadas en el estudio de la abundancia de los microRNAs (por ejemplo, ensayos de RT-PCR en tiempo real), no tienen esta versatilidad. Esto es debido a que las sondas son diseñadas para la identificación de la secuencia de referencia de miRBase y es posible que esta sonda no pueda detectar a los isomiRs más abundantes (Pillman *et al.* 2019) lo que puede traer como resultado interpretaciones erróneas acerca de la homeostasis de microRNAs

Por esto, es necesario el desarrollo de una manera de poder estudiar a los microRNAs en conjunto y a cada uno de sus isomiRs. Al tratar de comprender la heterogeneidad en las secuencias de los microRNAs revelando las proporciones de las poblaciones de isomiRs, es posible que podamos hacer una inferencia sobre su origen y su participación en la vía metabólica de los microRNAs.

Objetivo general

Desarrollar un método computacional que permita cuantificar el panorama completo de isomiRs con respecto a la secuencia más representada por cada microRNA de distintas bases de datos de secuenciación masiva de RNAs pequeños.

Objetivos específicos

1. Obtener la anotación de microRNAs de distintas bases de datos provenientes de diferentes muestras biológicas.
2. Desarrollar una herramienta bioinformática que permita representar de forma cuantitativa el contenido de microRNAs y el panorama completo de isomiRs.
3. Utilizar la herramienta para obtener información acerca del contenido de isomiRs de diferentes bases de datos relacionadas con estudios exploratorios de secuenciación masiva de RNAs pequeños.

Métodos

Selección de las bibliotecas públicas

Para la selección de las bibliotecas que se utilizaron en este estudio, se consultó Gene Expression Omnibus (GEO), en la sección de secuenciación de RNAs pequeños (sRNA-seq). Las bibliotecas elegidas provienen de diversas muestras biológicas (**Tabla 1**) y fueron generadas con sistemas de secuenciación masiva de Illumina, específicamente con el protocolo de secuenciación *single end*. Para las bases de datos de muestras de tejidos, se incluyeron los grupos control y para las derivadas de líneas celulares, se eligieron experimentos relacionados con la biogénesis de los microRNAs (**Tabla 1**).

Las bases de datos en formato FASTQ se descargaron del European Nucleotide Archive (ENA).

Tabla 1. Descripción general de las bibliotecas que se utilizaron en este estudio.

Especie	Tejido o línea celular	Kit de construcción de bibliotecas	Experimento	Muestra	Referencia
<i>Homo sapiens</i>	Tejido renal 1	TruSeq Small RNA Sample Prep kit (Illumina)	SRX1274995	SRR2487699	Nientiedt <i>et al.</i> 2016
	Tejido renal 2		SRX1274995	SRR2487698	
	HCT116 WT		SRX1589539	SRR3174964	Kim <i>et al.</i> 2016
	HCT116 Drosha KO		SRX1589538	SRR3174963	
	HCT116 Dicer KO		SRX1589543	SRR3174968	
	HCT116 XPO55 KO		SRX1589541	SRR3174966	
	Tejido neuronal		NEBNext Small RNA Library Prep Set (Illumina)	SRX5551137	SRR8760619
<i>Mus musculus</i>	Tejido hepático	SRX3322458	SRR6213580		

La secuencia del adaptador 3' para las bibliotecas generadas con Illumina TruSeq Small RNA Sample Prep kit es 5'-TGGAATTCTCGGGTGCCAAGG-3'. Mientras que el adaptador 3' para las bibliotecas generadas con Illumina NEBNext Small RNA Library Prep es:

5'-AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC-3'.

Pre-procesamiento de las bibliotecas

Las bibliotecas descargadas del ENA requieren de un pre-procesamiento antes de realizar el análisis del contenido de microRNAs. Este procesamiento previo al análisis consiste en 1) Realizar el control de calidad del llamado de base por posición en la secuencia y 2) remover los adaptadores 3' que se agregaron en la construcción de la biblioteca. Todos los programas fueron utilizados desde la línea de comando de la terminal del sistema operativo de distribución libre Bunsen Labs.

Control de calidad del llamado de base

El control de calidad del llamado de base sirve para conocer la calidad de la identidad de los nt por posición de las secuencias que se encuentran en la base de datos.

Para consultar la calidad del llamado de base por posición de las secuencias de los datos crudos de cada biblioteca, se utilizó FastQC v.0.11.6, (Andrews S, 2010).

Corte de adaptadores y eliminación de secuencias de baja calidad

Para remover los adaptadores 3' de las secuencias de cada biblioteca, se utilizó Cutadapt v.1.16, (Martin M, 2011) con las siguientes instrucciones:

```
$ cutadapt -a adapter_sequence -m 18 -- discard-untrimmed -o out.fastq in.fastq
```

En donde:

- a: Secuencia del adaptador 3'.
- m: Longitud mínima permitida de las secuencias.
- discard-untrimmed: Sólo conserva aquellas secuencias a las que se les recortó el adaptador.
- o: Archivo de salida.

- in.fastq: Archivo de entrada.

-q: 20: Sólo conserva aquellas secuencias que tengan una calidad superior a 20 en la escala de phred.

Anotación de microRNAs

Una vez que los datos crudos de las bibliotecas fueron pre-procesados, se realizó la anotación de los microRNAs de cada una. Para la anotación de microRNAs se utilizó el programa Unitas v.1.4.6 (Gebert *et al.* 2017) con las siguientes instrucciones:

```
$ perl unitas_v.1.4.6 -input in.fastq -tail 4 -species species_name
```

En donde:

- input: Base de datos pre-procesada con Cutadapt.

- tail 4: Número máximo de nucleótidos adicionales permitido.

- species: Nombre de la especie que corresponde a la base de datos Ej. *Homo_sapiens*.

Normalización de las lecturas totales

Unitas genera la anotación de microRNAs y registra las lecturas totales de cada microRNA en cada biblioteca y las organiza como lecturas por microRNA (RPmiR del inglés *Reads Per MicroRNA*). Para poder realizar una comparación entre la abundancia de un microRNA entre muestras, es necesario normalizar las lecturas. Para normalizar las lecturas totales de las bibliotecas en este estudio, realizamos la transformación de lecturas totales a lecturas por millón (RPM del inglés *Reads Per Million*). La transformación se realiza dividiendo las lecturas totales de cada microRNA entre las lecturas totales mapeadas en toda la base de datos y multiplicando el resultado por un millón.

Análisis estadístico

Para determinar la correlación entre las proporciones de MAS obtenidas por el isomiR-index, se utilizó el coeficiente de Spearman, la prueba se realizó con un valor de $p \leq 0.0005$, de dos colas y fue calculado por medio de GraphPad Prism v.6.01.

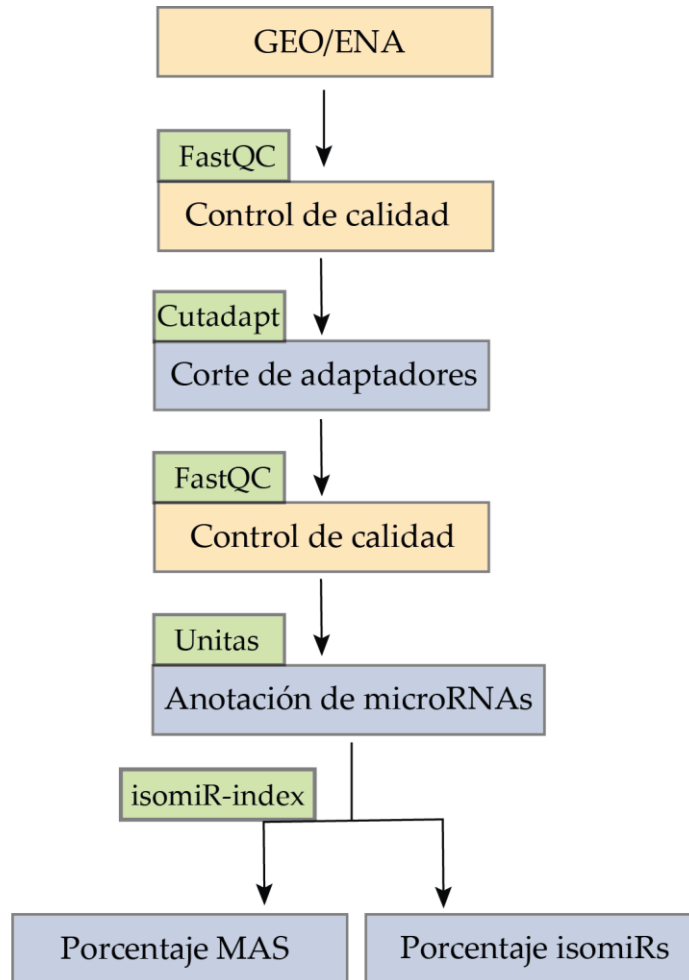


Figura 8. Flujo de trabajo para el análisis de las bibliotecas descargadas. En color verde se resaltan los programas utilizados para cada paso del procedimiento.

Ordenamiento de secuencias para calcular el isomiR-index

El isomiR-index utiliza como archivo de entrada la anotación de microRNAs creada por Unitas (**Figura 9 A**) para generar el perfil de isomiRs. Además, isomiR-index se encarga de ordenar primero las secuencias del perfil de acuerdo al microRNA (**Figura 9 B**), para posteriormente ordenar de mayor a menor de acuerdo al número de lecturas por microRNA en la muestra (RPmiR) (**Figura 9 C**).

