

Tesis Doctoral

Estudios de microARNs involucrados en la diferenciación cardíaca a partir de células pluripotentes humanas

Garate, Ximena

2018

Este documento forma parte de las colecciones digitales de la Biblioteca Central Dr. Luis Federico Leloir, disponible en bibliotecadigital.exactas.uba.ar. Su utilización debe ser acompañada por la cita bibliográfica con reconocimiento de la fuente.

This document is part of the digital collection of the Central Library Dr. Luis Federico Leloir, available in bibliotecadigital.exactas.uba.ar. It should be used accompanied by the corresponding citation acknowledging the source.

Cita tipo APA:

Garate, Ximena. (2018). Estudios de microARNs involucrados en la diferenciación cardíaca a partir de células pluripotentes humanas. Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires. https://hdl.handle.net/20.500.12110/tesis_n6396_Garate

Cita tipo Chicago:

Garate, Ximena. "Estudios de microARNs involucrados en la diferenciación cardíaca a partir de células pluripotentes humanas". Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires. 2018. https://hdl.handle.net/20.500.12110/tesis_n6396_Garate

EXACTAS UBA

Facultad de Ciencias Exactas y Naturales



UBA

Universidad de Buenos Aires



UNIVERSIDAD DE BUENOS AIRES

FACULTAD DE CIENCIAS EXACTAS Y NATURALES

DEPARTAMENTO DE QUÍMICA BIOLÓGICA

Estudio de los microARNs involucrados en la diferenciación cardíaca a partir de células pluripotentes humanas

Tesis presentada para optar al título de Doctor de la Universidad de
Buenos Aires en el área de Química Biológica

Ximena Garate

Director de tesis: Dr. Santiago Miriuka

Consejero de estudios: Dra. Silvia Rossi

Lugar de trabajo: Laboratorio de Biología Celular y Desarrollo. LIAN-FLENI-CONICET

Buenos Aires, 2018

Resumen

Estudio de los microARNs involucrados en la diferenciación cardíaca a partir de células pluripotentes humanas

Los microARNs son reguladores post-transcripcionales de la expresión génica que participan en la regulación de distintos procesos celulares como por ejemplo en el desarrollo embrionario. En el presente trabajo estudiamos el perfil de expresión de los microARNs durante la diferenciación a mesodermo temprano y al linaje cardíaco mediante secuenciación masiva de los ARN pequeños. A partir de estos resultados determinamos que hay aproximadamente 700 microARNs que se expresan diferencialmente en las células pluripotentes humanas, un progenitor temprano de mesodermo y cardiomiocitos. A su vez, analizamos la presencia de variantes de microARNs llamados isomiRs y determinamos la abundancia de las distintas isoformas en las distintas poblaciones celulares. A partir del análisis *in silico* de los genes *target* determinamos que estas isoformas amplían el repertorio de genes regulados por el microARN canónico, lo cual sugiere un rol funcional de los isomiRs durante el proceso de diferenciación celular. Por otro lado, analizamos el miRNoma de cada una de las tres poblaciones celulares en base a las familias y *clusters* de microARNs e identificamos que los microARNs de dichos grupos comparten un perfil de expresión y colaboran entre sí en la regulación de procesos relacionados con el desarrollo embrionario. Además, establecimos el miRNoma de los cardiomiocitos mediante la comparación de los microARNs que se expresan en otros tipos celulares. Finalmente, estudiamos el *cluster* C19MC, *up*-regulado en las células pluripotentes, y analizamos de manera *in silico* y experimental los genes *target* de dicho *cluster*. De esta manera, determinamos que el C19MC participa en la regulación del ciclo celular y de distintos procesos de diferenciación en el desarrollo embrionario. Por último, estudiamos el miR-205 en la diferenciación a mesodermo e identificamos que su sobreexpresión aumenta al doble el porcentaje del progenitor temprano de mesodermo. Los microARNs ejercen su función formando complejas redes de regulación, lo cual requiere de un análisis más global para poder entender mejor el rol que estos pequeños ARN tienen en las células. Nuestros resultados expanden el conocimiento de los microARNs en relación con el desarrollo cardíaco.

Palabras Clave: Células Pluripotentes Humanas; Diferenciación Cardíaca; Progenitor de Mesodermo Temprano; Cardiomiocitos; MicroARNs; IsomiRs; Cluster C19MC.

Abstract

Study of the microRNAs involved in the cardiac differentiation from human pluripotent stem cells

MicroRNAs are small non-coding RNAs involved in post-transcriptional regulation of gene expression related to many cellular functions, including human embryonic development. In this work we study the microRNAs involved in the cardiac differentiation by small RNA high-throughput sequencing of three specific cell populations: Pluripotent stem cells, an early mesoderm progenitor and cardiomyocytes. We identified 700 microRNAs involved in the differentiation process where some of them are clustered according to their expression level. We also studied the expression level of isomiRs and determined their abundance and predicted targets by an *in silico* analysis. These results suggest that these isoforms might have an important role that should be considered when studying the microRNAs involved the differentiation process. Moreover, we identified the microRNAs families and clusters involved in the establishment of cardiac lineage and define the mirRNome based on these groups. Furthermore, we were able to determine a more accurate miRNome associated with cardiomyocytes by comparing the expressed microRNAs with other mature cells. On the other hand, we studied the microRNA cluster C19MC which is highly expressed in pluripotent stem cells. Based on the predicted targets, as well as experimental targets and functional analysis, we determined that this cluster is involved in the pluripotent stem cells in the regulation of the cell cycle and in the regulation of different differentiation processes during early development. Finally, we studied the role of the miR-205 during differentiation towards mesoderm and determined that its overexpression increase twice the percentage of the early mesoderm progenitor population.

MicroRNAs exert their effect in a complex and interconnected way, making necessary a global analysis to better understand their role. Our data expands the knowledge of microRNAs and their implications in cardiomyogenesis.

Key words: Human pluripotent stem cells; Cardiac differentiation; Mesoderm Progenitor Cells; Cardiomyocytes; microRNAs; isomiRs; cluster C19MC.

Agradecimientos

Detrás de este trabajo de tesis hubo mucha gente que me acompañó y colaboró de una u otra manera para que yo pueda cumplir con los objetivos.

Sin duda quiero agradecerle a Santiago por haber confiado en mí y haberme guiado a lo largo de todo el doctorado con sus consejos y enseñanzas.

A la familia del LIAN-FLENI que cada vez somos más y sigue creciendo el grupo lindo de gente que hace que el ambiente de trabajo sea tan agradable. A los que estuvieron desde el inicio: a Gabi, mi compinche, con quien desde el principio estuvimos trabajando codo a codo para ir superando las distintas etapas, a Charly por tantos mates compartidos, consejos y palabras de aliento. A Darío, Willy, Leo y Marie, por estar siempre dispuestos a dar un consejo. Y a los que fueron llegando con el tiempo: Sole, Nati, Lu, Vero, Ari, Cyn, Agus, Ale, Alan, Anto, Celes, Sofi, Oli, Guada y Olivier. Porque detrás de cada experimento siempre hubo alguno de ustedes dándome una mano. Gracias Sole por las infinitas charlas, consejos e incondicional aguante.

Al grupo de Ale Guberman, por la buena onda y permitirme usar su espacio de trabajo todas las veces que lo necesité.

Al Dr. Mercola y su grupo de investigación en la Universidad de Stanford, quienes me abrieron las puertas y tan generosamente me permitieron trabajar unos meses con ellos.

A mi familia, por estar siempre tan presente, ayudarme y apoyarme en todos los proyectos que emprendo. Son un pilar fundamental en mi vida.

Especialmente a Vic, por su inmenso cariño, comprensión, paciencia y apoyo en todos estos años.

A mis amigas de la vida Gaby, Chula, Vicky, Celes, Estefi, Mel, Cheve, Guly, Maru, Flor, Caro y Ro. Gracias chicas por estar ahí siempre y acompañarme en todas las etapas de mi vida desde hace tanto años.

A mis amigos de la facu, los incondicionales que siempre están para compartir lindos momentos y charlas de aguante porque entienden perfectamente los momentos difíciles de esta profesión, Chi, Juan, Mati, Estefi, Euge, Dani, Sil, Osva y Puche.

Sin el apoyo, la contención y el cariño de mis seres queridos no hubiese podido alcanzar la meta que me propuse hace cinco años. A todos ellos estoy eternamente agradecida.

“Sólo aquellos que intentan lo absurdo pueden lograr lo imposible”

A. Einstein

Índice

Resumen	I
Abstract	II
Agradecimientos	III
Índice de figuras	XI
Índice de tablas	XII
Abreviaturas	XIII
1. Introducción	1
1.1. Células pluripotentes humanas: Características y definiciones	1
1.2. Diferenciación celular	4
1.2.1. Desarrollo embrionario del corazón	5
1.2.2. Formación del mesodermo: Transición epitelio-mesenquimal	7
1.2.3. Diferenciación <i>in vitro</i> al linaje cardíaco	9
1.2.4. Potencial uso de las células madre pluripotentes en la medicina	11
1.3. MicroARNs: pequeños reguladores génicos	13
1.3.1. Biogénesis y procesamiento de los microARNs	15
1.3.2. MicroARNs y la regulación de la expresión génica	16
1.3.3. Estudio de los genes blanco: análisis <i>in silico</i> y experimental	18
1.3.4. Clasificación de los microARNs en familias y clusters	21
1.3.5. IsomiRs: Variantes de los microARNs	23
1.4. Rol de los microARNs en la pluripotencia y la diferenciación celular	26
2. Hipótesis y Objetivos	32
3. Resultados	35
Capítulo 1: Diferenciación a mesodermo y al linaje cardíaco	35
1.1. Diferenciación al progenitor de mesodermo temprano	35
1.2. Diferenciación al linaje cardíaco	37

Capítulo 2: Secuenciación masiva de los ARN pequeños en la diferenciación cardíaca	46
2.1. Análisis del perfil de expresión de los microARNs en las células pluripotentes, el progenitor de mesodermo y los cardiomiocitos	46
2.2. Análisis de los isomiRs	53
2.3. Análisis de las familias y clusters de microARNs en la diferenciación cardíaca . .	61
2.4. MicroARNs asociados con el linaje cardíaco	71
Capítulo 3: Estudio del <i>cluster</i> C19MC y su rol en las células pluripotentes humanas	77
3.1. Análisis de la expresión del <i>cluster</i> C19MC en la diferenciación celular	78
3.2. Análisis <i>in silico</i> del <i>cluster</i> C19MC y sus genes blanco	80
3.3. Estudio del rol del <i>cluster</i> C19MC en las células pluripotentes	88
Capítulo 4: Estudio del miR-205 en la diferenciación a mesodermo temprano	112
4. Discusión	118
4.1. Diferenciación hacia un progenitor temprano de mesodermo y hacia el linaje cardíaco	118
4.2. Identificación del miRNoma relacionado con la diferenciación al progenitor de mesodermo temprano y al linaje cardíaco	120
4.3. Análisis del <i>cluster</i> C19MC en las células pluripotentes	125
4.4. Estudio del miR-205 en la diferenciación al progenitor de mesodermo temprano	128
5. Conclusiones generales	130
6. Materiales y Métodos	133
6.1. Células utilizadas	133
6.2. Co-Cultivo de CMEh sobre MEFi	133
6.3. Derivación de MEF	133
6.4. Cultivo y criopreservación de MEF	134
6.5. Cultivo de CMEh libre de MEFi	135
6.6. Inducción del progenitor temprano de mesodermo	135
6.7. Diferenciación cardíaca mediante la formación de EB	136
6.8. Diferenciación cardíaca dirigida en monocapa	136
6.9. Inmunomarcaciones y microscopía de fluorescencia	137

6.10. Citometría de Flujo	138
6.11. Extracción de ADN genómico	139
6.12. Extracción de ARN total	139
6.13. Extracción de pequeños ARN para secuenciación	140
6.14. Retrotranscripción de ARNm y microARNs maduros	140
6.15. PCR a tiempo final	141
6.16. PCR cuantitativa	142
6.17. Diseño de <i>pimers</i> para analizar los microARNs maduros	144
6.18. Diseño y clonado de los gRNA para CRISPR/Cas9	145
6.19. Preparación de plásmidos - Transformación de bacterias competentes	145
6.20. Preparación de plásmidos a gran escala (Maxi y midi-prep)	146
6.21. Transfección con pequeñas moléculas comerciales de ARN	146
6.22. Transfección de las CMEh con los gRNA para CRISPR/Cas9	148
6.23. Ensayo de incorporación de bromodeoxiuridina	148
6.24. Ensayo de apoptosis con Anexina V	149
6.25. Ensayo de ciclo celular con ioduro de propidio (IP)	149
6.26. Ensayo de co-inmunoprecipitación de Argonauta (AgoIP)	150
6.27. Secuenciación masiva de ARN (<i>RNA-Seq</i>)	151
6.28. Análisis bioinformático de los resultados de <i>RNA-Seq</i>	152
6.29. Análisis estadísticos	153

7. Bibliografía

155

Índice de figuras

I1.	Fases de la segmentación celular	2
I2.	Desarrollo embrionario del corazón	5
I3.	Mecanismos moleculares que regulan la formación del corazón	7
I4.	Etapas de la diferenciación cardíaca	10
I5.	Biogénesis y procesamiento de los microARNs	16
I6.	Reconocimiento de los ARNm blanco	17
I7.	Identificación experimental de los ARNm blanco	20
I8.	Cluster miR-17~92 y la regulación de genes relacionados entre sí	22
I9.	MicroARNs en la diferenciación celular	28
1.	Colonias de células madre embrionarias humanas	36
2.	Diferenciación a mesodermo	37
3.	Análisis de la expresión de genes en el progenitor de mesodermo	38
4.	Diferenciación dirigida al linaje cardíaco	39
5.	Niveles de expresión de los genes a lo largo de la diferenciación cardíaca	40
6.	Inmunofluorescencia de genes cardíacos	41
7.	Análisis de la expresión de los microARNs durante la diferenciación cardíaca	42
8.	Selección de células cardíacas	43
9.	Diferenciación cardíaca	44
10.	Análisis de los resultados de secuenciación masiva mediante el CAP-miRSeq pipeline	47
11.	Abundancia de los distintos tipos de ARN	47
12.	Niveles de expresión de los microARNs en las tres poblaciones celulares	49
13.	Análisis diferencial de la expresión de los microARNs	50
14.	Heatmap de distancia y PCA plot	52
15.	Abundancia y heterogeneidad de los isomiRs en la diferenciación cardíaca	54
16.	Análisis de los isomiRs: Posición y abundancia de las modificaciones	56
17.	Análisis del miR-302a-5p y el miR-302a-3p y sus respectivos isomiRs	57
18.	Análisis de Gene Ontology del miR-302a-5p e isomiRs	58
19.	Análisis de Gene Ontology del miR-302a-3p e isomiRs	59
20.	Análisis de los isomiRs con ediciones A-I	60