A	miRNA	Secuencia	RPmiR
	miR-1	5'-ucaugcacggacauauggac-3'	8,000
	miR-3	5'-acugcacgacauauguaca GGG -3'	50
	miR-2	5'-ugcacggacauauggac-3'	300
	miR-2	5'-ugcacggacauauggacag UUU -3'	900
	miR-3	5'-acugcacgacauauguacaug-3'	250
	miR-1	5'-ucaugcacggacauauggac AA -3'	1,000
	miR-2	5'-ugcacggacauauggacag AAA -3'	600
	miR-3	5'-acugcacgacauauguaca UUU -3'	150
	miR-1	5'-ucaugcacggacauaug-3'	3,000
B	miRNA	Secuencia	RPmiR
	miR-1	5'-ucaugcacggacauaug-3'	3,000
	miR-1	5'-ucaugcacggacauauggac-3'	8,000
	miR-1	5'-ucaugcacggacauauggac AA -3'	1,000
	miR-2	5'-ugcacggacauauggacag UUU -3'	900
	miR-2	5'-ugcacggacauauggac-3'	300
	miR-2	5'-ugcacggacauauggacag AAA -3'	600
	miR-3	5'-acugcacgacauauguaca GGG -3'	50
	miR-3	5'-acugcacgacauauguacaug-3'	250
	miR-3	5'-acugcacgacauauguaca UUU -3'	150
C	miRNA	Secuencia	RPmiR
	miR-1	5'-ucaugcacggacauauggac-3'	8,000
	miR-1	5'-ucaugcacggacauaug-3'	3,000
	miR-1	5'-ucaugcacggacauauggac AA -3'	1,000
	miR-2	5'-ugcacggacauauggacag UUU -3'	900
	miR-2	5'-ugcacggacauauggacag AAA -3'	600
	miR-2	5'-ugcacggacauauggac-3'	300
	miR-3	5'-acugcacgacauauguacaug-3'	250
	miR-3	5'-acugcacgacauauguaca UUU -3'	150
	miR-3	5'-acugcacgacauauguaca GGG -3'	50

Figura 9. Ejemplo del ordenamiento de las bibliotecas por isomiR-index. **A)** anotación de microRNAs por Unitas. **B)** organización por microRNA. **C)** organización por RPmiR.

Resultados

La motivación de la implementación del concepto de MAS, y el propósito de la herramienta *isomiR-index*

miRBase es el principal repositorio de referencia donde se recopila la información de microRNAs de diversas especies. Además, funge como la referencia para la búsqueda de microRNAs maduros y sus precursores (Kozomara *et al.* 2019). A pesar de ello, las secuencias de referencia contenidas miRBase no representan necesariamente a las más abundantes de cada muestra.

Para ejemplificar esto, en la **Figura 10**, se muestra la captura de pantalla de miRBase para el microRNA *hsa-let-7a-2*. En esta figura, las lecturas individuales más abundantes que corresponden al *hsa-let-7a-2-5p*, producto de experimentos de secuenciación masiva, no necesariamente corresponden a la secuencia de referencia de miRBase (ver flechas negras). Esta situación prácticamente depende de cada microRNA, y posiblemente de cada muestra.

En el ejemplo, la secuencia de referencia del microRNA *hsa-let-7a-2* según miRbase es 5'-GAGGUAGUAGGUUGUAUAGUU-3', que tiene 4,041,019 cuentas (5.95×10^3 rpm) (flechas negras); sin embargo, la secuencia individual más abundante es la carece de los 3 últimos nucleótidos, y se tiene en el registro como: 5'-UGAGGUAGUAGGUUGUAUA-3', la cual tiene 8,951,657 cuentas (2.52×10^4 rpm) (flechas rojas). Además, muchas otras secuencias, que en este caso representarían también a los isomiRs de ese microRNA, tienen abundancias considerables. En este ejemplo, la secuencia más abundante no es idéntica a la referencia; situación que entorpece el estudio de los microRNAs porque es ambigua la definición y anotación de secuencias específicas de microRNAs cuando muchas otras, inclusive de abundancias ampliamente diferentes, también podrían serlo. Dado que es ambiguo aún saber *a priori* si la secuencia de referencia corresponde además a la más abundante, sería razonable hacer un análisis simplemente desde la perspectiva de las secuencias más abundantes con respecto a las demás.

Deep sequencing reads for stem-loop sequence MI0000061

Stem-loop ID hsa-let-7a-2

Confidence Annotation confidence: high
Feedback: Do you believe this miRNA is real? Yes No Leave comment

Reads

hsa-let-7a-5p	hsa-let-7a-2-3p	Count	RPM (mean number of reads per million)
AGGUUCAGGUAAGGUUUU...		1	0.469
GGUUCAGGUAAGGUUUU...		13	0.378
GGUUCAGGUAAGGUUUU...		3	0.216
GGUUCAGGUAAGGUUUU...		2	1.68
GGUUCAGGUAAGGUUUU...		1	0.232
GUGAGGUAAGGUUUU...		157	2.19
GUGAGGUAAGGUUUU...		66	0.823
GUGAGGUAAGGUUUU...		36	0.613
GUGAGGUAAGGUUUU...		6	1.14
GUGAGGUAAGGUUUU...		6	1.91
GUGAGGUAAGGUUUU...		4	0.621
GUGAGGUAAGGUUUU...		2	0.279
GUGAGGUAAGGUUUU...		1	0.0484
GUGAGGUAAGGUUUU...		1	3.38
UGAGGUAAGGUUUU...		3038	16
UGAGGUAAGGUUUU...		451	3.23
UGAGGUAAGGUUUU...		341	5.29
UGAGGUAAGGUUUU...		297	5.19
UGAGGUAAGGUUUU...		152	1.68
UGAGGUAAGGUUUU...		151	3.14
UGAGGUAAGGUUUU...		31	1.41
UGAGGUAAGGUUUU...		28	0.614
UGAGGUAAGGUUUU...		24	0.254
UGAGGUAAGGUUUU...		1	0.186
UGAGGUAAGGUUUU...		8951657	2.52e+04
UGAGGUAAGGUUUU...		4041019	5.95e+03
UGAGGUAAGGUUUU...		2156141	5.53e+03
UGAGGUAAGGUUUU...		1499318	3.69e+03
UGAGGUAAGGUUUU...		1243513	3.39e+03
UGAGGUAAGGUUUU...		415854	882
UGAGGUAAGGUUUU...		171929	237
UGAGGUAAGGUUUU...		86885	300
UGAGGUAAGGUUUU...		2295	5.6
UGAGGUAAGGUUUU...		145	4.38
UGAGGUAAGGUUUU...		53	0.936
UGAGGUAAGGUUUU...		14	0.909
UGAGGUAAGGUUUU...		5	0.203
GAGGUAAGGUUUU...		14899	82
GAGGUAAGGUUUU...		4999	27.5
GAGGUAAGGUUUU...		494	46.4
GAGGUAAGGUUUU...		2845	12.1
GAGGUAAGGUUUU...		2179	24.9
GAGGUAAGGUUUU...		614	3.95
GAGGUAAGGUUUU...		210	2.94
GAGGUAAGGUUUU...		7	0.166
AGGUAAGGUUUU...		6763	27.1
AGGUAAGGUUUU...		6084	19.3
AGGUAAGGUUUU...		1244	9.46
AGGUAAGGUUUU...		1074	5.7
AGGUAAGGUUUU...		305	2.49
AGGUAAGGUUUU...		123	1.66
AGGUAAGGUUUU...		2	0.17
GGUAAGGUUUU...		20416	67.7
GGUAAGGUUUU...		14927	40.4
GGUAAGGUUUU...		682	9.42
GGUAAGGUUUU...		419	6.83
GGUAAGGUUUU...		236	2.75
GGUAAGGUUUU...		2	0.119
GUAGGUAAGGUUUU...		3509	26.6
GUAGGUAAGGUUUU...		1586	23.4
GUAGGUAAGGUUUU...		131	4.69
GUAGGUAAGGUUUU...		67	1.05
UAGGUAAGGUUUU...		4478	253
UAGGUAAGGUUUU...		2195	137
UAGGUAAGGUUUU...		112	5.14
UAGGUAAGGUUUU...		4	0.194
UAGGUAAGGUUUU...		1	1.2
UAGGUAAGGUUUU...		1	0.0484
UAGGUAAGGUUUU...		1	0.0951
UAGGUAAGGUUUU...		1	0.0484
AGUAGGUAAGGUUU...		1437	101
AGUAGGUAAGGUUU...		37	6.5
AGUAGGUAAGGUUU...		1	0.0484
GUAGGUAAGGUUUU...		57	1.38
UAGGUAAGGUUUU...		1	0.187
UAGAUAUACAAGGGGA...		45	1.08
UAGAUAUACAAGGGGA...		12	0.626
UAGAUAUACAAGGGGA...		10	0.585
UAGAUAUACAAGGGGA...		9	0.626
UAGAUAUACAAGGGGA...		4	0.499
UAGAUAUACAAGGGGA...		3	0.701
UAGAUAUACAAGGGGA...		2	0.33
AGAUAUACAAGGGGA...		1	0.232
UACUGUACAGGCUUAG...		3	0.211
UACUGUACAGGCUUAG...		3	0.33
UACUGUACAGGCUUAG...		2	0.505
UACUGUACAGGCUUAG...		1	3.06
ACUGUACAGGCUUAG...		2	0.245
ACUGUACAGGCUUAG...		1	1.47
CUUACAGGCUUAG...		30	5.98
CUUACAGGCUUAG...		23	2.76
CUUACAGGCUUAG...		10	0.589
CUUACAGGCUUAG...		6	6.6
CUUACAGGCUUAG...		2	4.64
CUUACAGGCUUAG...		1	0.677
CUUACAGGCUUAG...		1	0.232
GUACAGGCUUAG...		1	3.76
AGGUUCAGGUAAGGUUUU...UAGAUAUACAAGGGGAGUAAUGUACAGGCUUAGCUUUUUU...			