21. Perfiles de expresión de los <i>clusters</i> de microARNs y su relación con las familias de microARNs	62
22. Análisis de expresión de las familias y <i>clusters</i> de microARNs	63
23. Expresión de los microARNs en familias y <i>clusters</i> en la diferenciación cardíaca	64
24. Análisis de los genes <i>target</i> de la familia miR-302	66
25. Análisis de los genes <i>target</i> de la familia/cluster miR-17/92	68
26. Análisis de los genes <i>target</i> de los microARNs cardíacos	69
27. Análisis de los genes blanco de los microARNs individuales	70
28. Expresión de los microARNs en los CM en relación con otros tipos de células	72
29. Análisis de la expresión diferencial de los microARNs en los CM con respecto a los otros tipos celulares	73
30. Locus del <i>cluster</i> C19MC	78
31. Niveles de expresión de los microARNs del C19MC	79
32. Validación de la expresión de los microARNs del C19MC	80
33. Análisis de marcas de transcripción activa/inactiva en las regiones próximas al C19MC	81
34. Alineamiento múltiple de las secuencias de los microARNs del <i>cluster</i> C19MC	83
35. El <i>cluster</i> C19MC y la red de sus genes blanco predichos	84
36. Correlación de la afinidad de los microARN del C19MC por sus genes blanco	85
37. Análisis de <i>Gene Ontology</i> de los genes blanco predichos para los microARNs del C19MC	86
38. Comparación entre los cluster C19MC, miR-302 y miR-371-373	87
39. Términos de Gene Ontology asociados con los genes blanco en común entre los tres clusters	88
40. Esquema del protocolo de Ago-IP-Seq	90
41. Análisis de los niveles de expresión de los microARNs del C19MC	91
42. Sobreexpresión del miR-520a en las células pluripotentes inducidas y cardiomiocitos	91
43. Esquema del <i>pipeline</i> del análisis de los resultados de AgoIP-Seq	93
44. Control de la calidad de los resultados de secuenciación	94
45. Términos de Gene Ontology asociados con los genes blanco del miR-520a en las PSC	96

46. Términos de Gene Ontology asociados con los genes blanco del miR-520a en CM	98
47. Comparación de los genes blanco del miR-520a identificados <i>in silico</i> y experimentalmente	101
48. Ensayo de apoptosis al inhibir la subfamilia del miR-520	103
49. Análisis del perfil del ciclo celular al inhibir la subfamilia miR-520	104
50. Aplicaciones del sistema de edición génica CRISPR/Cas9	105
51. Esquema de la estrategia CRISPR/Cas9 para eliminar el C19MC.	106
52. Análisis de la delección del <i>cluster</i> C19MC	107
53. Análisis de la expresión de los microARNs del C19MC en las CMPi <i>C19MC-</i> . . .	108
54. Análisis del perfil del ciclo celular del clon CMPi C19MC-	109
55. Análisis del perfil de expresión de la familia del miR-200 y el miR-205 en la diferenciación al progenitor temprano de mesodermo	113
56. Sobreexpresión del miR-205 mediante la transfección del mimic miR-205	114
57. Modulación de la expresión del miR-205 en la diferenciación a mesodermo temprano	115
58. Análisis de la expresión de genes de indiferenciación y marcadores de la TEM al sobreexpresar el miR-205	115

Índice de tablas

1.	microARNs diferencialmente más expresados en PSC con respecto a CM	51
2.	microARNs diferencialmente más expresados en CM con respecto a PSC	51
3.	miRNoma de los cardiomiocitos	74
4.	Lista de genes <i>target</i> del miR-520a en las células pluripotentes	95
5.	Términos de Gene Ontology relacionados con los genes blanco del miR-520a en las PSC	97
6.	Términos de Gene Ontology relacionados con los genes blanco del miR-520a en los CM	99
7.	Anticuerpos primarios y secundarios utilizados en inmunofluorescencia	138
8.	Secuencias de los primers utilizados en la PCR cuantitativa en sentido 5'-3' . . .	143
9.	Secuencias de los primers utilizados para los microARNs en la RT-qPCR	144
10.	Secuencia de los gRNA diseñados para la edición génica del <i>cluster</i> C19MC . . .	145
11.	Nombre y número de catálogo de los oligos sintéticas de los microARNs	147

Abreviaturas

ADAR: Adenosine Deaminase Acting on RNA
ADN: Ácido desoxiribonucleico
Ago: Argonauta
Ago-IP: Argonauta Inmunoprecipitación
ARN: Ácido ribonucleico
ARNm: ARN mensajero
bFGF : basic fibroblast growth factor
BMP-4: Bone Morphogenetic Protein 4
BrdU: Bromodesoxiuridina
BSA: Bovine Serum Albumin
C19MC: Chromosome 19 microRNA Cluster
CDS: Coding DNA sequence
ChIP: Chromatin Immunoprecipitation
CM: Cardiomyocytes
CMEh : Células madre embrionarias humanas
CMPi : Células madre pluripotentes inducidas
CRISPR: Clustered regularly interspaced short palindromic repeats
cTnT : Cardiotroponina T
dNTPS: desoxi-ribonucleótidos-trifosfato (dATP, dCTP, dGTP, dTTP)
DMEM : Dulbecco's Modified Eagle's Medium
DMSO : Dimetilsulfóxido
EB: Embryoid Bodies
EpCAM: Epithelial Cell Adhesion Molecule
FACS: Fluorescence Activated Cell Sorting
F_NPC: Forebrain neural progenitor cells
FZD5: Frizzled class receptor 5
GO: Gene Ontology GSK3: Glucógeno-Sintetasa Quinasa 3
HLF: Fibroblastos embrionarios de pulmón
H_NPC: Hindbrain neural progenitor cells
Iwr-1: Wnt antagonist 1

KDR: Kinase insert Domain Receptor
 KSR : Knock Out Serum Replacement
 LNA: Locked Nucleic Acids
 lncRNA: long-non-coding RNA
 MAPK : Proteína quinasa activada por mitógenos
 MCI: Macizo Celular Interno
 MEFi : Murine embryonic fibroblast irradiated
 Mesp-1 : mesoderm posterior 1
 M_NPC: Midbrain progenitor cells
 MPC: Mesoderm progenitr cells
 MYH6: α -Myosin heavy chain
 NCAM: Neural cell adhesion molecule
 pb: pares de bases
 PCA: Principal Component Analysis
 PFA : Paraformaldehido
 PSC: pluripotent stem cells
 qPCR: Reacción en Cadena de la polimerasa Cuantitativa
 RIP: RNA Immunoprecipitation
 RISC: RNA-induced silencing complex
 RNA-Seq: RNA-sequencing
 SEM: Error estándar de la media
 SFB : Suero fetal bovino
 sgRNA: single guide RNA
 SLO: Stem Loop primer
 SNP: Nucleotide Polymorphism
 SOX 2: Sex Determining Region Y-Box 2
 TGF β : Transforming Growth Factor beta
 TME: Transición epitelio mesenquimal
 UTR: Untranslated region
 VEGF: Vascular Endothelial Growth Factor
 Wnt : Wingless-type

INTRODUCCIÓN

1. Introducción

Las células madre tienen la capacidad de autorenovarse y diferenciarse a distintos tipos celulares específicos, y se las clasifica en células madre adultas o células madre pluripotentes. Las células madre adultas se encuentran en todos los tejidos y además de autorenovarse, son multipotentes, es decir tienen la capacidad de diferenciarse y regenerar células específicas de un tejido frente a determinados estímulos. En ciertos órganos del sistema digestivo y en la médula ósea, estas células se dividen y diferencian regularmente para mantener la integridad de los órganos a lo largo de la vida, aunque con ciertas limitaciones ya que se vuelven quiescentes al alcanzar un número elevado de divisiones. Por otro lado, las células madre pluripotentes (CMP) incluyen a las células madre embrionarias y a las células madre pluripotentes inducidas. Estas células se las conoce como pluripotentes porque tienen la capacidad de diferenciarse a los tres linajes embrionarios: endodermo, mesodermo y ectodermo.

1.1. Células pluripotentes humanas: Características y definiciones

Durante del desarrollo embrionario temprano, luego de la fecundación, el cigoto se divide para dar origen a un embrión multicelular. En la fase de segmentación, luego de la fase de la mórula, se forma el blastocisto, el cual está formado por una masa celular interna situado dentro de la cavidad denominada blastocelo rodeado por una capa de células externas denominada trofoblasto, las cuales darán origen a todos los tejidos extraembrionarios (placenta y membranas amnióticas). Una vez formado el blastocisto, la división celular se enlentece y las células de la masa celular interna dan origen a las tres capas embrionarias, endodermo, mesodermo y ectodermo, en un proceso que se conoce como gastrulación. Las células de las tres capas interactúan entre sí y se reorganizan dando lugar a la formación de tejidos y órganos, proceso denominado organogénesis. Muchos órganos contienen células provenientes de más de una capa germinal. Durante el desarrollo embrionario se denomina células totipotentes a las células que tienen la capacidad de formar todos los tejidos del organismo, mientras que se denomina células pluripotentes a las células que tienen la capacidad de dar origen a los tres linajes embrionarios (lo cual excluye a los tejidos extraembrionarios formados por el trofoblasto). De esta manera, totipotente solo es el cigoto y posiblemente las primeras células de la mórula. En cambio, existen varias células pluripotentes en esta etapa, dependiendo del momento en el que se las aísle.

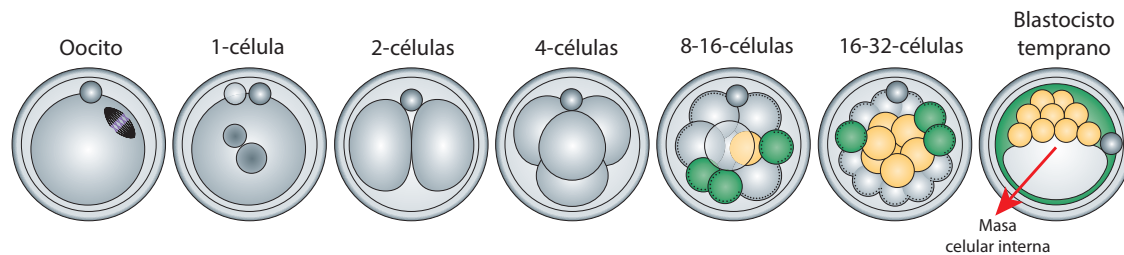


Figura I1: Fases de la segmentación celular: Durante el desarrollo embrionario las células se van dividiendo hasta dar origen al blastocisto, compuesto por la masa celular interna, el trofoblasto que lo rodea y un espacio llamado blastocele. Las CME habitualmente cultivadas corresponden a la masa interna del blastocisto temprano. Figura adaptada de (Zernicka-Goetz et al., 2009)

Las células madre embrionarias (CME) se obtienen a partir de la masa celular interna (Figura I1) y se caracterizan por mantener a lo largo del tiempo una alta tasa de división celular y la capacidad de diferenciarse *in vitro* a todos los tejidos que derivan de los tres linajes embrionarios (Thomson et al., 1998; Amit et al., 2000). Esta capacidad se evidencia mediante la formación de tumores llamados teratomas al inyectar las CME en ratones inmuno deficientes, los cuales se componen de masas celulares que están formadas por distintos tipos de células provenientes del endodermo, mesodermo y ectodermo. Por otro lado, las CME se caracterizan también por tener una alta actividad de telomerasa, lo cual les confiere la capacidad de ser mantenidas durante muchos pasajes en cultivo. El cultivo de la primera línea celular se realizó en 1981, pero recién en 1998 se logró cultivar células madre humanas (Thomson et al., 1998). El principal inconveniente que impidió inicialmente el cultivo de estas células fue la diferencia en los requerimientos nutricionales entre las células madre de ratón y las células madre humanas (CMEh). Sin embargo, en la actualidad ya es posible mantener CMEh en cultivo por un tiempo prolongado mediante el uso de diferentes factores extracelulares y sobre diferentes sustratos, tales como en co-cultivo con fibroblastos embrionarios murinos inactivados o sobre sustratos comerciales más o menos definidos, constituidos por proteínas de la matriz extracelular, principalmente por laminina.

En la actualidad se conocen varios genes que poseen un rol clave en el mantenimiento de la pluripotencia y la identidad de estas células, incluyendo a Oct-4, Nanog y Sox2. Estos factores regulan la transcripción de una batería de genes que permiten mantener las características fundamentales de las CME, tales como la pluripotencia y la auto-renovación (Boyer et al.,

2005). Para ello, los factores de transcripción activan principalmente genes involucrados en las vías de TGF β /Activina/Nodal y la vía de Wnt (James, 2005; Sato et al., 2003). Además, reprimen genes involucrados en la diferenciación celular, tales como Pax6, Meis1, Hoxb1, Esx11, Lhx5, Lbx1, Myf5, Onecut1 (Boyer et al., 2005). Inicialmente, las CME eran mantenidas en cultivo por tiempos prolongados preservando sus características de pluripotencia mediante el co-cultivo con MEFi (del ingles, *Murine Embryonic Fibroblasts inactivated*) suplementado con b-FGF (Amit et al., 2000). Luego, gracias al avance en el campo de las células madre, fue posible cultivar las células sin perder su estado de indiferenciación en medios condicionados por MEFi y sobre *coatings* de matrigel o laminina (Xu et al., 2001). Posteriormente, se desarrollaron medios definidos químicamente que poseen los factores necesarios para el cultivo de las células en condiciones libres de MEFi (Chen et al., 2011).

Por otro lado, existe otro tipo de células madre pluripotentes que comparten características con las CMEh, denominadas células madre pluripotentes inducidas (CMPi). Estas células fueron generadas por primera vez por el grupo de Shinya Yamanaka en el 2007 (Takahashi et al., 2007) mediante la reprogramación de células somáticas adultas. Brevemente, la reprogramación se realiza mediante la expresión ectópica de factores de transcripción que son claves en el desarrollo embrionario tales como Oct-4, Sox2, Klf4, c-Myc o incluso ciertos microARNs como el miR-302 (Subramanyam et al., 2011). Luego de la incorporación de los factores, un número pequeño de células comienzan a cambiar su identidad celular a un estado de pluripotencia, formando colonias, y al cabo de algunos pasajes varias de estas colonias semejan a las colonias de células madre embrionarias. Sin embargo, presentan algunas diferencias como por ejemplo, una memoria epigenética que se mantiene luego de la reprogramación (Kim et al., 2010). Las primeras células que fueron reprogramadas fueron fibroblastos de piel pero en la actualidad se ha logrado reprogramar células de otros tejidos, incluso de aquellos obtenidas mediante métodos poco invasivos como son las células hematopoyéticas de sangre periférica (Staerk et al., 2010).

Tanto las CME como las CMPi mantienen la capacidad de diferenciarse *in vitro* y madurar a todos los tejidos adultos que derivan de los tres linajes embrionarios. Actualmente, existen diversos protocolos que permiten manipular la diferenciación celular y obtener así un determinado fenotipo adulto, aunque a veces de manera poco eficiente.

El estudio de las células pluripotentes humanas ha abierto un campo en la investigación que está en auge desde fines de los años 90. Estas células representan un modelo de estudio

único para conocer más acerca del desarrollo embrionario humano y además tienen un gran potencial para ser utilizadas en la medicina regenerativa y en el desarrollo de nuevos compuestos farmacológicos. Ampliar el conocimiento acerca de los mecanismos que regulan el mantenimiento de la pluripotencia y la diferenciación celular permitirá desarrollar protocolos de diferenciación *in vitro* más eficientes que generen poblaciones celulares de fenotipos bien definidos y en gran cantidad, condiciones esenciales para su uso en el desarrollo de terapias celulares.

1.2. Diferenciación celular

La capacidad de diferenciarse *in vitro* a todo tipo de célula madura proveniente de las tres capas germinales, fue determinado en las CMEh poco después del desarrollo de metodologías para el cultivo de las mismas. Inicialmente, los protocolos de diferenciación se basaban en tres estrategias diferentes (Keller, 2005). El primero y más utilizado, se basaba en cultivar las CMEh como agregados celulares, conocidos como cuerpos embrioides o EB (del inglés *Embryoid Bodies*). Los EB son generados mediante métodos enzimáticos y representan estructuras tridimensionales que facilitan el intercambio de señales moleculares entre las células, permitiendo de esta manera recapitular lo que ocurre durante la embriogénesis (Itskovitz-Eldor et al., 2000; Denning et al., 2006; Kurosawa, 2007). Estos EB eran crecidos en suspensión durante varios días en medios suplementados con suero fetal bovino. Estas condiciones de cultivo recrean la formación de un teratoma *in vitro*. Luego, con el paso de los días de iniciada la diferenciación, se forman células adultas específicas, entre ellas, cardiomiocitos con actividad contráctil espontánea (Itskovitz-Eldor et al., 2000). Los otros dos protocolos se basaban en el co-cultivo de las CME con células estromales murinas y en la diferenciación de las células crecidas en adhesión sobre proteínas de la matriz extracelular (Keller, 2005).

Parte del conocimiento que se tiene hoy en día acerca del desarrollo cardíaco humano durante la embriogénesis se basa en resultados experimentales obtenidos a partir de la diferenciación cardíaca de las CMEh. Los trabajos de (Kehat and Kenyagin-Karsenti, 2001) y (Mummery, 2003) caracterizaron por primera vez el fenotipo cardíaco obtenido a partir de la diferenciación espontánea de los EB derivados de las CMEh. Dichos cuerpos embrioides presentaban características funcionales y estructurales semejantes a las de cardiomiocitos embrionarios. Años después, varios trabajos lograron establecer cuáles son las señales extra e

intracelulares necesarias para llevar a cabo el proceso de diferenciación de manera más dirigida. En los siguientes párrafos se describirán brevemente estos procesos.

1.2.1. Desarrollo embrionario del corazón

Durante el desarrollo embrionario, el corazón comienza a formarse aproximadamente en el día 15 de gestación, a partir del área cardiogénica del mesodermo. Esta área del mesodermo da origen a dos tipos de células que forman los dos tubos endocárdicos, conocidos como FHF y SHF de las siglas en ingles *First Heart Field* y *Second Heart Field*. Estos dos tubos se fusionan en el día 20 para generar el tubo cardíaco primitivo el cual consiste en una capa interna de células que forman el endocardio y una externa que forman el miocardio. Durante el posterior desarrollo embrionario, las células del FHF contribuyen a la formación del ventrículo izquierdo y el SHF contribuye primero a la formación del ventrículo derecho y luego a la formación de los senos venosos y las aurículas derecha e izquierda (Figura I2).

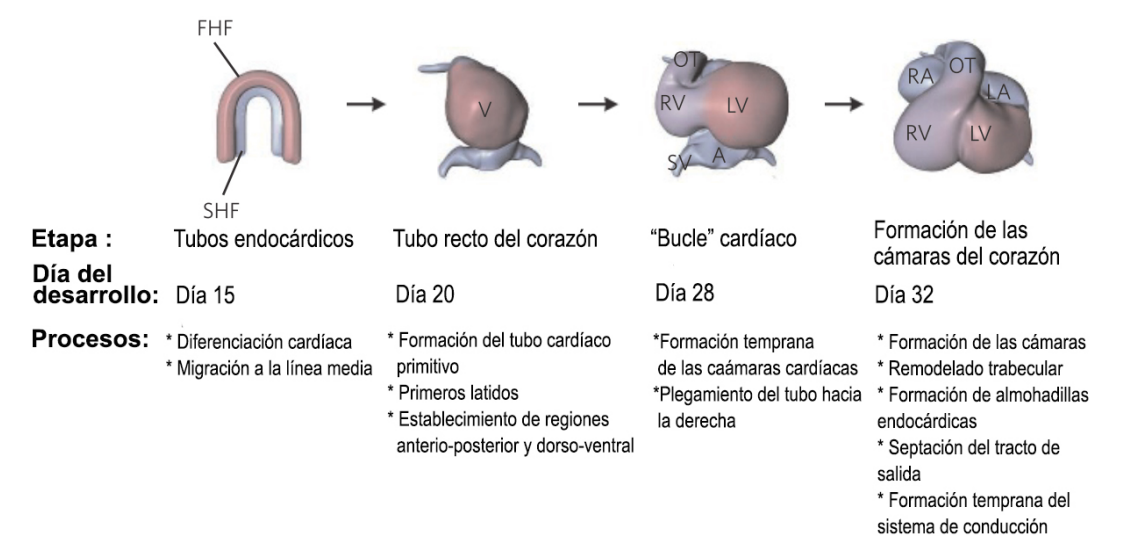


Figura I2: Desarrollo embrionario del corazón: Etapas tempranas del desarrollo del corazón. FHF: First Heart Field; SHF: Second Heart Field; V: Ventricle; OT: Out Flow Tract; RV: Right Ventricle; LV: Left Ventricle; A; Atria; SV: Sinus Venosus; RA: Right Atria; LA: Left Atria. Figura adaptada de (Bruneau, 2008).

El desarrollo cardíaco es un proceso muy complejo que está regulado por muchos genes y cascadas de señalización que controlan los cambios genéticos y morfológicos que permiten la formación del órgano completo y su correcto funcionamiento. Estudios de mutaciones en distintos genes que generan malformaciones cardíacas han permitido identificar muchas de las vías de señalización involucradas en la cardiomiogénesis de distintos organismos (Yin and Pacifici, 2001). Varios factores se han identificado como esenciales para la formación del corazón, como el factor de respuesta a suero, SRF (de las siglas en inglés *serum response factor*), el cual participa de la diferenciación y formación del miocardio, y factores secretados por células del endodermo que incluyen a miembros de la familia de BMPs (del inglés *Bone Morphogenetic Protein*), la familia de FGFs (del inglés *Fibroblast Growth Factor*) (principalmente el factor FGF8), y miembros de la familia de TGF β , particularmente las activinas (Brand, 2003). Mutaciones en los genes que codifican para dichos factores generan aberraciones morfológicas muy severas en el corazón (Brand, 2003). Además, la vía de Wnt también tiene un rol importante en la cardiomiogénesis, pero su efecto es dual dependiendo de la etapa del desarrollo en la que actúe. Se ha demostrado que la activación de la vía de β -Catenina/Wnt en las primeras etapas del desarrollo permite que se establezca el linaje cardíaco, pero luego la inactivación de dicha vía es necesaria para el correcto desarrollo y maduración del corazón (Vijayaragavan and Bhatia, 2007).

Por otro lado, se han estudiado numerosos factores de transcripción que tienen un rol clave en las etapas tempranas de la cardiomiogénesis. La formación del corazón comienza con el desarrollo del mesoendodermo durante el proceso de gastrulación. Durante la primera etapa, las células del mesodermo en desarrollo expresan los factores de transcripción *Brachyury T* y *Eomes*. Luego, se activan en la línea primitiva en desarrollo genes involucrados en la diferenciación a mesodermo y la transición epitelio-mesenquimal, tales como FGF-4 y FGF8. Esto lleva a la aparición del mesodermo precardiaco en donde comienza a expresarse el factor de mesodermo posterior *Mesp-1*, denominado regulador *master* de ésta etapa (Saga et al., 1996). Estas células dan origen a las células precursoras cardíacas pero también a otros tejidos como aquellos que derivan del mesodermo paraxial y en músculo esquelético (Annis et al., 2009). Una vez que se forman las primeras células progenitoras cardíacas se expresan genes que tienen un rol clave en el desarrollo cardíaco. Unas de las primeras vías de señalización que se activan incluyen a la familia de los genes *homeobox NK* (Nkx2.3, Nkx2.5, Nkx2.6) y los factores de transcripción *zinc-fingers GATA* (GATA 4-6). Deficiencias en algunos de estos factores ge-

formación del mismo (Tam and Behringer, 1997). Esta transición se caracteriza por cambios genéticos y morfológicos que permiten que las células pierdan adhesión y adquieran la capacidad de migrar. En la línea primitiva, ciertas células del epiblasto migran entre el epiblasto y el endoblasto para formar el mesoendodermo primitivo, el cual luego dará lugar al mesodermo y el endodermo. La TEM es un proceso descrito en tres circunstancias: en el desarrollo embrionario, en la cicatrización de heridas, y en el desarrollo de tumores, particularmente en la invasión y metástasis. Este proceso ha sido muy caracterizado en los tumores metastásicos. Con la TEM, las células cancerígenas se vuelven más agresivas e invasivas, pierden las características de células epiteliales y adquieren características de mesénquima (Guarino et al., 2007). Las células embrionarias, en similitud con células epiteliales tumorales, establecen una relación apical-basal con las células en las que se co-cultivan y tienen una alta expresión de moléculas de adhesión célula-célula, tales como la conexina-43 y la E-cadherina (Krtolica et al., 2007). Las principales marcas que caracterizan la TEM son la pérdida de E-cadherina y la expresión de proteínas mesenquimales tales como la Vimentina y N-cadherina (CD56). Además, hay varios factores de transcripción tales como Snail, Slug, ZEB1, ZEB2 y Twist que inhiben la expresión de E-cadherina y juegan un rol importante en la inducción de la TEM y el mantenimiento del fenotipo invasivo y migratorio de las células cancerígenas (Heldin et al., 2012). Por otro lado, varios trabajos plantean que la vía de señalización de TGF β cumple un rol fundamental en la inducción de la TEM mediante la cascada de fosforilación de factores como las SMADs, reguladores positivos de la expresión de los inhibidores de E-cadherina mencionados previamente (Derynck and Zhang, 2003; Roberts and Wakefield, 2003).

En el 2010 Denis Evseenko y col. (Evseenko et al., 2010) lograron identificar durante la diferenciación a mesodermo, un progenitor temprano que tiene la capacidad de generar todos los tipos celulares originados a partir de ésta capa germinal, es decir, el linaje hematopoyético, cardiovascular y mesenquimal y sus respectivos derivados. Este progenitor se induce mediante un protocolo de diferenciación a mesoendodermo, creciendo las células en adhesión en un medio suplementado con los factores de crecimientos BMP-4, Activina A, VEGF (del inglés *Vascular Endothelial Growth Factor*) y FGF2. En el día 3,5 de iniciada la diferenciación aparece un progenitor temprano de mesodermo que se caracteriza por haber perdido la molécula de adhesión EpCAM (CD326), presente en células indiferenciadas, y haber adquirido la molécula de membrana NCAM (CD56). Este progenitor presenta un perfil de expresión de genes característico de células que sufrieron la transición epitelio mesenquimal, tales como

Snail-1, Snail-2, Twist, Eomes, Mesp-1/2 y Vimentina y la inactivación de E-cadherina y otros genes que participan de las uniones celulares ocluyentes. Además, al analizar la expresión de genes de los tres linajes embrionarios, determinaron que el progenitor presenta una expresión alta de genes de mesodermo y una expresión baja de genes característicos de endodermo y ectodermo. Este progenitor temprano de mesodermo representa una población única y bien definida que permite estudiar los mecanismos que regulan la formación de mesodermo sin la presencia de otros tipos celulares, lo cual podría enmascarar ciertos resultados.

1.2.3. Diferenciación *in vitro* al linaje cardíaco

Los primeros protocolos de diferenciación *in vitro* de células madre pluripotentes se basaron en la diferenciación espontánea mediante el cultivo de las células en EB en un medio suplementado con 20 % suero fetal bovino. Los EB eran crecidos en suspensión durante los primeros 4 días y luego adheridos entre los días 4 y 7 de iniciada la diferenciación. A partir del día 13 aproximadamente era posible encontrar en los EB ciertas regiones con actividad contráctil. Luego, se desarrolló un protocolo más dirigido hacia el linaje cardíaco, el cual se basaba en el co-cultivo de las células con una línea celular murina de endodermo visceral (END-2, *visceral endoderm-like cells* (Mummary, 2003). Sin embargo, el porcentaje de células positivas para marcadores del linaje cardíaco era bajo (entre un 20 a 40 % dependiendo de la línea celular), y de una composición celular compleja y heterogénea. Esto llevó a que trabajos posteriores buscaran desarrollar protocolos más eficientes que permitieran obtener cultivos más homogéneos y enriquecidos en células cardíacas.

La principal estrategia de estos protocolos fue activar determinadas vías moleculares específicas de cada etapa de la cardiomiogénesis e inhibir las vías de señalización que dirigen la diferenciación hacia otros linajes, recapitulando de esta manera las etapas claves del desarrollo cardíaco. La mayoría de los protocolos fueron desarrollados de esta manera, los cuales si bien presentaban ciertas diferencias, todos se basaban en la regulación de las mismas vías de señalización (Burridge et al., 2012). Estos protocolos consistían en crecer las células en cuerpos embrioides o cultivando las células en adhesión en monocapa y en medios suplementados con distintos factores de crecimiento. Los primeros protocolos activaban en las etapas iniciales las vías de señalización de BMP, Wnt y TGF β /Activin/NODAL para estimular la formación del mesodermo cardíaco. Básicamente, se suplementaba el medio de cultivo con BMP-4, Activina

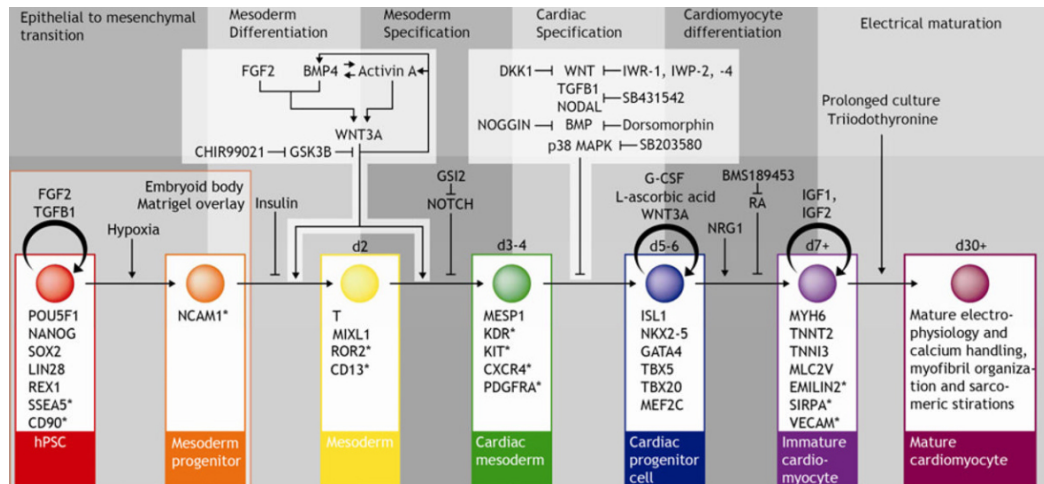


Figura I4: Etapas de la diferenciación cardíaca: En este esquema se resumen algunas de las poblaciones celulares que se forman a lo largo de la diferenciación cardíaca a partir de las células madre pluripotentes. Figura adaptada de (Burridge et al., 2012).