(-25..28)

Figura 10. Captura de pantalla de miRBase para el microRNA hsa-let-7a-2. Las flechas negras señalan la abundancia en rpm y la secuencia de referencia. Las flechas rojas señalan al isomiR más abundante que la secuencia de referencia de miRBase. La fotografía pertenece a miRBase release 22.1, octubre 2018.

Por esta razón, diseñamos una herramienta computacional, *isomiR-index*, que permite analizar el panorama completo de isomiRs con respecto a la secuencia más abundante por microRNA en cada base de datos. Como primer paso identificamos a la secuencia más abundante del microRNA dentro de una muestra, independientemente de si corresponde o no a la secuencia de referencia reportada en miRBase, y la denominamos MAS (del inglés *Most Abundant Sequence*). Con excepción del MAS, los isomiRs representan la suma de la abundancia de lecturas de todas las secuencias detectadas correspondientes a un microRNA (**Figura 11**).

miR-x	RPmiR	
5'-ucaugcacggacauauggac-3'	8,000	— MAS
5'-ucaugcacggacauauggac G -3'	250	} isomiRs
5'-ucaugcacggacauau-3'	150	
5'-ucaugcacggacauauggac CC -3'	50	
5'-ucaugcacggacauauggac UAG -3'	30	
5'-ucaugcacggacauauggac UUUU -3'	10	
5'-ucaugcacggacauauggac AAA -3'	10	

Figura 11. Ejemplo de la selección del MAS y la fracción de isomiRs. (RPmiR significa lecturas por microRNA).

Una vez que se ordenan las secuencias del perfil de isomiRs, el script del *isomiR-index* selecciona la secuencia MAS y posteriormente realiza el cálculo de la abundancia en porcentaje de MAS y el restante de la fracción de isomiRs por cada microRNA contenido en la base de datos (**Figura 12**). Por lo tanto, el *isomiR-index* es el cálculo de la obtención de las proporciones de MAS vs el resto de las secuencias correspondientes a un microRNA dentro de una muestra.

	Número de lecturas	Proporción (%)
MAS	8,000	94.1
IsomiRs	500	5.9
Total	8,500	100

Figura 12. Cálculo del porcentaje del MAS y la fracción de isomiRs.

Con esta sencilla representación ordenada es posible asignar un valor porcentual a la abundancia de la proporción de isomiRs que tiene un microRNA. Un acercamiento de este tipo no se ha reportado en la literatura hasta la fecha. Con la obtención del valor de MAS, será posible poder graficar todos los valores correspondientes dentro de una muestra y visualizarlos para tener una noción general de cuáles y cuántos microRNAs tienen un alto contenido porcentual de isomiRs.

El isomiR-index es útil para comparar el contenido de isomiRs entre diferentes muestras

Para poner a prueba la utilidad del cálculo de MAS a través del isomiR-index, ocupamos datos de secuenciación masiva de acceso público de microRNAs provenientes de la base de datos del *Gene Expression Omnibus* (GEO).

Las primeras bibliotecas analizadas pertenecen a dos muestras independientes de tejido renal humano (Nientiedt *et al.* 2016). Con el resultado del análisis del isomiR-index, se construyó una gráfica que representa el porcentaje del MAS y el porcentaje de la fracción de isomiRs para todos los microRNAs de la base de datos. Adicionalmente, se representa la abundancia en la misma gráfica categorizada conforme a la distribución de las lecturas por microRNA (*reads per miR*, *RPmiR*) en percentiles de la siguiente manera (**Figura 13**):

microRNAs de abundancia baja, 0-25%; puntos en color gris

microRNAs de abundancia media, 26-75%; puntos en color azul

microRNAs de abundancia alta, 76-100%; puntos en color rojo

Los datos están presentados de tal manera que denotan el porcentaje de MAS; a medida que aumenta el porcentaje de MAS, los puntos se ubican hacia la izquierda (región en color verde). El porcentaje de isomiRs en cada muestra es el recíproco del MAS; a medida que aumenta la proporción de isomiRs, disminuye el de MAS (región en amarillo). Dado que por definición el conteo de isomiRs se hace a partir del MAS, conceptualmente no puede haber valores de MAS en cero, sin embargo sí es posible tener valores de 100 cuyo significado es que las secuencias de esos microRNAs son totalmente uniformes y no tienen variación en esa muestra.

Del resultado del análisis de ambas bibliotecas de secuenciación es evidente que los porcentajes de MAS calculados a partir del isomiR-index son variados. Algo interesante es que parece haber una ligera tendencia de los microRNAs menos abundantes (gris) a tener mayor contenido de isomiRs, con respecto a los más abundantes. Sin embargo, los diez microRNAs más abundantes aún

conservan una distribución de MAS variada, del 40-100% de MAS. Esto sugiere que, a pesar de que el isomiR-index se calcula a partir de las abundancias de cada MAS y sus isomiRs, la abundancia global de cada microRNA no parece influenciar al isomiR-index. Es importante recalcar que ambas muestras provienen de individuos diferentes, y aun así el patrón de MAS en ambas gráficas es aparentemente similar (**Figura 13A y 13B**).

La similitud de la distribución de puntos en ambas gráficas brinda la oportunidad de comparar los porcentajes de MAS presentes en distintas muestras. Por ello, el siguiente paso fue comparar los índices del MAS de todos los microRNAs presentes en la muestra de tejido renal 1, con las contrapartes en la muestra de tejido renal 2 y calcular su correlación (**Figura 14**). El coeficiente de correlación de Spearman para los valores del porcentaje del MAS asignado a cada microRNA de cada muestra de tejido renal tiene una correlación alta entre ambas muestras ($r = 0.9$). Este resultado tiene varias implicaciones. Primero, indica que la frecuencia y abundancia los isomiRs para cada microRNA pueden usarse para comparar muestras entre sí, ya que tejidos de individuos independientes conservan una alta correlación; típicamente esto solo se había hecho comparando la abundancia total de los microRNAs en lecturas por millón (reads per million, RPM), pero no de sus isomiRs ni sus proporciones. Segundo, dada la alta correlación entre las muestras, se apoya la idea de que los isomiRs y sus abundancias no son mera casualidad o resultado de degradaciones; muestras independientes no tendrían por qué conservar proporciones con una alta correlación.

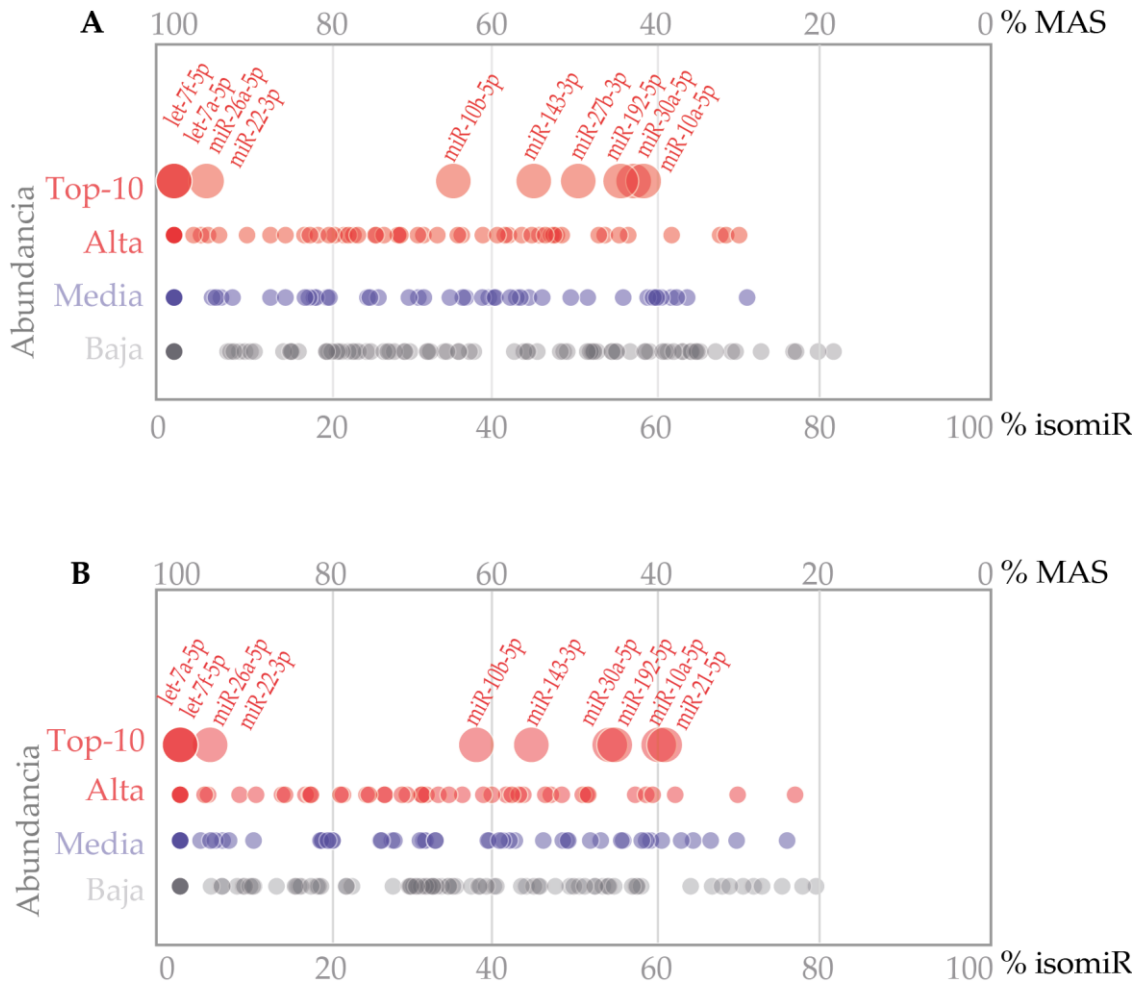


Figura 13. Porcentaje de MAS por cada microRNA en muestras de tejido renal. La abundancia corresponde a la distribución de las lecturas por microRNA (RPmiR). Baja, 0–25% (puntos grises); media 26–75% puntos azules; alta 76–100% puntos rojos. Los círculos rojos más grandes representan los 10 microRNAs más abundantes contenidos en cada muestra. **A)** tejido renal 1; **B)** tejido renal 2.