A y FGF2. La concentración adecuada de cada uno de estos factores es dependiente de la línea de célula madre pluripotente, lo cual requiere entonces que sean optimizadas para cada línea en particular. Luego, una vez establecido el linaje cardíaco, se inhibía la vía de Wnt mediante el uso de pequeñas moléculas como IWR-1, IWP-1, IWP-2 o factores de crecimiento inhibidores de esta vía como DKK-1. Trabajos posteriores lograron determinar mejoras en los protocolos que favorecían la diferenciación cardíaca, tales como agregar ácido ascórbico al medio y crecer las células sin insulina durante los primeros días (Freund et al., 2008). Posteriormente, en el 2012 el grupo de Palecek desarrolló un protocolo en el cual se obtenía una alta eficiencia en la diferenciación cardíaca únicamente mediante la modulación de la vía de Wnt con pequeñas moléculas (Lian et al., 2012). Este protocolo consistía en cultivar las células durante las primeras 24hs con CHIR99021, un inhibidor de GSK3B, y luego desde el día 3 al 5 inhibir la vía de Wnt mediante el uso del inhibidor IWP-4. La inhibición de la vía de GSK3B (del inglés *Glycogen synthase kinase 3 beta*) permite la activación de BMP-4 y la vía Activin/NODAL.

En la figura I4 se resumen algunos de los tipos celulares que marcan las distintas etapas de la diferenciación cardíaca y las vías de señalización involucradas en cada una.

Para determinar el progreso de la diferenciación cardíaca y la pureza del cultivo, se analiza la expresión de distintos marcadores cardíacos. Entre los más utilizados, se encuentran proteínas estructurales como la troponina cardíaca (cTNT). Para identificar poblaciones pro-

genitoras de etapas intermedias en la diferenciación cardíaca, se mide habitualmente factores de transcripción pro-cardíacos, tales como Brachyury/T o Nkx2.5, que son característicos del progenitor de mesodermo y progenitor cardíaco, respectivamente (Burridge et al., 2012).

A pesar de los esfuerzos realizados para desarrollar protocolos que permitan obtener cultivos enriquecidos en células cardíacas, fue necesario establecer métodos que permitan aislar los precursores cardíacos y/o eliminar los otros tipos celulares no deseados. Para esto, se utilizaron varias estrategias. La primera se basa en la generación de líneas celulares que expresen una proteína fluorescente o resistencia a antibiótico bajo la regulación de promotores cardíacos como el promotor de cTNT y de Nkx2.5. La segunda estrategia, que evita la modificación genética de las células, se basa en el uso de marcadores moleculares de membrana que permitan identificar y aislar las células mediante anticuerpos acoplados a *beads* magnéticas o por FACS (del inglés, *Fluorescent-Activated Cell Sorting*). Varios marcadores de membrana específicos para cardiomiocitos han sido identificados, tales como EMILIN2, SIRPA y VCAM (Burridge et al., 2012). Finalmente, en el trabajo de Tohyama y col. basaron su estrategia de purificación en un tipo particular de metabolismo que tienen los cardiomiocitos (Tohyama et al., 2013). Las células cardíacas tienen la capacidad de utilizar distintos azúcares como fuente de energía, tales como glucosa, ácidos grasos y lactato. Este grupo se basó entonces en la ventaja metabólica que proporcionan los cardiomiocitos al poder metabolizar el lactato en ausencia de glucosa. De esta manera, lograron enriquecer el cultivo en más de un 90% de cardiomiocitos mediante su cultivo en un medio sin glucosa y rico en lactato.

1.2.4. Potencial uso de las células madre pluripotentes en la medicina

Las células madre pluripotentes (CMP) representan una herramienta con mucho potencial en lo que refiere a la medicina traslacional. En los primeros años, muchas fueron las expectativas acerca del uso de estas células en el desarrollo de terapias regenerativas, pero mediante el avance del conocimiento en el área se descubrieron varios aspectos que hicieron que el camino hacia su uso en la medicina clínica fuera más difícil de lo que se pensaba. Lograr un control preciso de la diferenciación dirigida, escalar la producción de células de un tipo de tejido específico y lograr que éstas células se interconecten al tejido preexistente, son algunos de los desafíos actuales para poder utilizar las células madre pluripotentes en terapias celulares. Sin embargo, hoy en día algunas terapias que utilizan células diferenciadas a partir de éstas célu-

las ya están siendo evaluadas en ensayos clínicos. Tal es el caso de tratamientos para daño en la médula espinal, degeneración del epitelio pigmentario de retina, problemas cardiovasculares, diabetes y Parkinson (Thies and Murry, 2015). Además, muy recientemente se describió el primer implante exitoso de cardiomiocitos derivados de células madre embrionarias en seis pacientes que fueron sometidos a cirugía de revascularización coronaria. Es de destacar que no existieron efectos adversos atribuibles a la inyección de las células (Menasché et al., 2018). Un objetivo aún más a largo plazo sería utilizar en las terapias celulares CMPi propias de cada paciente para evitar así problemas de rechazo inmunológico debido a diferencias en el sistema de mayor de histocompatibilidad entre las células donantes y el receptor.

Dos aplicaciones de las CMPih han sido ya desarrolladas y utilizadas de manera exitosa. La primera es el modelado *in vitro* de enfermedades de causas genéticas. Diferenciar células pluripotentes inducidas que posean la mutación responsable de la patogénesis brinda una herramienta inigualable para estudiar el rol que tiene dicho gen mutado en el progreso de la enfermedad y poder desarrollar así tratamientos adecuados y eficientes para tratarla. Ejemplos de estos casos son el estudio de la atrofia muscular espinal (de las siglas en inglés SMA), el estudio del síndrome de Lesch-Nyhan (Zhu and Huangfu, 2013) y el estudio del síndrome del QT largo (Moretti et al., 2010). En los tres casos el modelado de la enfermedad utilizando células de pacientes brindó ventajas importantes sobre los modelos murinos utilizados hasta entonces.

La segunda aplicación relacionada con la medicina se basa en el uso de las CMPih para el *screening* y desarrollo de nuevos fármacos. Mediante el desarrollo de las nuevas tecnologías *high-throughput* es posible evaluar compuestos químicos de interés farmacológico en células de tejidos específicos y analizar sus efectos. Un ejemplo de esto es el uso de cardiomiocitos derivados de CMPi para el testeo de fármacos pre-clínicos para cardiopatías y el análisis de efectos no deseados como arritmias cardíacas (Liang et al., 2013). Evaluar futuros fármacos comerciales en células específicas de corazón, brinda ventajas incomparables con respecto al uso de otras líneas celulares que muchas veces no presentan los mismos efectos frente al tratamiento, lo cual representa un gran riesgo y una alta probabilidad de que la droga desarrollada no sea efectiva o incluso peligrosa para la salud.

Dado el gran impacto que tienen las CMP en la medicina moderna, desde hace más de 15 años muchos grupos de investigación trabajan en el desarrollo de nuevas técnicas para el uso aplicado de estas células, como así también en el estudio de la biología de las mismas.

Un mayor entendimiento de los procesos celulares y mecanismos moleculares involucrados en la diferenciación celular es imperativo para poder avanzar en el desarrollo de tecnología con CMP y su aplicación en la medicina traslacional.

1.3. MicroARNs: pequeños reguladores génicos

La transcripción del ADN en las células eucariotas da lugar a diferentes moléculas de ARN con múltiples funciones. Estas funciones comienzan a ser comprendidas a partir de los trabajos del consorcio ENCODE (Djebali et al., 2012). Uno de los hallazgos más notables es que la mayoría de los ARNs producidos son no codificantes: sólo el 2% de los ARNs transcriptos son traducidos a proteínas. El resto de los ARNs transcriptos tienen funciones estructurales, enzimáticas o de regulación. Arbitrariamente, los ARNs no codificantes han sido divididos, según su longitud, en los ARN de cadena larga (mayor a 200 bases, llamados *long-non-coding-ARNs*), y los de cadena corta (menor a 200 bases, llamados *short-non-coding-RNAs*). Las diferencias, sin embargo, no son sólo en la longitud: los grupos de ARNs presentan características funcionales muy disimiles.

En los animales existen tres tipos de pequeños ARNs no codificantes: los siARNs, los piARNs y los microARNs. Los siARN (del inglés *small interfering RNA*) derivan de ARNs largos de doble cadena, tienen un largo de ~21 nucleótidos y regulan de manera post-transcripcional transcriptos y transposones, y participan de la defensa anti viral. Los piARN tienen un tamaño entre 24 a 30 bases (o más dependiendo de la especie) y provienen del procesamiento de moléculas de ARN simple cadena que son clivadas por la enzima *Zucchini*. Su función es poco conocida en células adultas, pero en las células germinales tienen por función regular los transposones y evitar las alteraciones genéticas durante la recombinación (Ha and Kim, 2014). Los microARNs son pequeñas moléculas de ARN no codificante, simple cadena y evolutivamente muy conservados que regulan de manera postranscripcional la expresión génica. Fueron descritos por primera vez en el año 1993 en el trabajo de Lee y col. en el que demostraron que el microARN *lin-4* participa en el desarrollo temprano de *C. elegans* (Lee, 1993). Inicialmente, se consideró que esa pequeña molécula de ARN era única en su tipo. Más tarde, en los trabajos de Reinhart y col. y Slack y col. se descubrió el segundo microARN, *let-7* lo cual marcó el inicio de cientos de trabajos posteriores que estudiaron la presencia de los microARNs en distintos organismos (Reinhart et al., 2000; Slack et al., 2000). Desde entonces, han

adquirido una gran importancia y en el genoma humano ya se han descubierto alrededor de 2800 microARNs (<http://www.mirbase.org>) que regularían, al menos, el 60 % de las proteínas (Friedman et al., 2009).

Los microARN reconocen por complementariedad de secuencia a sus ARN mensajeros (ARNm) *target* y regulan su estabilidad mediante distintos mecanismos. A pesar de que sean considerados reguladores finos de la expresión génica, se ha demostrado que el *knock-out* de enzimas clave de la biogénesis y procesamiento de los microARNs en su conjunto genera letalidad en etapas tempranas del desarrollo (Wang et al., 2010; Bernstein et al., 2003). Los microARNs pueden tener una expresión ubicua o patrones de expresión bien definidos en diversos tejidos o tipos celulares, ejerciendo su función en el control de numerosos procesos celulares, ya sean normales o patológicos (Bartel, 2009).

En la última década, el estudio de los microARNs ha adquirido una gran relevancia en la medicina y en el desarrollo de terapias. Se ha determinado que numerosos microARNs son desregulados en distintas etapas del progreso tumoral, siendo así eficientes biomarcadores que permiten establecer tanto el diagnóstico como pronóstico de la enfermedad (Schwarzenbach et al., 2014). La principal ventaja que presentan es su gran estabilidad y fácil detección en fluidos como la sangre y la orina. A su vez, se han desarrollado tratamientos contra diferentes enfermedades basados en la sobreexpresión de los microARNs con moléculas sintéticas que mimetizan su función o bien bloqueado su actividad con antagomiRs o LNA (de las siglas en inglés *Locked Nucleic Acids*). Tal es el caso de las terapias contra el dolor, tratamientos contra enfermedades cardiovasculares y tratamientos contra el cáncer, entre otros (Tan et al., 2013; Thum, 2012; Nana-Sinkam and Croce, 2011).

Además, los microARNs tienen un rol fundamental en el control del desarrollo embrionario y el mantenimiento de la identidad de los tejidos tanto en la embriogénesis como en tejidos adultos (Fazi and Nervi, 2008). Un mayor entendimiento del rol que los microARNs cumplen tanto en las células madre pluripotentes como en la diferenciación celular permitirá conocer y comprender mejor los mecanismos moleculares que controlan los complejos procesos involucrados en las primeras etapas del desarrollo.

1.3.1. Biogénesis y procesamiento de los microARNs

Los genes que codifican para los microARNs se encuentran dispersos a lo largo de todo el genoma y pueden encontrarse en *clusters* que producen transcritos policistrónicos, en regiones intergénicas o en regiones génicas, dentro de intrones o incluso dentro de exones (Fazi and Nervi, 2008). La biogénesis de los microARNs es un proceso complejo muy regulado que involucra varios pasos, los cuales pueden ser clasificados en 3 etapas (Winter et al., 2009). Las primeras dos etapas ocurren dentro del núcleo y la tercera y última en el citoplasma (Figura 15). En la primera etapa, los microARNs se transcriben principalmente por la ARN polimerasa II (Pol II) la cual produce el pri-microARN. Este precursor consiste en una molécula de ARN que puede medir cientos de nucleótidos hasta algunas kilobases (Saini et al., 2007), posee las modificaciones características de los transcritos de la pol II (5' 7-methylguanosine cap y cola de poliA) y contiene una estructura doble cadena conocida como *hairpin*.

Luego en la siguiente etapa, el pri-microARN es clivado por el complejo microprocesador nuclear formado por las enzimas DGCR8 y la endonucleasa ARNasa III, Drosha. La enzima DGCR8 se une al pri-microARN y marca el sitio específico de clivado para que Drosha corte y produzca de esta manera el pre-microARN, lo cual ocurre de manera co-transcripcional. Para ciertos microARNs que se encuentran dentro de intrones, se cree que este procesamiento ocurre durante el proceso de *splicing* y mediante la maquinaria del mismo (Winter et al., 2009). El pre-microARN consiste en una ARN doble cadena de ~65 nucleótidos que posee la forma de *hairpin* y es exportado hacia el citoplasma mediante el complejo de transporte nuclear formado por la exportina 5 (EXP5) y la proteína Ran-GTP.