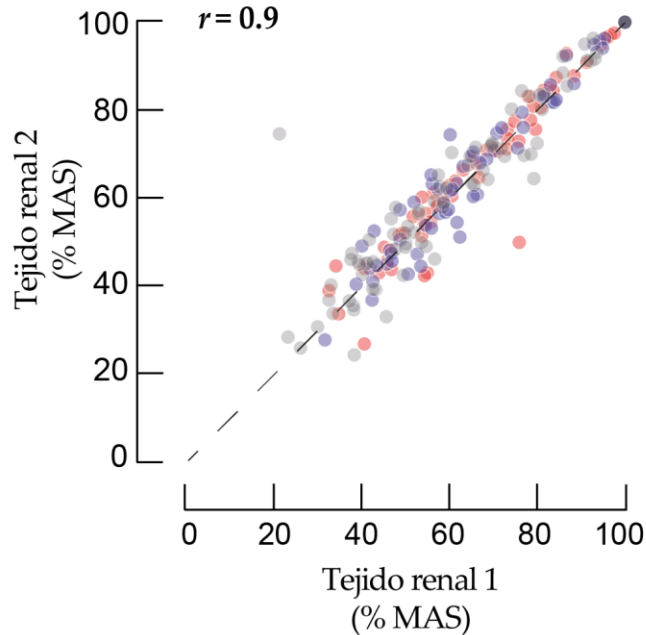


Figura 14. Correlación de porcentaje de MAS por cada microRNA en muestras independientes de tejido renal. El valor de correlación se obtuvo mediante la prueba de Spearman de dos colas con un valor de significancia de $p < 0.0005$. La abundancia corresponde a la distribución de las lecturas por microRNA (RPmiR). Baja 0–25% puntos grises; media 26–75% puntos azules; alta 76–100% puntos rojos. Solo se muestran los microRNAs y sus proporciones presentes en ambas muestras.

La abundancia de microRNAs dentro de una muestra es independiente de su porcentaje de MAS

A pesar de que el isomiR-index ayuda a asignar valores a las proporciones de isomiRs y MAS dentro de una muestra basándose en la abundancia, el porcentaje de MAS no es impactado por el rango de abundancia de un microRNA vs algún otro. La **figura 15** muestra un ejemplo que ilustra lo anterior. Si tomamos la correlación de MAS de las muestras de tejido renal de la **figura 14** y extraemos los puntos correspondientes a uno de los cuadrantes, al graficar nuevamente su correlación pero esta vez tomando únicamente su abundancia se observa que los puntos con bajo porcentaje de MAS (cuadrante en amarillo) se redistribuyen de acuerdo a sus abundancias (**Figura 15 A**). Del modo inverso, cuando se grafican los microRNAs por su nivel de abundancia y se toman únicamente aquellos cuya abundancia es baja (**Figura 15 B**, cuadrante amarillo), al re-graficar dichos puntos con respecto a sus porcentajes de MAS, observamos que nuevamente se

redistribuyen. Esta demostración es importante debido a que de ella podemos concluir que el isomiR-index refleja cambios en las proporciones de cada microRNA *vs* sus isomiRs, pero no en la abundancia entre microRNAs dentro de la muestra; los colores asignados a cada punto pueden o no ser incluidos en la gráfica.

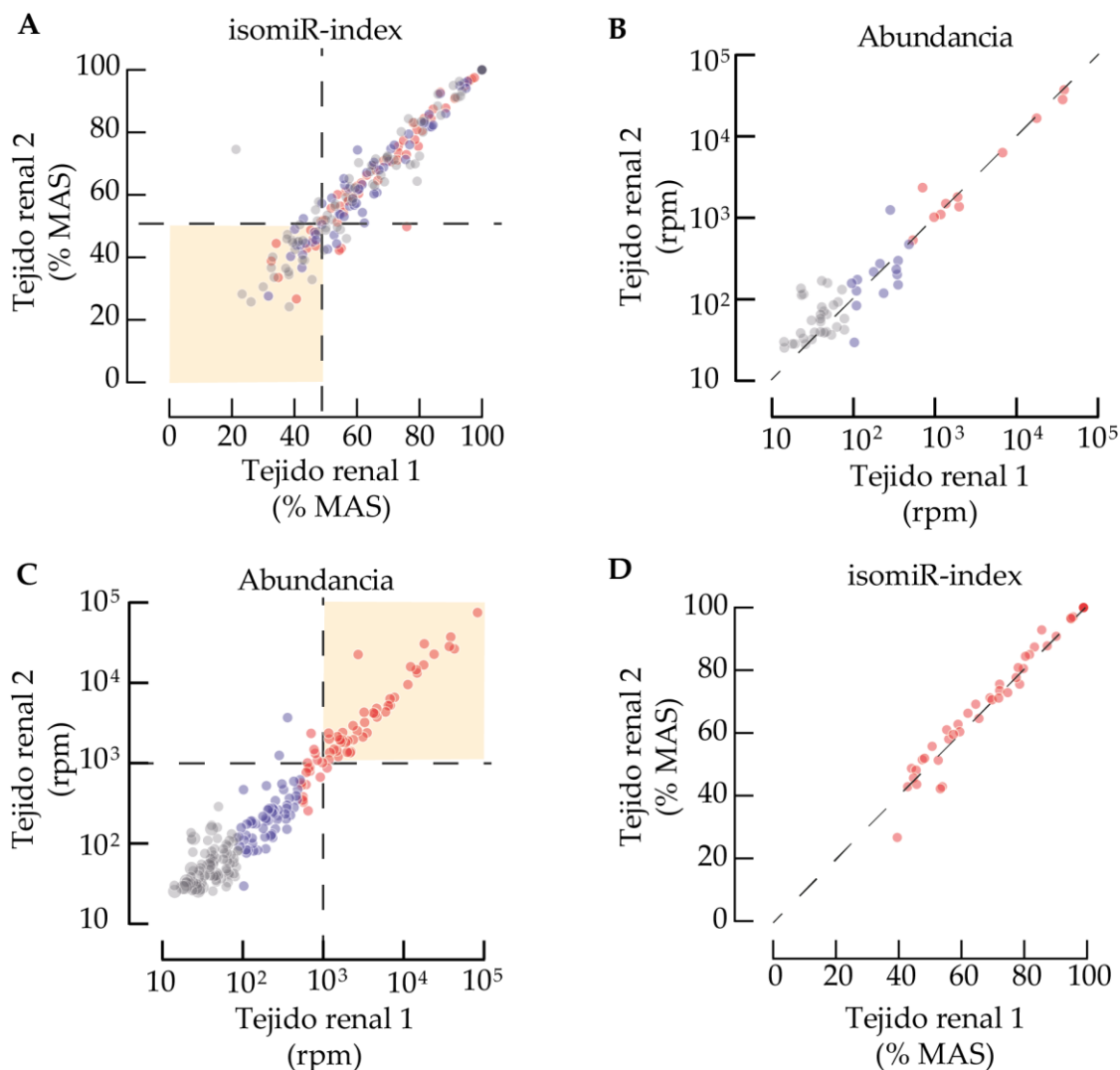


Figura 15. Comparación entre la correlación del porcentaje de MAS y la correlación de abundancia. **A)** Correlación entre los valores porcentuales de MAS generados por el isomiR-index para las muestras de tejido renal 1 y 2. **B)** Los puntos del cuadrante en amarillo se re-graficaron de acuerdo a su abundancia en RPmiR para las muestras de tejido renal 1 y 2. El código de colores en **B)** es el mismo que en **A)**. **C)** y **D)**, muestra en experimento inverso a **A)** y **B)**. Solo se muestran los microRNAs y sus proporciones presentes en ambas muestras.

Las proporciones del MAS e isomiRs son diferentes entre tejidos de diferentes especies

Una vez comprobado que el cálculo de obtención de MAS mediante el isomiR-index es útil para comparar muestras similares, decidimos preguntar si además sirve para discernir entre muestras cuya proporción de isomiRs no se espera que tengan alta correlación. Dado que los microRNAs son RNAs altamente conservados entre especies, se espera que haya una expresión similar en muchos de estos, aunque no necesariamente en el porcentaje de sus isomiRs. Además, ya que el isomiR-index es independiente de la abundancia de la expresión diferencial de microRNAs, las muestras por comparar deberían ser aquellas que cambien en su porcentaje de MAS. Por estas razones, decidimos usar muestras de tejido altamente especializados ya que la expresión de sus mRNAs regulables por microRNAs podría ser muy distinta, y abriría la oportunidad a que isomiRs funcionales reflejaran cambios en sus proporciones. Para este análisis buscamos muestras de estudios depositados en el ENCODE 2012 donde se reportaran experimentos de secuenciación masiva de microRNAs en tejidos altamente especializados, por ejemplo tejido neuronal y hepático. Por su grado de especialización, sería de esperarse que muchos de los mRNAs presentes, incluyendo microRNAs, fueran diferentes entre muestras. Por lo tanto quisimos preguntar si el contenido de isomiRs, representado por el porcentaje de MAS, es diferente si se comparan ambas muestras. Por otro lado, la diferencia entre muestras podría acentuarse si los microRNAs provienen de tejidos de diferentes especies, por lo que para este análisis comparamos los datos provenientes de neuronas de humano contra hígado de ratón.