Por último, la tercer etapa consiste en la formación del microARN maduro, el cual se genera a partir del clivado por parte de la endonucleasa Dicer. Esta enzima forma un complejo junto con las proteínas TRBP y PACT, las cuales se unen al microARN y facilitan la actividad de Dicer. De esta manera, se escinde la porción que forma el *loop* y se produce una molécula de ARN doble cadena de ~22 pares de bases que es cargada en el complejo RISC (de las siglas en inglés *RNA-induced silencing complex*) formado principalmente por la proteína Argonauta (en humanos son cuatro, AGO 1-4). Luego, de manera selectiva, la cadena *sense* o la *antisense* (conocidas como -5p y -3p, respectivamente) del ARN doble cadena es eliminada, para que sólo una hebra forme parte del complejo y cumpla su función de regular post-transcripcionalmente a sus ARNm *target*. La otra cadena "no funcional" puede ser o bien retenida o enviada a degra-

dación. Varios trabajos postularon que la hebra que efectivamente ejerce su función depende del contexto celular o incluso puede que ambas estén presentes pero que sean cargadas en RISC con distinta afinidad (Ha and Kim, 2014).

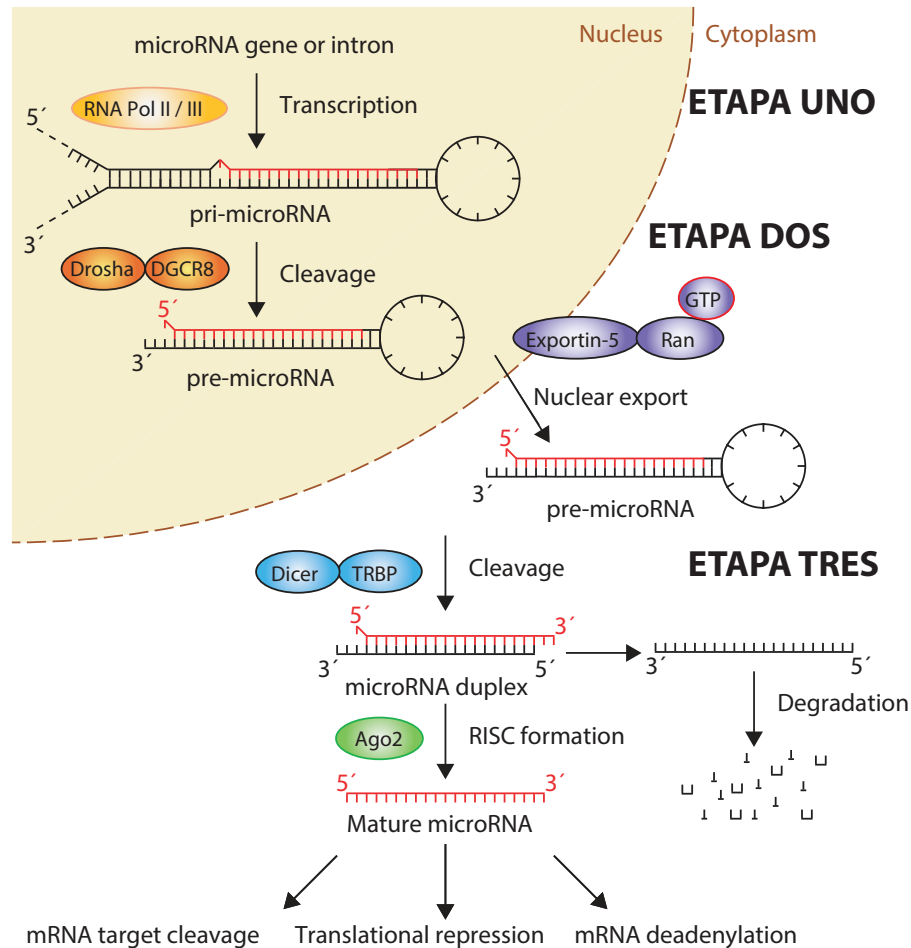


Figura I5: Biogénesis y procesamiento de los microARNs: Representación esquemática de las distintas etapas de la formación y procesamiento de los microARN en mamíferos. Figura adaptada de (Winter et al., 2009).

1.3.2. MicroARNs y la regulación de la expresión génica

La actividad reguladora que ejercen los microARNs se inicia con el reconocimiento de sus ARNm blanco mediante complementariedad de secuencia. Dentro de la secuencia de los ARNm se encuentran sitios de unión al microARN, conocidas como MRE (de las siglas en inglés, *MicroRNA Recognition Element*). Generalmente los ARNm contienen varios MRE reconocidos por uno o más microARNs, hecho que potencia marcadamente su capacidad de regulación. Ini-

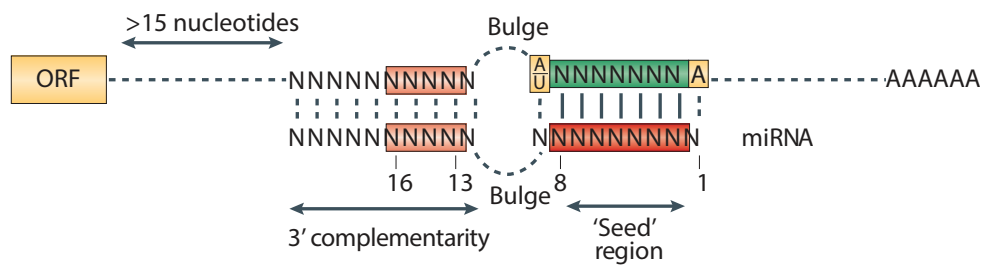


Figura I6: Reconocimiento de los ARNm blanco: En el esquema se pueden observar las distintas regiones de la secuencia del microARN y el tipo de apareamiento que tienen con la región de reconocimiento en el 3'UTR del ARN blanco. Figura adaptada de (Filipowicz et al., 2008).

cialmente, se había postulado que el apareamiento entre el microARN y su mensajero blanco se daba entre la región 5' del microARN (entre los nucleótidos 2-8), denominada secuencia semilla (*seed sequence*) y la región 3' no codificante del mensajero blanco (3'UTR) (Lewis et al., 2003). Sin embargo, trabajos posteriores postularon mecanismos de reconocimiento alternativos. Por un lado, se determinó que la región 3' del microARN ejercería un rol importante y necesario para el reconocimiento del mensajero *target* (Bartel, 2009). A esta región que comprende a los nucleótidos 13-17 se la denomina "*3'-compensatory site*" (Figura I6). Y por otro lado, se han identificado también sitios de interacción microARN-ARNm en las regiones 5' UTR e incluso dentro de la secuencia codificante de los mensajeros, aunque con menor frecuencia (Lytle et al., 2007; Tay et al., 2008). Por lo tanto, el reconocimiento canónico del 3'UTR debe ser considerado orientador pero no necesariamente efectivo.

El mecanismo por el cual los microARNs silencian sus ARNm blanco ha sido un tema de mucho estudio y diferentes modelos han sido propuestos. Una vez que el microARN ha sido cargado en el complejo RISC, reconoce a sus mensajeros blanco mediante complementariedad de secuencia y regula de manera negativa su expresión. Dependiendo si el apareamiento es perfecto o imperfecto, el ARN mensajero será degradado o su traducción será bloqueada, respectivamente. En el caso de que se bloquee la traducción, puede ocurrir en cualquiera de las siguientes etapas: iniciación, elongación o generando terminaciones prematuras (Huntzinger and Izaurralde, 2011). A su vez, el ARN mensajero puede ser acumulado en cuerpos de procesamiento ubicados en el citoplasma, conocidos como *p-bodies* (Fazi and Nervi, 2008).

Mediante el desarrollo de técnicas *genome-wide* como la secuenciación masiva de ARN y estudios de proteómica, varios trabajos han analizado el porcentaje de genes que son regulados

vía degradación del mensajero y vía inhibición de la traducción. En el trabajo de Huntzinger y col. se resumen los hallazgos encontrados al respecto (Huntzinger and Izaurralde, 2011). A pesar de ciertas contradicciones en los resultados, la conclusión fue que la degradación del ARNm explica una gran parte de la regulación por parte de los microARNs en los animales (en las plantas existen ciertas diferencias en los mecanismos de regulación).

Por otro lado, se ha propuesto y demostrado que ciertos microARNs pueden tener efectos opuestos en la regulación génica, siendo en realidad activadores de la expresión de genes. En el trabajo de Place y col. determinaron que el miR-373 se une a los promotores de E-caderina y CSDC2 activando su expresión (Place et al., 2008), y el trabajo de Vasudevan y col. postuló que ciertos microARNs cargados en el complejo RISC reclutan ribonucleoproteínas como *FXR1* que favorecen la traducción de los ARNm blanco (Vasudevan et al., 2007). Todos estos mecanismos de regulación propuestos contribuyen a la complejidad asociada con los microARNs y la manera de ejercer su rol. Probablemente, futuros trabajos van a aportar nuevas evidencias que permitirán comprender mejor los mecanismos moleculares involucrados en la regulación post-transcripcional por parte de los microARNs.

1.3.3. Estudio de los genes blanco: análisis *in silico* y experimental

Dado que un mismo microARN puede regular cientos de genes, poder identificar los mensajeros blanco de manera experimental es una tarea ardua y laboriosa. Como consecuencia, diversas herramientas bioinformáticas han sido desarrolladas con el fin de poder predecir, con un alto nivel de certeza, las interacciones microARN-ARNm, y ayudar de esta manera a la interpretación de los resultados experimentales. Cada programa se basa en un algoritmo de búsqueda diferente, pero la mayoría tienen en cuenta ciertas características tales como el tamaño de la secuencia semilla, la conservación evolutiva de los sitios MRE entre especies, la energía de unión entre el microARN y el mensajero, la accesibilidad al sitio de unión y el número de MRE (Akhtar et al., 2016). Algunos de los más utilizados son TargetScan, miRanda, PicTar, PITA, RNA22, entre otros. Cada uno de estos programas tiene en cuenta las características enumeradas antes, pero le da distinto peso a cada una de ellas, prediciendo de esta manera genes blanco con distinta probabilidad. Muchos genes son identificados por varios de los programas, pero otros son predichos únicamente por algunos de ellos. Por ejemplo, TargetScan prioriza la conservación evolutiva del sitio de unión, mientras que PITA prioriza

la accesibilidad del microARN al sitio de unión en el ARNm blanco, evaluando el delta de energía libre al efectuarse la interacción. Para poder identificar con mayor certeza y confiabilidad los genes blanco, es necesario utilizar en simultáneo varias herramientas de predicción. Algunos programas disponibles utilizan la conjunción de los resultados de los algoritmos mencionados para aumentar la capacidad de predicción, como por ejemplo el programa miRwalk (<http://mirwalk.uni-hd.de/>). Además, se han desarrollado herramientas que permiten predecir genes *target* de microARNs nuevos (no identificados previamente). Algunos de ellos son DIANA-microT v3.0, TargetScan v5.2 y RNAhybrid (Akhtar et al., 2016).

A pesar de que el análisis *in silico* de los genes *target* permite identificar muchas veces de manera certera los genes regulados por un microARN específico, el estudio y validación experimental de los mismos es necesaria. Además, es importante aclarar que los análisis *in silico* predicen todos los *targets* posibles, y no únicamente los expresados en un contexto celular específico. Tal como se mencionó anteriormente, cada microARN puede regular hasta cientos de ARNm, lo cual muchas veces genera que el análisis de los niveles de expresión de sólo algunos genes sea poco informativo. Es por esto que mediante el avance de las tecnologías *high-throughput*, se han podido desarrollar técnicas experimentales para estudiar los genes *target* de manera más global. Estas técnicas se las puede clasificar en dos grupos, uno basado en el análisis de expresión de los genes luego de modular la expresión de un determinado microARN y el otro basado en la co-inmunoprecipitación de los mensajeros junto con proteínas del complejo RISC y el microARN. En el primer caso, la técnica consiste en sobreexpresar o inhibir un microARN mediante moléculas sintéticas tales como *mimics* o *inhibitors/LNA* y luego analizar cambios en la expresión de los genes mediante microarray o RNA-seq, en el caso de analizar los niveles de ARNm, o mediante ensayos de proteómica o de *ribosome profiling*, en el caso de analizar niveles de proteínas (Hausser and Zavolan, 2014). En el segundo grupo, las técnicas consisten en el análisis de los ARNm que interactúan con las proteínas del complejo RISC. En este caso, hay dos metodologías diferentes, aquellas en las que se realiza un *cross-linking* (HITS/PAR-CLIP y CLASH) y aquellas en las que no (RIP). Las técnicas HITS/PAR-CLIP y CLASH se basan en generar una unión estable entre las proteínas del complejo RISC, el microARN y el ARNm mediante *cross-linking* y luego precipitar con un anticuerpo específico una de las proteínas del complejo. Comúnmente, se inmunoprecipita con un anticuerpo anti-Argonata. Luego, se fragmenta el ARN y mediante secuenciación masiva se identifican los sitios de unión ARNm-microARN. A continuación, por medio de programas bioinformáticos, se

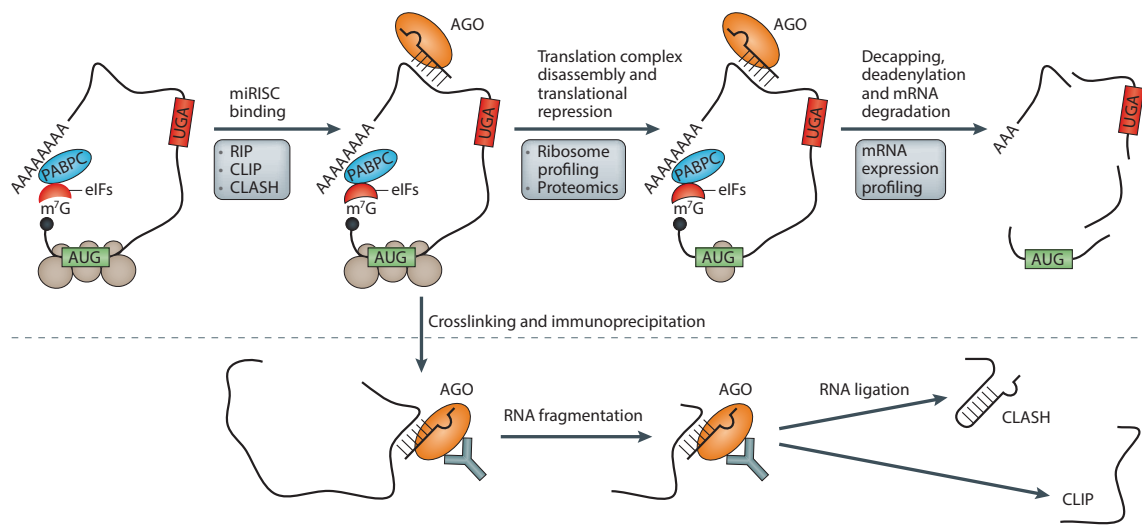


Figura I7: Identificación experimental de los ARNm blanco: En la figura se observa de manera resumida algunas de las técnicas que permiten identificar de manera experimental los genes *target* regulados por los microARNs. Figura adaptada de (Hausser and Zavolan, 2014).

determina la identidad de los microARNs y los ARNm secuenciados y por complementariedad de secuencia se estiman las interacciones. La diferencia entre el método CLIP y CLASH, se basa en que este último requiere un paso de ligación que permite determinar con mayor precisión las interacciones microARN-ARNm. Por otro lado, la técnica RIP (del inglés, *RNA Immuno Precipitation*) presenta ciertas diferencias. La principal, se basa en analizar los genes blanco de un microARN en particular. Para ello, se sobreexpresa el microARN en estudio mediante la transfección con un oligonucleótido sintético conocido como *mimic*. Luego, se lisan las células mediante el uso de un *buffer* de lisis suave que permite preservar las interacciones proteína-ARN, se extrae el contenido celular, se co-inmunoprecipita la proteína Argonauta junto con los ARNm y al igual que las otras técnicas, se identifican los ARNm cargados en RISC mediante secuenciación masiva (Hausser and Zavolan, 2014). En la figura I7 se resumen de manera esquemática las distintas técnicas para identificar los genes blanco de los microARNs.