Al comparar los porcentajes de MAS de las muestras de humano vs ratón, obtuvimos un coeficiente de correlación de Spearman bajo ($r = 0.05$) (**Figura 16**), indicando que la proporción de isomiRs de una muestra es diferente de la otra. Algo importante de señalar es que en tejido neuronal existe más uniformidad de secuencias en comparación con el hígado: los valores de 100 significan que el 100% de las lecturas de cada microRNA corresponden a una sola secuencia –ese microRNA no tiene isomiRs en esa muestra; por el contrario, en la muestra de hígado, los mismos puntos correspondientes contienen diversos porcentajes de MAS. Este es un hallazgo interesante, hasta el momento no se había reportado ningún estudio en donde se comparara la proporción de isomiRs en una muestra vs otra, y en donde se mostrara además que un tejido tiene más uniformidad que otro en sus secuencias de microRNAs. Algo de resaltar es que parece haber también un número importante de microRNAs en tejido neuronal cuya proporciones de MAS e isomiRs se acercan al 50%; este fenómeno no se observa en las muestras de tejido hepático. Es importante aclarar, que hasta el momento, el

isomiR-index no considera si la secuencia MAS es idéntica a la de referencia provista por miRBase.

En conjunto, los resultados presentados hasta esta sección muestran que la forma de analizar el contenido de isomiRs en diferentes muestras mediante la aplicación del isomiR-index, podría ser útil para analizar el contenido global de microRNAs y sus variaciones dentro de diferentes muestras.

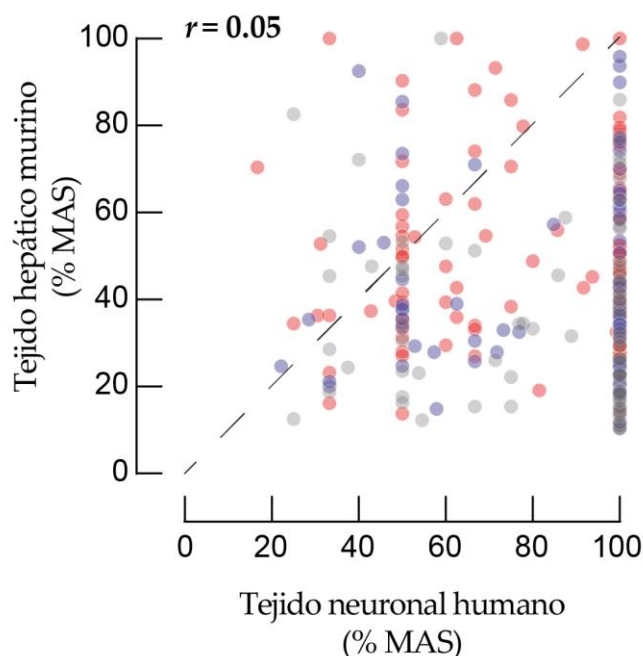


Figura 16. Correlación entre muestras de tejido neuronal de humano contra tejido hepático de ratón. Los colores corresponden a la abundancia en RPmiR de cada muestra. Baja 0–25% puntos grises; media 26–75% puntos azules; alta 76–100% puntos rojos. Solo se muestran los microRNAs y sus proporciones presentes en ambas muestras.

La proporción global de isomiRs se incrementa en células carentes de Dicer, pero no en Drosha o XPO 5

Para poder indagar si cambios en la homeostasis celular y la expresión global de las proporciones de microRNAs y sus isomiRs pueden verse reflejadas mediante el uso del isomiR-index, lo siguiente que hicimos fue analizar los datos de secuenciación masiva de células *knock-out* de proteínas involucradas en la biogénesis de microRNAs. Para esto, usamos unas bibliotecas de secuenciación

masiva de la línea celular HCT116, las mismas con las que Kim y colaboradores lograron hacer líneas *knock-out* de Dicer, Drosha, y XPO 5 (Kim *et al.* 2016). Como se explicó anteriormente, estas tres proteínas son importantes para el procesamiento de microRNAs hasta justo antes de su ensamblaje en RISC, –por lo que perturbaciones en el procesamiento de pri-, o pre-microRNAs podrían tener un efecto sobre la proporción de isomiRs (Starega-Roslan *et al.* 2015).

Al usar el isomiR-index y mostrar todos los microRNAs de las células parentales (HCT116 WT), se observa una diferencia notable en el número total de microRNAs que contiene con respecto a las que no expresan Dicer (Dicer KO) (**Figura 17 A y 17 B**). Otro elemento por resaltar es que parece haber una tendencia a que haya más isomiRs en las células knock-out de Dicer, ya que proporcionalmente hay más puntos en la región amarilla, en comparación con las células parentales: siete de los diez microRNAs más abundantes de las HCT116 WT, presentan un MAS por arriba del 50% (**Figura 17 A**), mientras que para las HCT116 Dicer KO, siete microRNAs presentan un MAS por debajo del 50% (**Figura 17 B**). La correlación entre los microRNAs compartidos entre las dos muestras es baja ($p = 0.3$), indicando que la producción de isomiRs de muchos microRNAs es dependiente de Dicer. Sin embargo, existen otros microRNAs cuyas proporciones de MAS se mantuvieron constantes, apoyando la idea de que el procesamiento por Dicer no es la única fuente de isomiRs (**Figura 18**).

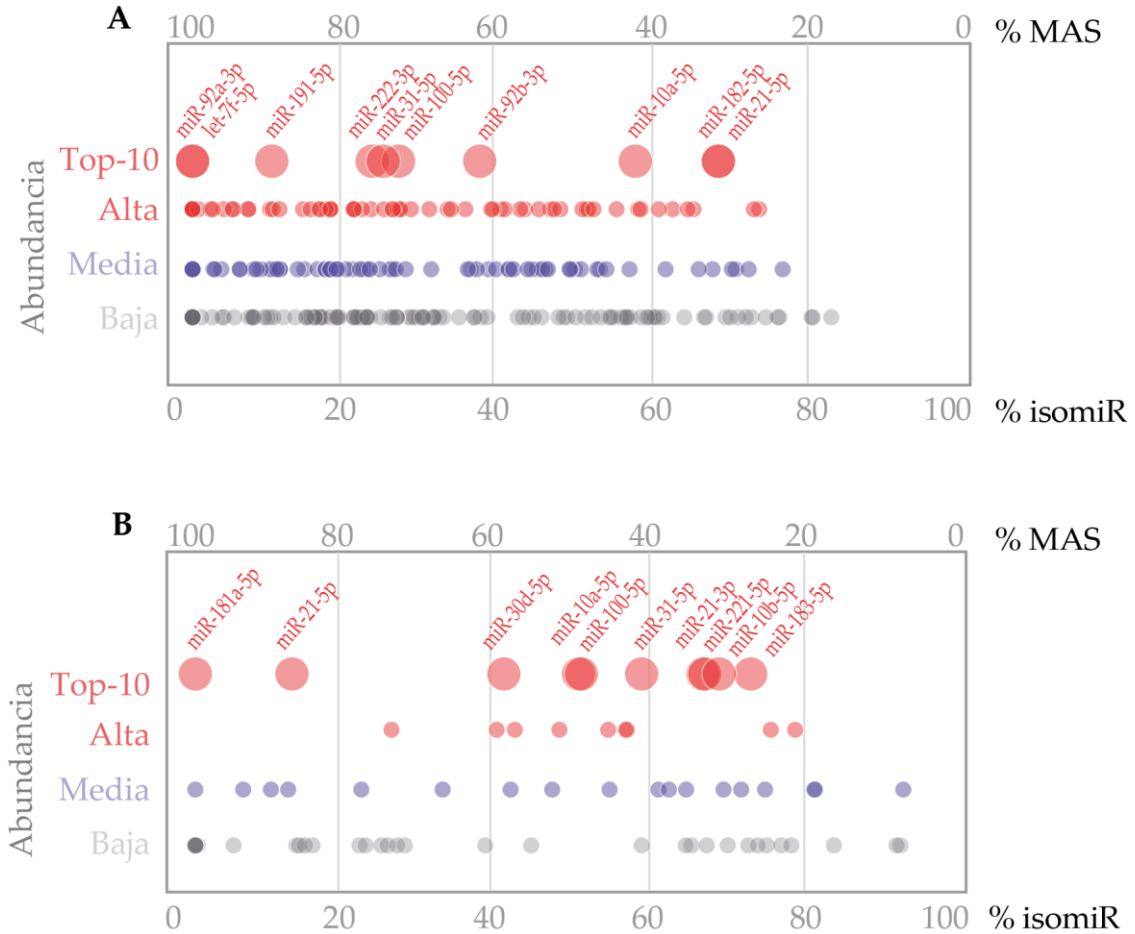


Figura 17. Porcentaje de MAS por cada microRNA en muestras de la línea celular HCT116 WT y HCT116 Dicer KO. La abundancia corresponde a la distribución de las lecturas por microRNA (RPmiR). Baja, 0–25% (puntos grises); media 26–75% puntos azules; alta 76–100% puntos rojos. Los círculos rojos más grandes representan los 10 microRNAs más abundantes contenidos en cada muestra. **A)** HCT116 WT. **B)** HCT116 Dicer KO.