Finalmente, cabe destacar que existen programas que permiten analizar las interacciones entre microRNA y mRNA por medio de la unión de las bases de datos *in silico* con bases de datos formadas a partir de los resultados experimentales realizados con HITS/PAR-CLIP y CLASH (Fan et al., 2016).

1.3.4. Clasificación de los microARNs en familias y clusters

Los microARNs se caracterizan por formar complejas redes de regulación génica, en donde un mismo microARN regula cientos de genes blanco y un solo ARNm puede ser regulado en simultáneo por varios microARNs diferentes. Gracias al mayor conocimiento que se tiene acerca de la biología de los microARNs y la manera en la que ejercen su función, surgió la clasificación en dos tipos de grupos, las familias y *clusters* de microARNs, simplificando de cierta manera el abordaje de su estudio.

Por un lado, aquellos microARNs que compartan una misma secuencia semilla se agrupan en una familia, y dada la función que tiene esa región, los miembros de una misma familia van a regular *pool* de genes similares. En un principio, se planteó que los microARNs de una familia regularían la expresión génica de manera redundante, pero luego se determinó que muchas veces los miembros de una familia regulan los mismos genes pero se expresan en diferentes contextos celulares (en distintos tipos de tejido, frente a distintos estímulos, *etc*). Por ejemplo, tal es el caso de la familia del miR-125, formado por el miR-125a, miR-125b y miR-125-2, la cual participa de distintas vías de señalización y su desregulación se ha asociado con distintas patologías. Muchas veces estos microARNs actúan de manera cooperativa pero otras, dependiendo del contexto, actúan cada uno de ellos de manera independiente (Sun et al., 2013).

Por otro lado, los microARNs se agrupan en *clusters* cuando se encuentran ubicados muy próximos en el genoma y dada esta cercanía comúnmente están bajo el control de los mismos elementos regulatorios (como por ejemplo promotores y *enhancers*). En general, dadas las diferencias que existen entre sus secuencias semilla no regulan a los mismos genes, sin embargo muchos de los genes que silencian pertenecen a las mismas vías de señalización, actuando de esta manera en cooperación frente a un mismo estímulo. Un ejemplo es el caso del *cluster* miR-17~92. Este *cluster* está formado por el miR-17, miR-18a, miR-19a, miR-20a, miR-19b y miR-92a, los cuales a su vez forman parte de cuatro familias distintas. Estos microARNs se expresan tanto en células embrionarias como en distintos tipos de cáncer, en donde los miembros del *cluster* regulan distintos genes que participan principalmente de la proliferación celular (Hausser and Zavolan, 2014) (Figura I8).

Si bien estas dos categorizaciones son comúnmente aceptadas, es importante aclarar que presentan ciertas falencias. Por un lado, la definición de familia es algo imprecisa en cuanto a

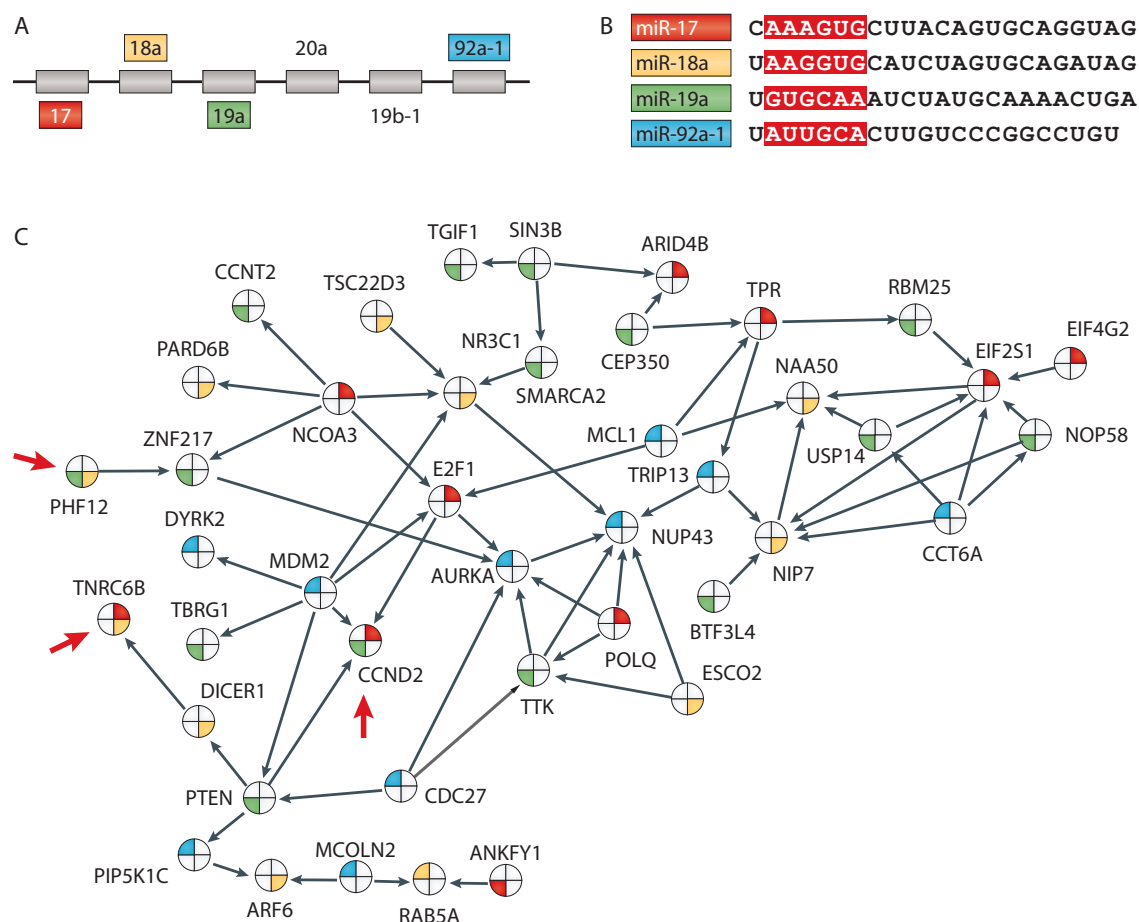


Figura I8: Cluster miR-17~92 y la red de regulación de genes: (A) El cluster miR-17~92 está formado por los microARNs miR-17, miR-18a, miR-19a, miR-20a, miR-19b y miR-92a. (B) En rojo se representan las cuatro secuencias semillas que poseen los microARNs del cluster. (C) Los genes target de cada microARN esta representado por los distintos colores y las flechas rojas señalan los genes target en común. Figura adaptada de (Hausser and Zavolan, 2014).

la extensión establecida en la similitud de las secuencias, y no se ha definido claramente el rol que tendría el resto de la secuencia que no forma parte de la región semilla (la cual podría tener también un rol regulatorio). Asimismo, la clasificación de *cluster* es arbitraria, los microARNs pertenecerían al mismo *cluster* cuando se encuentran codificados a menos de 10.000 bases. Así, algunos *clusters* están formados por dos o tres microARNs y otros por decenas. Por otro lado, esta clasificación no considera la función de los microARNs, sino que los agrupa por similitud de secuencia (familia) o de regulación (*cluster*).

Finalmente, para poder estudiar la función que tienen los microARNs se han desarrollado

distintas estrategias basadas en la modulación de su expresión. Por un lado para sobreexpresar el microARN se puede utilizar un oligonucleótido sintético denominado *mimic-microRNA*, el cual mimetiza la secuencia del pre-microARN o el microARN maduro, aumentando ectópicamente su expresión. Mientras que para los ensayos de pérdida de función, se utilizan generalmente los oligonucleótidos antisentido *miR-inhibitors* o LNAs. Ambos oligos consisten en una molécula de ARN modificada que se une complementariamente al microARN de manera muy estable, bloqueando así la interacción con sus genes *target*. Sin embargo, si bien estas estrategias son utilizadas comúnmente, presentan como limitante que regulan los microARNs de manera individual. Por ende, dado que en muchos casos los microARNs forman parte de familias/*clusters* relacionados funcionalmente, se han desarrollado diferentes estrategias que permiten regular simultáneamente los miembros de un mismo grupo. Por ejemplo, se han diseñado transgenes denominados "esponjas" que codifican sitios de unión a los distintos microARNs de la familia/cluster. De esta manera, el oligonucleótido diseñado "secuestra" los microARNs disminuyendo la interacción de éstos con sus genes *target* (Ebert et al., 2013; Kluiver et al., 2012). Otra estrategia se basa en el diseño de una variante de los LNAs, la cual es más pequeña y sólo consiste en una secuencia de 8 nucleótidos complementaria a la secuencia semilla. De esta manera, estos "*tiny LNAs*" reconocen y regulan a todos los microARNs de una misma familia (Obad et al., 2013). Además, se han desarrollado diferentes estrategias basadas en la técnica de edición génica CRISPR/Cas9, mediante la cual se pueden mutar tanto *clusters* de microARNs policistronicos (Chang et al., 2016) como también familias de microARNs (Narayanan et al., 2016).

En conclusión, dado que en muchos casos los microARNs están relacionados funcionalmente con otros microARNs formando distintas familias y *clusters*, es importante considerar en el estudio del miRNoma de las células el comportamiento de éstos grupos de microARNs.

1.3.5. IsomiRs: Variantes de los microARNs

Mediante el avance en las técnicas de secuenciación masiva y el desarrollo de programas cada vez más sofisticados para analizar los resultados de secuenciación, se ha logrado identificar variantes de los microARNs. Previamente la resolución de las técnicas no alcanzaba para detectarlas o se las consideraba errores de la secuenciación *per se* (Nielsen et al., 2012). A estas isoformas se las denominó isomiRs y se determinó que en algunos casos los niveles de

expresión de los isomiRs son incluso equivalentes al de los microARNs canónicos y pueden estar asociados a la maquinaria de silenciamiento RISC (Cloonan et al., 2011).

Los isomiRs son pequeñas moléculas de ARN doble cadena que difieren ligeramente en la secuencia con respecto a su microARN canónico. Estas diferencias pueden estar en el extremo 5' (iso-5'), en el extremo 3' (iso-3') o incluso dentro de la secuencia como SNPs (*Single Nucleotide Polymorphism*). Tanto los iso-5' como los iso-3' pueden variar en la composición de la secuencia o en el largo. Se ha postulado que las diferencias en el largo de las secuencias se debe al clivaje impreciso de las endonucleasas Drosha y Dicer durante el procesamiento del microARN hacia su forma madura. Pero dado que los iso-3' son más abundantes que los iso-5', se cree que deben existir mecanismos adicionales que regulan la aparición de estas modificaciones. A su vez, otros mecanismos han sido propuestos, tal como el agregado de nucleótidos mediante enzimas *nucleotidyl transferases* las cuales no requieren de un templado y generalmente agregan uracilos y adenosinas en los extremos 3' (su actividad polimerasa es en dirección 5'-3'). Otro mecanismo que explicaría la modificación de los extremos 3', se basa en la acción de enzimas exoribonucleasas 3'-5' que podrían "recortar" los microARNs ya cargados en el complejo RISC, como por ejemplo la enzima *Nibbler* en *Drosophila* (Neilsen et al., 2012). Por otro lado, también se han detectado modificaciones dentro de las secuencias de los microARNs, las cuales podrían deberse a mecanismos de edición como por ejemplo la deaminación de adenosina en inosina. Este tipo de edición de ARN ha sido muy estudiado en distintos organismos y detectado tanto en secuencias codificantes como en no codificantes, particularmente en secuencias repetitivas invertidas doble cadena *Alu* y en precursores de los microARNs (pri- y pre-microARNs). Este proceso de edición es llevado a cabo por la familia de enzimas ADAR (*Adenosin Deaminase Acting on RNA*), de las cuales se conocen tres genes en vertebrados (ADAR1, ADAR2 y ADAR3) y se han determinado distintos mecanismos de regulación, tanto transcripcionales como post-transcripcionales (Nishikura, 2016). Estas enzimas se unen a los precursores doble cadena de los microARNs, generan modificaciones A-I y como consecuencia afectan los sitios de reconocimiento de las enzimas encargadas de la maduración y correcto procesamiento de los mismos. En el cerebro se ha identificado que el 20 % de los pri-microARNs sufren ediciones A-I (Nishikura, 2016). Además, ciertas modificaciones A-I pueden no afectar la biogénesis del microARN pero sí alterar la secuencia del mismo y el posterior reconocimiento de los mensajeros blanco (Nishikura, 2016).

El hecho de que exista una gran diversidad de isomiRs, de secuencias específicas, presentes

en distintos tejidos y con perfiles de expresión característicos, ha abierto la posibilidad de que tengan una función en las células. Es por esto que diversos mecanismos han sido postulados acerca del rol que los isomiRs tienen. Por un lado, se cree que las modificaciones en el extremo 5' podrían afectar a la secuencia semilla, alterando así el repertorio de genes blanco. Por ejemplo, en el trabajo de Tan y col. se estudió un isomiR en particular, el isomiR-9-1, el cual es cargado al complejo RISC y silencia *in vitro* la expresión de genes distintos a su microARN canónico (Tan et al., 2014). A su vez, modificaciones en el extremo 5' puede afectar la selección de la hebra que es cargada por AGO en el complejo RISC, generando así un gran impacto en los genes regulados (Neilsen et al., 2012). Sin embargo, muchas de estas modificaciones no llegan a alterar significativamente la afinidad de los genes blanco y en consecuencia el microARN canónico y sus isomiRs tienen en común muchos ARNm permitiendo que la regulación de estos sea más efectiva (Guo and Liang, 2016). Por otro lado, si bien las modificaciones en el extremo 5' son las que tendrían un mayor impacto en la regulación de genes, los iso-3' son más abundantes y se cree que podrían afectar tanto la interacción de la hebra *sense/antisense* con AGO como la estabilidad y el tiempo de vida del microARN, afectando de esta manera la función y efecto de éstos en la regulación de la expresión génica (Neilsen et al., 2012). Además, se ha demostrado que los isomiRs interactúan al igual que los microARNs con distintas moléculas de ARN no codificante, tales como pseudogenes, lncRNAs y ARN circulares, los cuales en muchos casos actúan como esponjas de microARNs bloqueando la acción de los mismos (Guo and Liang, 2016).

Dado la importante participación que pueden tener los isomiRs en la red de regulación de los microARNs es necesario que futuros trabajos busquen comprender mejor los mecanismos mediante los cuales los isomiRs ejercen su función. En la actualidad, ya se han desarrollado algunos programas bioinformáticos que permiten analizar la presencia y niveles de expresión de los isomiRs a partir de resultados de secuenciación masiva. Mediante estos programas se ha logrado detectar las isoformas de los microARNs que típicamente eran descartados del análisis debido a los algoritmos utilizados. Por ejemplo, un *pipeline* muy usado en el análisis de *RNA-Seq* es el CAP-miRSeq, el cuál utiliza el *software* miRDeep para el mapeo de los *reads* al genoma de referencia (Sun et al., 2014). En este mapeo se descartan aquellas secuencias que presentan más de un *missmatch* y una extensión mayor a dos nucleótidos en el extremo 5' y mayor a cinco en el extremo 3'. Todas las demás secuencias que cumplen con estos requisitos son considerados un mismo microARN, sin tener en cuenta la existencia de isoformas. Sin embargo,

en los últimos años se han desarrollado programas que permiten estudiar a los isomiRs, tal como el IsomiR, el Chimira y el isomiR-SEA (Pantano et al., 2009; Vitsios and Enright, 2015; Urgese et al., 2016). Estos programas aceptan en el análisis la presencia de *missmatchs* en las secuencias (con ciertas restricciones) y determinan los niveles de expresión tanto de los microARNs como de los isomiRs. El programa Chimira, a diferencia del IsomiR y el isomiR-SEA, permite distinguir en el análisis la presencia de modificaciones A-I generadas por ADARs.

1.4. Rol de los microARNs en la pluripotencia y la diferenciación celular

Los microARNs son reguladores post-transcripcionales de diversos procesos biológicos, entre los cuales se ha determinado que cumplen un rol esencial en el mantenimiento de la pluripotencia y en el proceso de diferenciación celular. Tal como se mencionó al principio de la sección, ratones mutantes para la enzima *Dicer* generan letalidad embrionaria y CME murinas con deficiencia en esta endonucleasa también presentan un fenotipo con alteraciones tanto en la diferenciación celular como en ciclo celular (Wang et al., 2010). A su vez, CME mutantes para DGCR8 presentan complicaciones en la diferenciación celular y los marcadores de pluripotencia tales como Oct-4, Nanog, Rex1 y Nanog presentan un retraso en la disminución de sus niveles de expresión. Estas evidencias, demuestran la importancia que tienen los microARNs en la regulación de la pluripotencia y en la capacidad de éstas células para diferenciarse.

En las CME humanas, la familia del miR-302 es una de las más estudiadas y su rol es fundamental en la regulación del mantenimiento de la pluripotencia. Esta familia esta formada por los microARNs miR-302a, miR-302b, miR-302c, miR-302d, miR-302e y miR-302f. Los dos últimos no se encuentran expresados en embrionarias, y se desconocen sus funciones. Además, los primeros cuatro se encuentran agrupados en un *cluster* de aproximadamente mil bases localizado en el cromosoma 4, y en el cual está también presente el miR-367. Se ha determinado que esta familia forma un *loop* positivo de regulación con factores clave de pluripotencia como Oct-4, Nanog y Sox-2, los cuales se unen al promotor de la familia miR-302 y activan su transcripción. A su vez, esta familia de microARNs participa de la regulación del ciclo celular, silenciando reguladores del ciclo como la ciclina D1 y promoviendo así la transición G1/S. Dada la gran importancia de ésta familia en las CME, se ha estudiado y demostrado que la sobreexpresión de la misma favorece la reprogramación de células somáticas a células pluripotentes inducidas (Ong et al., 2015). Otros microARNs que presentan una expresión alta en

las células indiferenciadas son la familia del miR-371, el *cluster* C19MC, el *cluster* de miR-17/92 y el miR-106 (Laurent et al., 2008). La desregulación de algunos de estos microARNs ha sido relacionada también con ciertos tipos de cáncer, tal como el *cluster* de miR-17/92 que cumpliría un rol importante en el progreso de distintos tipos de tumores (Mogilyansky and Rigoutsos, 2013). Asimismo, ciertos microARNs participan de la regulación del inicio de la diferenciación celular, tales como el miR-145 que regula a los genes Oct-4, Klf y Sox2 y la familia de let-7 y los miR-125 y miR-181 (Ong et al., 2015).

Una vez iniciada la diferenciación celular, los microARNs cumplen un rol esencial en el control post-transcripcional de determinados genes del desarrollo que permiten establecer los diferentes tipos celulares. Se han encontrado microARNs que participan en el desarrollo de neuronas, células hematopoyéticas, osteocitos, condrocitos, adipocitos, células endoteliales y cardiomiocitos. En la figura I9 se puede observar una lista de algunos microARNs relacionados con la diferenciación hacia los distintos linajes.

Con respecto al desarrollo cardíaco, varios microARNs han sido estudiados en profundidad y se ha logrado determinar el rol que éstos cumplen durante la especificación cardíaca. Durante las etapas tempranas, en la formación del mesodermo, ciertos microARNs regulan la expresión de genes involucrados en la transición epitelio-mesenquimal. Por ejemplo, la familia del miR-200, formada por el miR-200a, miR-200b, miR-200c, miR-429 y miR-141, es bien conocida por regular genes inductores de la TEM tales como Snail1, Zeb1 y Zeb2 tanto en células indiferenciadas como en células tumorales (Gill et al., 2011; Park et al., 2008; Gregory et al., 2008a). Una vez iniciada la diferenciación, Gill y col. han demostrado que el inhibidor de E-cadherina, Snail1 reprime la expresión de la familia del miR-200. Además, se ha determinado que ésta familia también participaría durante la reprogramación celular, del proceso inverso de la TEM, conocido como transición mesénquima-epitelial (TME), en la cual reprime a los inhibidores de E-cadherina (Wang et al., 2013).

A su vez, el miR-205 ha sido identificado junto con la familia del miR-200 como regulador clave de la expresión de los factores activadores de la TEM, Zeb1, Zeb2 y TGF β 2 en el contexto tumoral. La desregulación de la expresión del miR-205 ha sido relacionada con el desarrollo de tumores en el cáncer de mama, cáncer de ovario, cáncer de esófago y cáncer de vejiga (Gregory et al., 2008b). Sin embargo, el rol que cumplen estos microARNs dependería de cada contexto celular, dado que en muchos casos se ha observado que la disminución de sus niveles de expresión favorece el desarrollo tumoral, y otras veces el aumento de su expre-

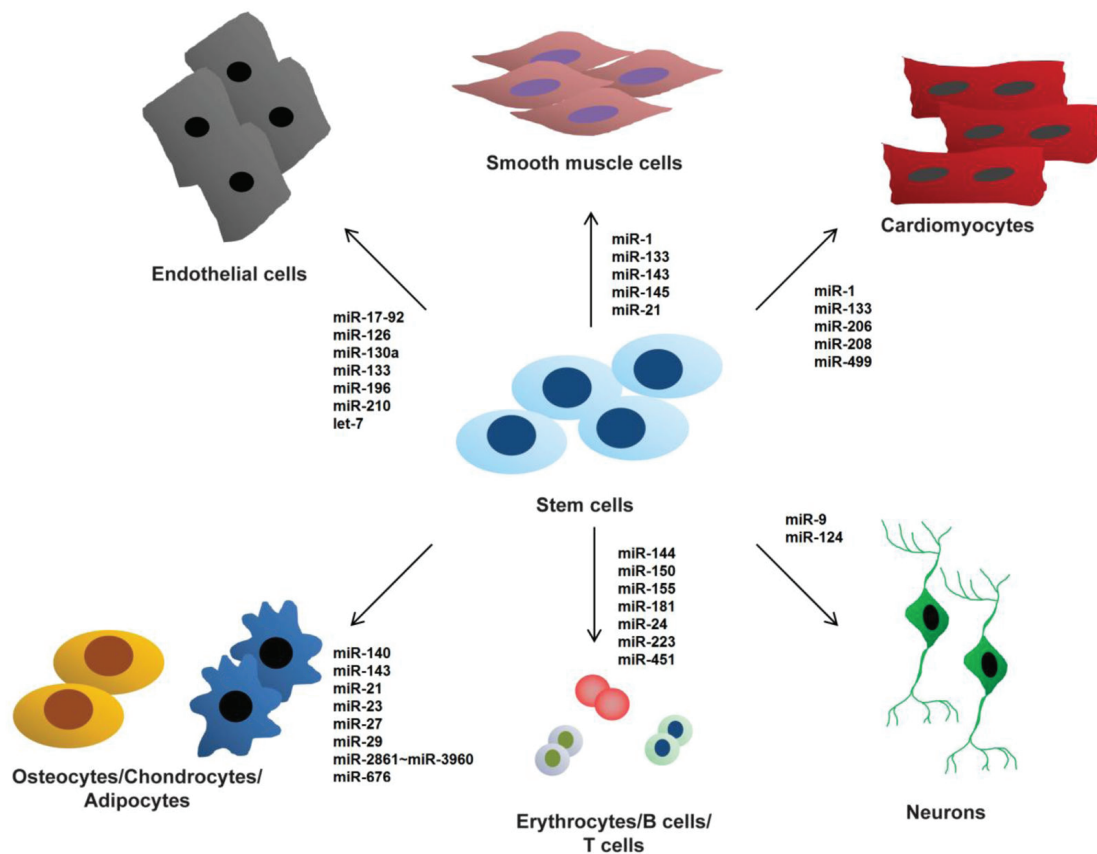


Figura I9: MicroARNs involucrados en la diferenciación celular: En la figura se puede observar un lista con algunos de los microARNs que participan de la diferenciación hacia los distintos tipos celulares. Figura adaptada de (Ong et al., 2015).

sión es lo que favorece el avance de la enfermedad. El efecto dual del miR-205 ha sido incluso reportado en la diferenciación celular, ya que por un lado se demostró que en las células indiferenciadas presenta niveles altos de expresión, pero en el trabajo de Li y col. se demostró que la sobreexpresión del miR-205 favorece la diferenciación hacia un progenitor de endodermo extraembionario y este aumento activa genes relacionados con la adhesión y migración celular tales como Cdh2 (N-Cadherina), Cdh5, Cdh6 y Cdh11 (Li et al., 2013). Experimentos futuros en el estudio del rol que cumple el miR-205 en la diferenciación celular permitirán comprender mejor su función.

Por otro lado, determinados microARNs han sido identificados como reguladores clave de la diferenciación cardíaca. Los principales identificados son el miR-1, el miR-133, el miR-206, miR-208 y miR-499. El miR-1 fue el primer microARN identificado en el desarrollo cardíaco. Tanto el miR-1 como el miR-133 inhiben la diferenciación a endodermo y ectodermo, promo-

viendo la formación de mesodermo (Ong et al., 2015). En particular el miR-1, junto con el miR-206, promueven la diferenciación de los mioblastos en miotubos, mientras que el miR-133 promueve la proliferación de los mioblastos (Townley-Tilson et al., 2010). A su vez, se ha determinado que el miR-1 participa de la maduración electrofisiológica de los cardiomiocitos, mientras que el miR-499 dirige la diferenciación cardíaca de CMEh hacia la generación de cardiomiocitos ventriculares (Ong et al., 2015). El miR-208 esta codificado en un intrón del gen de la cadena pesada de miosina (α MHC) y participa de la adaptación asociada al estrés en el corazón.

El rol fundamental que los microARNs cardíacos tienen en el desarrollo del corazón y en su correcto funcionamiento se refleja también en la participación que éstos tienen en distintas cardiopatías, en las cuales la desregulación de la expresión de alguno de ellos tiene efectos severos en el desarrollo de la patología. Es por esto que varios microARNs cardíacos son blancos terapéuticos o incluso biomarcadores del estadio de enfermedades cardiovasculares (Barwari et al., 2016). Por ejemplo, la expresión disminuida del miR-133 genera hipertrofia en el corazón la cual es restituida con la sobreexpresión del mismo, los niveles anormales del miR-1 tiene efectos en electrofisiología del corazón, y el miR-208 regula el balance entre α y β MHC siendo que un aumento del isotipo β se asocia con una mala adaptación en respuesta al estrés y una contractilidad reducida.

Dada la importancia de los microARNs en las enfermedades cardiovasculares, distintos trabajos han estudiado la posibilidad de modular la expresión de los mismos como posibles tratamientos. En la actualidad, ciertos microARNs están siendo evaluados para el desarrollo de terapias, como por ejemplo el miR-92a, cuya inhibición reduce la inflamación endotelial y promueve la angiogénesis y la recuperación del miocardio isquémico, o como por ejemplo la inhibición del miR-133 y miR-21 que reduce la fibrosis y mejora el funcionamiento cardíaco en modelos de ratón con infarto de corazón (Barwari et al., 2016). Por último, dada la gran estabilidad y la presencia de los microARNs en sangre, actualmente se considera analizar los niveles de microARNs en sangre para el diagnóstico temprano de infarto de miocardio. Se ha determinado que la evaluación de un sólo microARN no es lo suficiente preciso para el diagnóstico, pero sí el análisis simultáneo de un grupo de microARNs. Se cree que incluso se podría mejorar el diagnostico obtenido mediante el tradicional modelo de riesgo de *Framingham* (Barwari et al., 2016).

En conclusión, estos resultados demuestran la importancia que tiene poder comprender

mejor la participación de los microARNs en el desarrollo cardíaco tanto para conocer mejor los mecanismos involucrados en la regulación de la cardiomiogénesis como para el potencial uso de los microARNs en la medicina traslacional.

HIPÓTESIS Y OBJETIVOS

2. Hipótesis y Objetivos

A lo largo de la introducción se describieron las características de los microARNs y su participación en la regulación post-transcripcional de un gran número de genes relacionados con diversos procesos biológicos, entre ellos el desarrollo embrionario. En particular, el correcto funcionamiento y biogénesis de los mismos es necesario para la cardiomiogénesis, siendo que alteraciones en estos provoca fenotipos patológicos o incluso letalidad embrionaria. En este sentido, en el presente trabajo nos propusimos analizar y estudiar los microARNs involucrados en la diferenciación cardíaca a partir de células pluripotentes humanas, con el objetivo principal de delucidar la participación que los microARNs tienen tanto en el mantenimiento del estado de indiferenciación como en el proceso de diferenciación celular.