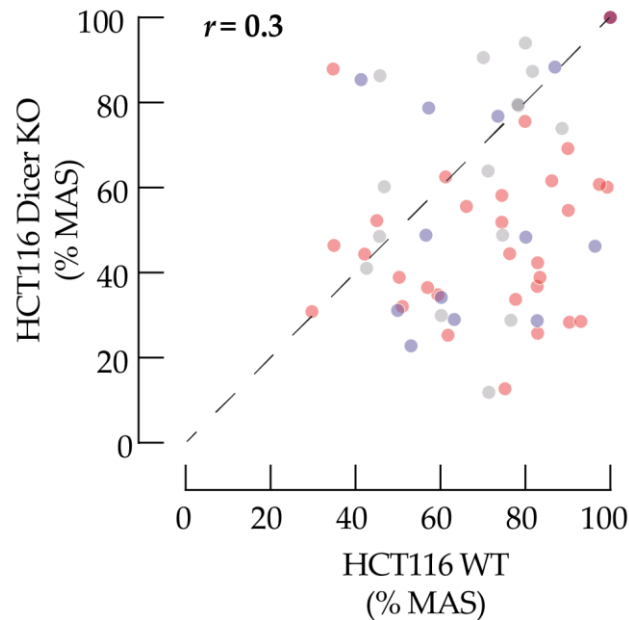


Figura 18. Correlación entre muestras de la línea celular HCT116 WT y HCT116 Dicer KO. Los colores corresponden a la abundancia en RPmiR de cada muestra. Baja 0–25% puntos grises; media 26–75% puntos azules; alta 76–100% puntos rojos. Solo se muestran los microRNAs y sus proporciones presentes en ambas muestras.

Al obtener el isomiR-index de las células Droscha KO y correlacionar su porcentaje de MAS con la línea parental, no se observa un efecto tan marcado en el sesgo de la distribución de puntos, ya que su correlación es relativamente alta en comparación con las células Dicer KO (**Figura 19 A**, $p = 0.8$), sugiriendo que Droscha tiene una participación más modesta en la generación de isomiRs. Es importante señalar que, al igual que las células carentes en Dicer, hay una marcada reducción en el número total de microRNAs que coinciden en ambas muestras, lo que indica que la biblioteca de Droscha KO tiene muy pocos microRNAs totales en comparación con la línea parental.

La XPO5 es una proteína cuya función dentro de la biogénesis de microRNAs se ha asociado al transporte de los mismos del núcleo al citoplasma, por lo que no se esperaría que su ausencia tuviera una influencia en las proporciones de MAS e isomiRs. Consecuentemente, la correlación de MAS entre la línea WT y las células XPO5 KO es alta (**Figura 19 B**, $p = 0.8$). Sin embargo, a diferencia de lo que ocurre con la correlación observada entre Droscha y su línea parental, el número total de microRNAs no baja tan drásticamente, lo cual sugiere que otro mecanismo de exportación de microRNAs puede estar supliendo al de la XPO5. Estudios recientes han reportado que además de la XPO5, la XPO 1 puede mediar el transporte de pre-microRNAs al citoplasma bajo ciertas condiciones (Martínez *et al.* 2017) por lo que es posible que en estas muestras ocurra un fenómeno similar.

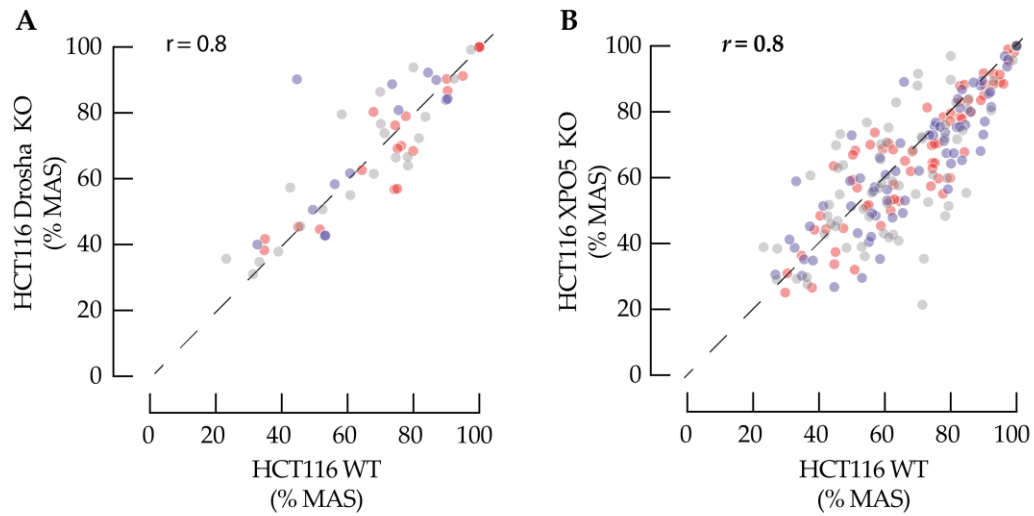


Figura 19. Correlación entre células parentales vs Drosha KO o XPO5 KO. **A)** células WT vs Drosha KO. **B)** Células WT vs XPO5 KO. Los colores corresponden a la abundancia en RPmiR de cada muestra. Baja 0–25% puntos grises; media 26–75% puntos azules; alta 76–100% puntos rojos. Solo se muestran los microRNAs y sus proporciones presentes en ambas muestras.

Discusión

El objetivo de este trabajo fue diseñar una manera de graficar cuantitativamente las proporciones de isomiRs que corresponden a un microRNA de manera global dentro de una muestra. Hasta la fecha, el estudio de los isomiRs había estado limitado a estudiar los microRNAs y sus isoformas uno por uno dentro de una sola representación gráfica por microRNA. Esto, aunque no es erróneo, es poco práctico, ya que dentro de una muestra pueden existir hasta 400 microRNAs diferentes, cada uno con sus propios isomiRs; sería altamente impráctico el realizar 400 graficas diferentes y compararlas –eso sin contar el número de las mismas si es que se tienen diferentes muestras experimentales.

El isomiR-index provee una forma sencilla de organizar las proporciones de isomiRs para cada microRNA dentro de una muestra. A través de la correlación del MAS obtenido, es particularmente fácil identificar condiciones experimentales o fisiológicas en las que las proporciones de isomiRs cambian para un grupo determinado de microRNAs. En este trabajo se puso a prueba la utilidad del isomiR-index al comparar muestras similares y disimilares en su contenido de isomiRs, de tal manera que la alta o baja correlación de los porcentajes de MAS fueron útiles para evaluar si las muestras tienen o no parecido. Un factor importante es que el isomiR-index y la correlación de MAS no equivalen meramente a la correlación de la expresión o abundancia de los microRNAs, ya que se demostró que su distribución es independiente, por lo que el isomiR-index puede usarse como un complemento al análisis de expresión de muestras, con el que se pueden identificar patrones que de otra manera no podrían explorarse.

Los resultados que muestran las correlaciones de MAS de las muestras de células parentales vs knock-out en Dicer y XPO5, correlacionan con lo esperado y conocido de la biogénesis de microRNAs. La ausencia de la RNasa Dicer tiene un impacto en la producción de isomiRs; la ausencia de XPO 5, al no ser esta una RNasa, no tiene un impacto apreciable en la producción de isomiRs al comparar con la línea parental. Un resultado inesperado es que la línea carente de Drosha tampoco tiene un impacto aparente en la producción de isomiRs. Esto puede deberse a que uno de los mecanismos naturales de generación de microRNAs es independiente de Drosha (**Figura 4A**). En esta vía, los microRNAs producidos no son procesados por Drosha, sino por el espliceosoma –que es un complejo multiproteínico encargado del *splicing* de RNAs primarios, por lo que cambios o variaciones en la precisión del *splicing* tendrían un impacto muy profundo en la

expresión de mRNAs (Okamura *et al.* 2007; Rubi *et al.* 2007), razón por la que el espliceosoma es una maquinaria altamente precisa. La interpretación de la alta correlación en la ausencia de Drosha es que los microRNAs que se observan en esa muestra son derivados del procesamiento llevado a cabo por el espliceosoma, y por lo tanto son miRtrons. El que presenten una alta correlación sería indicativo de la alta precisión que tiene el espliceosoma para procesar RNA en sitios específicos.

Por el contrario, el contenido de isomiRs aumenta en la línea carente de Dicer. Este resultado puede deberse a que el precursor recién exportado al citoplasma, para poder ensamblarse en RISC, requiere ser cortado para remover su asa del pre-microRNA. Al no existir Dicer, la única RNasa funcional restante dentro de la vía es Ago2, la cual podría ser la encargada del corte. Existe al menos un ejemplo en el que un microRNA, el miR-451, es procesado directamente por Ago2, sin pasar por Dicer. Ante la ausencia de Dicer, Ago2 es enfrentado con precursores de tallo y asa. Sin embargo, dado su mecanismo de corte y función, Ago2 no tiene la precisión de Dicer al procesar los pre-microRNAs y su corte es desordenado y variable, inclusive al procesar moléculas de RNA idénticas en secuencia y estructura, viéndose reflejado en el contenido de isomiRs.

El que la ausencia de la XPO5 no tenga una influencia aparente en el contenido de isomiRs con respecto a la línea parental, confirma su función como proteína exportadora, pero no procesadora. Es interesante, sin embargo, que un número importante de microRNAs sigue siendo exportado al citoplasma en ausencia de XPO5. Esto sugiere que otras proteínas, por ejemplo XPO1, puedan estar involucradas en el transporte del núcleo. Sin duda, este es un área que ha sido muy poco explorada.

Finalmente, es importante recalcar que la generación de isomiRs es un fenómeno común dentro del proceso de biogénesis y maduración de los microRNAs (Tan *et al.* 2015). Algunos estudios han encontrado que enzimas, como TUTasas, son capaces de regular el intercambio dinámico de los microRNAs, pues pueden añadir nucleótidos en el extremo 3' (en su mayoría As ó Us). También se conoce que los mRNAs blancos inducen la generación de isomiRs, añadiendo nucleótidos al extremo 3' de los microRNAs y funcionando así como marca molecular para su posterior decaimiento (Sanei *et al.* 2015). Por lo tanto, la generación de isomiRs ocurre en un amplio panorama de posibilidades que involucran desde el procesamiento y maduración del microRNA, hasta su decaimiento. Por lo tanto es importante expandir la visión que se tiene de su homeostasis entendiendo a los isomiRs como una consecuencia del fenómeno

biológico que los está originando, ya que pueden ofrecernos información acerca de la función y la regulación de los microRNAs. Dada la utilidad del isomiR-index como herramienta para el estudio y clasificación de cambios en las proporciones de isomiRs mostradas en este trabajo, es posible que en estudios futuros el contenido de isomiRs pueda arrojar hallazgos importantes en el campo de la investigación sobre la función y los mecanismos de acción de los microRNAs.

Conclusiones

En el presente trabajo planteamos una nueva estrategia para el estudio bioinformático global de isomiRs. Esta nueva estrategia permite obtener un índice porcentual de las proporciones de isomiRs y la secuencia más abundante de cada microRNA por biblioteca, por lo que se puede comparar la composición de un microRNA en distintas muestras.

Utilizando el isomiR-index, comparamos el contenido porcentual de isomiRs de diferentes bases de datos de tejidos y líneas celulares. Constatamos que tejidos parecidos pero de distintos individuos, comparten casi completamente la proporción de isomiRs y también pudimos corroborar que tejidos diferentes de humano y ratón tienen un contenido distinto de isomiRs.

El análisis global de isomiRs es diferente al análisis de abundancia que se realiza convencionalmente. Evaluar la proporción de isomiRs de cada microRNA por muestra permite distinguir y comparar la heterogeneidad de las secuencias que componen a los microRNA bajo diferentes condiciones.

Por último, utilizamos isomiR-index para determinar el efecto del KO de enzimas importantes para la maduración de los microRNAs en mamíferos. El análisis mostró que los microRNAs tienen mayor proporción de isomiRs cuando la RNAsa tipo III, Dicer, está ausente. Existe una disminución de la producción de microRNAs cuando Drosha se inhibe, mientras que con la ausencia de la XPO5 no muestra ninguna alteración, sugiriendo que esta enzima no juega un papel en la generación de la secuencia de un microRNA maduro.

Referencias

- Agarwal, V., Bell, G. W., Nam, J.-W., & Bartel, D. P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. *ELife*, 4.
- Ambros, V. (2004). The functions of animal microRNAs. *Nature*, 431(7006), 350–355.
- Ameres, S. L., & Zamore, P. D. (2013). Diversifying microRNA sequence and function. *Nature Reviews Molecular Cell Biology*, 14(8), 475–488.
- Amsel, D., Vilcinskas, A., & Billion, A. (2017). Evaluation of high-throughput isomiR identification tools: illuminating the early isomiRome of *Tribolium castaneum*. *BMC Bioinformatics*, 18(1), 359.
- Andrews S (2010) FastQC: A quality control tool for high throughput sequence data.
- Aryani, A., & Denecke, B. (2015). *In vitro* application of ribonucleases: comparison of the effects on mRNA and miRNA stability. *BMC Research Notes*, 8, 164.
- Bartel, D. P. (2018). Metazoan MicroRNAs. *Cell*, 173(1), 20–51
- Bartel D. P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell*, 136(2), 215–33.
- Boele, J., Persson, H., Shin, J. W., Ishizu, Y., Newie, I. S., Søkilde, R., de Hoon, M. J. L. (2014). PAPD5-mediated 3' adenylation and subsequent degradation of miR-21 is disrupted in proliferative disease. *Proceedings of the National Academy of Sciences of the United States of America*, 111(31), 11467–11472.
- Burns, D. M., D'Ambrogio, A., Nottrott, S. & Richter, J. D. (2011) CPEB and two poly(A) polymerases control miR-122 stability and p53 mRNA translation. *Nature* 473, 105–108.
- Cai, X., Hagedorn, C. H. & Cullen, B. R. (2004). Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA* 10, 1957–1966.
- Catalanotto, C., Cogoni, C., & Zardo, G. (2016). MicroRNA in Control of Gene Expression: An Overview of Nuclear Functions. *International Journal of Molecular Sciences*, 17(10).
- Chitwood, D. H., & Timmermans, M. C. P. (2007). Target mimics modulate miRNAs. *Nature Genetics*, 39(8), 935–936.

- de Rie, D., Abugessaisa, I., Alam, T., Arner, E., Arner, P., Ashoor, H., ... de Hoon, M. J. L. (2017). An integrated expression atlas of miRNAs and their promoters in human and mouse. *Nature Biotechnology* 35 (9).
- Ebert, M. S., Neilson, J. R., & Sharp, P. A. (2007). MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. *Nature Methods*, 4(9), 721–726.
- ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414), 57–74.
- Fernandez-Valverde, S. L., Taft, R. J., & Mattick, J. S. (2010). Dynamic isomiR regulation in *Drosophila* development. *RNA*, 16(10), 1881–1888.
- Flores-Jasso, C. F., Arenas-huertero, C., Reyes, J. L., Contreras-cubas, C., Covarrubias, A., & Vaca, L. (2009). First step in pre-miRNAs processing by human Dicer. *Acta Pharmacologica Sinica*, 30(8), 1177–1185.
- 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.
- Gebert D, Hewel C, Rosenkranz D (2017). unitas: the universal tool for annotation of small RNAs. *BMC Genomics* 18.
- Gebert, L. F. R., & MacRae, I. J. (2019). Regulation of microRNA function in animals. *Nature Reviews Molecular Cell Biology*, 20(1), 21–37.
- Guo, L., & Chen, F. (2014). A challenge for miRNA: multiple isomiRs in miRNAomics. *Gene*, 544(1), 1–7.
- Guo, L., Liang, T., Yu, J., & Zou, Q. (2016). A Comprehensive Analysis of miRNA/isomiR Expression with Gender Difference. *PLOS ONE*, 11(5), e0154955.
- Ha, M., & Kim, V. N. (2014). Regulation of microRNA biogenesis. *Nature Reviews Molecular Cell Biology*, 15(8), 509–524.
- Havens, M. A., Reich, A. A., Duelli, D. M., & Hastings, M. L. (2012). Biogenesis of mammalian microRNAs by a non-canonical processing pathway. *Nucleic acids research*, 40(10), 4626–40.
- Höck, J., & Meister, G. (2008). The Argonaute protein family. *Genome Biology*, 9(2), 210.
- Hutvagner, G., McLachlan, J., Pasquinelli, A. E., Bálint, E., Tuschl, T., & Zamore, P. D. (2001). A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science (New York, N.Y.)*

- Janssen, H. L. A., Reesink, H. W., Lawitz, E. J., Zeuzem, S., Rodriguez-Torres, M., Patel, K., Hodges, M. R. (2013). Treatment of HCV Infection by Targeting MicroRNA. *New England Journal of Medicine*, 368(18), 1685–1694.
- Jin, F., Hu, H., Xu, M., Zhan, S., Wang, Y., Zhang, H., & Chen, X. (2018). Serum microRNA Profiles Serve as Novel Biomarkers for Autoimmune Diseases. *Frontiers in Immunology*, 9, 2381.
- Jones, M. R. et al. (2009) Zcchc11-dependent uridylation of microRNA directs cytokine expression. *Nat. Cell Biol.* 11, 1157–1163.
- Katoh, T. et al. (2009) Selective stabilization of mammalian microRNAs by 3' adenylation mediated by the cytoplasmic poly(A) polymerase GLD-2. *Genes Dev.* 23, 433–438.
- Ketting, R. F., Fischer, S. E., Bernstein, E., Sijen, T., Hannon, G. J., & Plasterk, R. H. (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes & development*, 15(20), 2654–2659.
- Kim, Y.-K., Kim, B., & Kim, V. N. (2016). Re-evaluation of the roles of DROSHA, Exportin 5, and DICER in microRNA biogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, 113(13), E1881-9.
- Kozomara, A., Birgaoanu, M., & Griffiths-Jones, S. (2019). miRBase: from microRNA sequences to function. *Nucleic Acids Research*, 47(D1), D155–D162.
- Kurihara, Y. & Watanabe, Y. (2004) *Arabidopsis* micro-RNA biogenesis through Dicer-like 1 protein functions. *Proc. Natl Acad. Sci. USA* 101, 12753–12758.
- Kwon, S. C., Nguyen, T. A., Choi, Y.-G., Jo, M. H., Hohng, S., Kim, V. N., & Woo, J.-S. (2016). Structure of Human DROSHA. *Cell*, 164(1–2), 81–90.
- Lagos-Quintana, M., Rauhut, R., Lendeckel, W., & Tuschl, T. (2001). Identification of Novel Genes Coding for Small Expressed RNAs. *Science*, 294(5543), 853–858.
- Lan, C., Peng, H., McGowan, E. M., Hutvagner, G., & Li, J. (2018). An isomiR expression panel based novel breast cancer classification approach using improved mutual information. *BMC Medical Genomics*, 11(Suppl 6), 118.
- Lares, M. R., Rossi, J. J., & Ouellet, D. L. (2010). RNAi and small interfering RNAs in human disease therapeutic applications. *Trends in Biotechnology*, 28(11), 570–579.

Lau, N. C., Lim, L. P., Weinstein, E. G., & Bartel, D. P. (2001). An Abundant Class of Tiny RNAs with Probable Regulatory Roles in *Caenorhabditis elegans*. *Science*, 294.

Lee, C.-T., Risom, T., & Strauss, W. M. (2007). Evolutionary Conservation of MicroRNA Regulatory Circuits: An Examination of MicroRNA Gene Complexity and Conserved MicroRNA-Target Interactions through Metazoan Phylogeny. *DNA and Cell Biology*, 26(4).

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(5), 843–854.

Lee, R. C., & Ambros, V. (2001). An Extensive Class of Small RNAs in *Caenorhabditis elegans*. *Science*, 294(5543), 862–864.

Lee, Y., Jeon, K., Lee, J.-T., Kim, S., & Kim, V. N. (2002). MicroRNA maturation: stepwise processing and subcellular localization. *The EMBO Journal*, 21(17), 4663–4670.

Leung, A. K. L. (2015). The Whereabouts of microRNA Actions: Cytoplasm and Beyond. *Trends in Cell Biology*, 25(10), 601–610.

Lewis, B. P., Burge, C. B., & Bartel, D. P. (2005). Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets. *Cell*, 120(1), 15–20.

Li, S., & Patel, D. J. (2016). Drosha and Dicer: Slicers cut from the same cloth. *Cell Research*, 26(5), 511–512.

Ludwig Nicole, Petra Leidinger, Kurt Becker, Christina Backes, Tobias Fehlmann, Christian Pallasch, Steffi Rheinheimer, Benjamin Meder, Cord Stähler, Eckart Meese, Andreas Keller (2016) Distribution of miRNA expression across human tissues, *Nucleic Acids Research*, Volume 44, Issue 8, Pages 3865–3877

Lukasik, A., Wójcikowski, M., & Zielenkiewicz, P. (2016). Tools4miRs - one place to gather all the tools for miRNA analysis. *Bioinformatics (Oxford, England)*, 32(17), 2722–2724.

Macrae, I. J., Zhou, K., Li, F., Repic, A., Brooks, A. N., Cande, W. Z., Doudna, J. A. (2006). Structural basis for double-stranded RNA processing by Dicer. *Science (New York, N.Y.)*, 311(5758), 195–198.

Martin, M. (2011) Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads. *EMBnet Journal*, 17, 10-12.

- Martinez, I., Hayes, K. E., Barr, J. A., Harold, A. D., Xie, M., Bukhari, S. I. A., DiMaio, D. (2017). An Exportin-1-dependent microRNA biogenesis pathway during human cell quiescence. *Proceedings of the National Academy of Sciences of the United States of America*, 114(25).
- Morin, R. D., O'Connor, M. D., Griffith, M., Kuchenbauer, F., Delaney, A., Prabhu, A.-L., Marra, M. A. (2008). Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome Research*, 18(4).
- Neilsen, C. T., Goodall, G. J., & Bracken, C. P. (2012). IsomiRs - the overlooked repertoire in the dynamic microRNAome. *Trends in Genetics*, 28(11), 544–549.
- Nientiedt, M., Deng, M., Schmidt, D., Perner, S., Müller, S. C., & Ellinger, J. (2016). Identification of aberrant tRNA-halves expression patterns in clear cell renal cell carcinoma. *Scientific reports*, 6.
- Nguyen, T. A., Jo, M. H., Choi, Y.-G., Park, J., Kwon, S. C., Hohng, S., Woo, J.-S. (2015). Functional Anatomy of the Human Microprocessor. *Cell*, 161(6), 1374–1387.
- O'Brien, J., Hayder, H., Zayed, Y., & Peng, C. (2018). Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. *Frontiers in Endocrinology*, 9, 402.
- Okamura, K., Hagen, J. W., Duan, H., Tyler, D. M., & Lai, E. C. (2007). The mirtron pathway generates microRNA-class regulatory RNAs in *Drosophila*. *Cell*, 130(1), 89–100.
- Park, J.-E., Heo, I., Tian, Y., Simanshu, D. K., Chang, H., Jee, D., Kim, V. N. (2011). Dicer recognizes the 5' end of RNA for efficient and accurate processing. *Nature*, 475(7355), 201–205.
- Pillman, K. A., Goodall, G. J., Bracken, C. P., & Gantier, M. P. (2019). miRNA length variation during macrophage stimulation confounds the interpretation of results: implications for miRNA quantification by RT-qPCR. *RNA (New York, N.Y.)*, 25(2), 232–238.
- Pillai, R. S., Artus, C. G., & Filipowicz, W. (2004). Tethering of human Ago proteins to mRNA mimics the miRNA-mediated repression of protein synthesis. *RNA (New York, N.Y.)*, 10(10), 1518–1525.
- Quick-Cleveland, J., Jacob, J. P., Weitz, S. H., Shoffner, G., Senturia, R., & Guo, F. (2014). The DGCR8 RNA-Binding Heme Domain Recognizes Primary MicroRNAs by Clamping the Hairpin. *Cell Reports*, 7(6), 1994–2005.

- Roberts, T. C. (2014). The MicroRNA Biology of the Mammalian Nucleus. *Molecular Therapy - Nucleic Acids*, 3, e188.
- Roth, B. M., Ishimaru, D., & Hennig, M. (2013). The Core Microprocessor Component DiGeorge Syndrome Critical Region 8 (DGCR8) Is a Nonspecific RNA-binding Protein. *Journal of Biological Chemistry*, 288(37), 26785–26799.
- Ruby, J. G., Jan, C. H., & Bartel, D. P. (2007). Intronic microRNA precursors that bypass Drosha processing. *Nature*, 448(7149), 83–86.
- Sanei, M., & Chen, X. (2015). Mechanisms of microRNA turnover. *Current Opinion in Plant Biology*, 27, 199–206.
- Schirle, N. T., Sheu-Gruttadauria, J., & MacRae, I. J. (2014). Structural basis for microRNA targeting. *Science* 346, 608–613.
- Schwab, R., Palatnik, J. F., Riester, M., Schommer, C., Schmid, M., & Weigel, D. (2005). Specific Effects of MicroRNAs on the Plant Transcriptome. *Developmental Cell*, 8(4), 517–527.
- Seitz, H. (2009). Redefining microRNA targets. *Current Biology: CB*, 19(10), 870–873.
- Starega-Roslan, J., Witkos, T. M., Galka-Marciniak, P., & Krzyzosiak, W. J. (2015). Sequence features of Drosha and Dicer cleavage sites affect the complexity of isomiRs. *International Journal of Molecular Sciences*, 16(4), 8110–8127.
- Tan, G. C., & Dibb, N. (2015). IsomiRs have functional importance. *The Malaysian Journal of Pathology*, 37(2), 73–81.
- Telonis, A. G., Magee, R., Loher, P., Chervoneva, I., Londin, E., & Rigoutsos, I. (2017). Knowledge about the presence or absence of miRNA isoforms (isomiRs) can successfully discriminate amongst 32 TCGA cancer types. *Nucleic Acids Research*, 45(6), 2973–2985.
- Tomari, Y., Matranga, C., Haley, B., Martinez, N., & Zamore, P. D. (2004). A protein sensor for siRNA asymmetry. *Science (New York, N.Y.)*, 306(5700), 1377–1380.
- Ventura, A., Young, A. G., Winslow, M. M., Lintault, L., Meissner, A., Erkeland, S. J., Newman, J., Bronson, R. T., Crowley, D., Stone, J. R., Jaenisch, R., Sharp, P. A., Jacks, T. (2008). Targeted deletion reveals essential and overlapping functions of the miR-17 through 92 family of miRNA clusters. *Cell*, 132(5), 875–86.
- Wang, S., Zheng, Z., Chen, P., & Wu, M. (2019). Tumor classification and biomarker discovery based on the 5'isomiR expression level. *BMC Cancer*, 19(1), 127.

Wee, L. M., Flores-Jasso, C. F., Salomon, W. E., & Zamore, P. D. (2012). Argonaute divides its RNA guide into domains with distinct functions and RNA-binding properties. *Cell*, 151(5), 1055–1067.

Yi, R., Qin, Y., Macara, I. G., & Cullen, B. R. (2003). Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes & Development*, 17(24), 3011–3016.

Zamudio, J. R., Kelly, T. J., & Sharp, P. A. (2014). Argonaute-bound small RNAs from promoter-proximal RNA polymerase II. *Cell*, 156(5), 920–934.

Zhang, H., Kolb, F. A., Jaskiewicz, L., Westhof, E., & Filipowicz, W. (2004). Single Processing Center Models for Human Dicer and Bacterial RNase III. *Cell*, 118(1), 57–68.

Zhang, Y., Zang, Q., Xu, B., Zheng, W., Ban, R., Zhang, H., Shi, Q. (2016). IsomiR Bank: a research resource for tracking IsomiRs. *Bioinformatics*, 32(13), 2069–2071.