La hipótesis del proyecto se basa en que los microARNs estarían involucrados tanto en la formación del progenitor temprano de mesodermo como en el establecimiento del linaje cardíaco durante la diferenciación celular a partir de células pluripotentes humanas, y formarían complejas redes de regulación de vías de señalización fundamentales tanto para el mantenimiento de la pluripotencia como para la diferenciación celular. En base a ésta hipótesis, los objetivos que se desprenden son los siguientes:

- **Objetivo 1:** El primer objetivo de este trabajo es establecer el modelo de estudio, el cual consiste en tres poblaciones celulares bien definidas y relacionadas con la diferenciación cardíaca *in vitro*. Estas poblaciones corresponden a células pluripotentes, un progenitor de mesodermo temprano y cardiomiocitos derivados de pluripotentes. A partir de éstas poblaciones celulares nos proponemos obtener los microARNs y estudiar el perfil de expresión de los mismos. En particular nos proponemos generar y caracterizar las tres poblaciones celulares a partir de la expresión de genes específicos de cada una y analizar la expresión de microARNs ya descritos en la bibliografía a modo de poner a punto el protocolo de extracción y el análisis de los mismos.
- **Objetivo 2:** El segundo objetivo se basa en el análisis de los microARNs presentes en cada una de las tres poblaciones mediante secuenciación masiva y determinar su dinámica de expresión. Nos proponemos estudiar y establecer el miRNoma característico de la diferenciación cardíaca a partir de las células pluripotentes humanas, teniendo en cuenta la clasificación funcional de los microARNs en familias y *clusters*. Además, nos

proponemos analizar la presencia y abundancia de isomiRs y compararlos entre las tres poblaciones celulares.

- Objetivo 3: Además, nos proponemos estudiar el *cluster* de microARNs C19MC, exclusivo de primates, para entender el rol que cumple en las células pluripotentes humanas y en la diferenciación celular. Para esto, analizaremos de manera predictiva con programas bioinformáticos los posibles genes regulados por los microARNs del *cluster* y luego de manera experimental modulando la expresión de los microARNs. Por un lado, nos proponemos modular la expresión de algún/os microARNs mediante el uso de oligos sintéticos que permiten aumentar/disminuir la expresión de los mismos y por otro lado nos proponemos eliminar el *cluster* entero mediante edición génica con CRISPR/Cas9 y evaluar el efecto del *knock-out* del C19MC en las células pluripotentes.
- Objetivo 4: Por último, nos proponemos analizar el rol del miR-205 en las células pluripotentes y en la diferenciación hacia el progenitor temprano de mesodermo. El objetivo es entender la posible relación que tiene el miR-205 en el proceso de transición epitelio mesenquimal (TEM) que sufren las células pluripotentes en el desarrollo del mesodermo, y comparar los resultados con la función que tiene el microARN en la TEM durante el desarrollo tumoral.

RESULTADOS

3. Resultados

Capítulo 1:

Diferenciación a mesodermo y al linaje cardíaco

El primer objetivo de este trabajo fue estudiar e identificar los microARNs involucrados en la diferenciación cardíaca a partir de las células pluripotentes humanas. Para ello, analizamos la expresión de estos pequeños ARNs mediante secuenciación masiva (*small RNA-Seq*) en tres poblaciones celulares bien definidas: células madre pluripotentes, un progenitor temprano de mesodermo y cardiomiocitos derivados de CMP.

Las células madre pluripotentes humanas (H9) fueron crecidas sobre una matriz comercial compuesta por una mezcla de proteínas que semejan la composición de la matriz extracelular, y cultivadas en un medio definido que permitió mantener su capacidad de pluripotencia. Las células fueron mantenidas en estas condiciones durante muchos pasajes, manteniendo una alta tasa de división y aspecto morfológico característico de células indiferenciadas. En el laboratorio se analiza una vez por semana mediante PCR la presencia de *Mycoplasma sp*, ya que esta contaminación puede afectar sustancialmente la pluripotencia de las células. En la figura 1 se observa una colonia típica de CMEh crecida sobre Geltrex y el análisis de la expresión del marcador de pluripotencia Oct-4, mediante citometría de flujo. A partir de estas células aislamos los ARN pequeños para el análisis por secuenciación masiva. Publicaciones previas de nuestro grupo han validado la pluripotencia de esta línea celular y la capacidad de diferenciarse al linaje cardíaco (Romorini et al., 2016; Blugüermann et al., 2017).

1.1. Diferenciación al progenitor de mesodermo temprano

Para obtener el progenitor de mesodermo temprano, las células fueron crecidas en adhesión sobre Geltrex y en un medio suplementado con distintos factores de crecimiento durante 3.5 días. El progenitor de mesodermo temprano (PM) fue identificado por primera vez en el trabajo de Evseenko en el 2010 (Evseenko et al., 2010) y se caracteriza por haber sufrido la

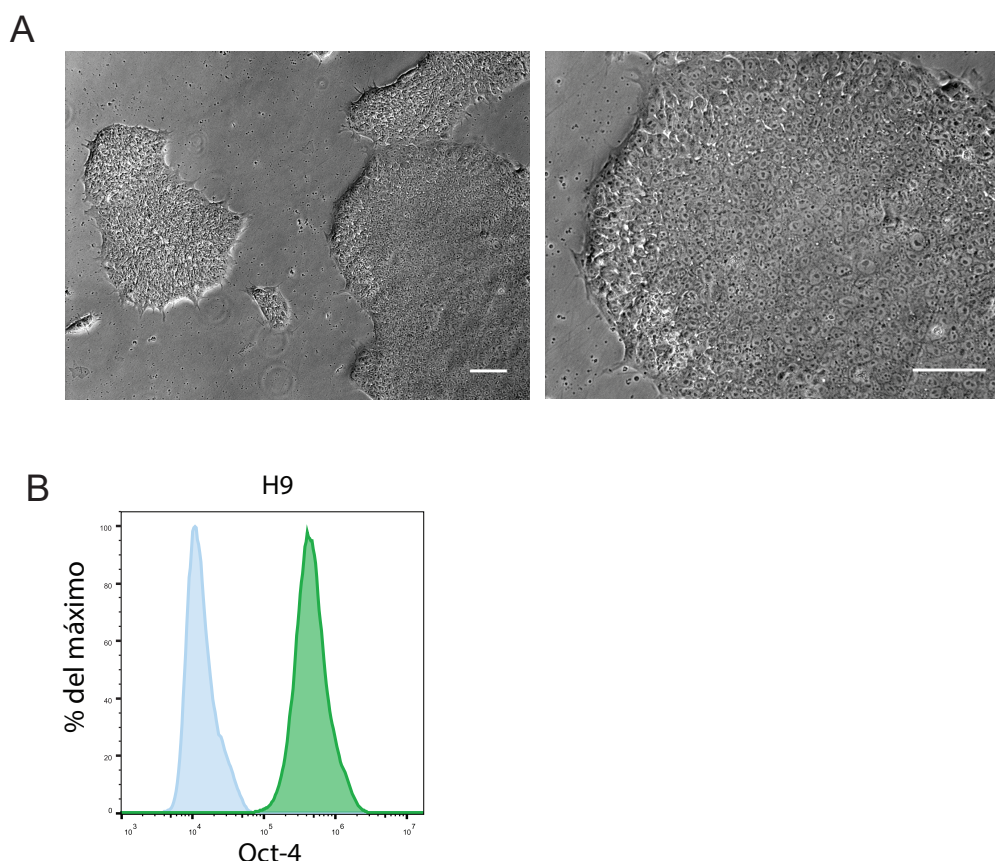


Figura 1: Colonias de células madre embrionarias humanas: (A) En las imágenes se observan colonias de las células madre embrionarias humanas crecidas sobre Geltrex y cultivadas en un medio definido. El aspecto compacto y bordes irregulares son características típicas de las células en estado de indiferenciación. La imagen de la derecha fue tomada con un aumento 10X y la de la izquierda con un aumento 20X (la barra representa $100\mu\text{m}$). (B) Análisis por citometría de flujo de la expresión de la proteína Oct-4. En verde se muestra el histograma correspondiente a las células marcadas con un anticuerpo primario α -Oct-4 y en azul el control sin el anticuerpo primario. El promedio de las células positivas para Oct-4 fue de $99,1\% \pm 0,4$, considerando tres réplicas biológicas.

transición epitelio-mesenquimal y por la capacidad de generar todos los tejidos que derivan del mesodermo (el linaje cardiovascular, mesenquimal y hemoendotelial). La formación del PM se determinó mediante el análisis de expresión por citometría de flujo de dos marcadores de membrana, EpCAM (CD326) y NCAM (CD56). Tal como se puede observar en la figura 2, en el día 3.5 de iniciada la diferenciación disminuye la expresión de CD326 (molécula característica de las células indiferenciadas) y aumenta la expresión de CD56. El porcentaje esperado del PM es minoritario y varía entre un 10 % y un 20 %. Para obtener la muestra de los ARN pequeños del PM, se aisló la población CD56+/CD326- por FACS (*Fluorescence Activated Cell Sorting*)

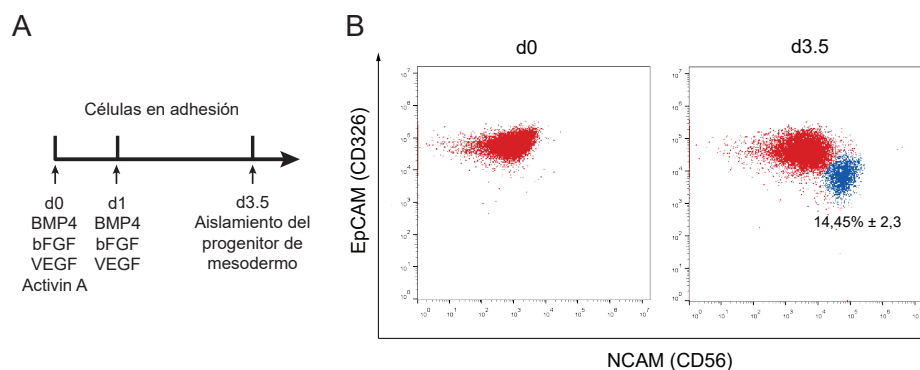


Figura 2: Diferenciación al progenitor de mesodermo temprano: (A) Esquema del protocolo de diferenciación a mesodermo a partir de las células pluripotentes. Las concentraciones de cada uno de los factores se detallan en la sección de materiales y métodos. (B) Mediante citometría de flujo se monitoreó la formación del progenitor de mesodermo temprano. El *dot plot* de la izquierda corresponde al día 0 de la diferenciación y el de la derecha al día 3.5. En rojo se muestra la población CD326+/CD56- y en azul el progenitor (CD326-/CD56+). El porcentaje corresponde al promedio de tres réplicas biológicas $\pm$ el SEM.

y se extrajo la fracción de ARN pequeños de las mismas.

A partir del PM aislado por FACS, se analizó por PCR en tiempo real la expresión de genes característicos de mesodermo y mesodermo cardíaco para evaluar así el compromiso de estas células hacia este linaje (figura 3A). Como se puede observar en la figura, los genes *Mixl1*, *Brachyury* y *Mesp1* aumentan significativamente sus niveles de expresión con respecto a las células en estado de indiferenciación (día 0), mientras que los genes *KDR* y *Nkx2.5* no cambiaron significativamente sus niveles de expresión. A su vez, dado que el PM ha sufrido la transición epitelio-mesenquimal (TEM), analizamos la expresión de distintos marcadores característicos de éste proceso por PCR en tiempo real. Tal como se puede observar en la figura 3B, el nivel de expresión de E-cadherina disminuye (principal molécula de adhesión célula-célula que se pierde la ocurrir la transición), y aumentan significativamente los niveles de expresión de los inhibidores de E-cadherina, *Zeb1*, *Zeb2*, *Snail1* y *Snail2* y la proteína estructural Vimentina, típico marcador de la TEM.

1.2. Diferenciación al linaje cardíaco

La población de cardiomiocitos fue obtenida mediante la diferenciación de las CMEh con un protocolo dirigido específicamente hacia el linaje cardíaco. En este caso, las células fueron

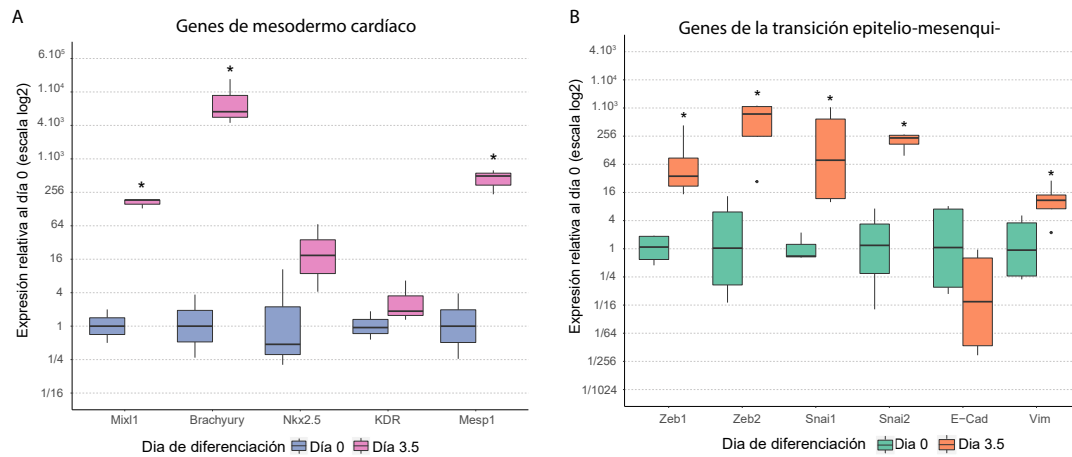


Figura 3: Análisis de la expresión de genes en el progenitor de mesodermo: Se analizó la expresión de genes específicos para el mesodermo cardíaco (A) y genes de la transición epitelio mesenquimal (B), mediante PCR en tiempo real. El nivel de expresión se relativizó con el día 0 y en el *boxplot* se muestra la mediana de al menos 4 replicas biológicas y sus respectivos cuartiles. Los asteriscos representan diferencias significativas con un $p < 0,05$.

agregadas de manera forzada por centrifugación para formar lo que se conoce como cuerpos embrioides (EB, en ingles *embryoid bodies*). Los EB fueron crecidos y mantenidos en suspensión durante los primeros cuatro días y luego en adhesión, en un medio suplementado con distintos factores que permiten dirigir la diferenciación primero hacia mesodermo y luego hacia el linaje cardíaco (figura 4). Luego de aproximadamente 9-10 días de diferenciación, se aprecia la aparición de zonas con contractilidad espontánea (como archivo suplementario hay un video en donde se observa un EB con actividad contráctil y expresión de GFP, ver video).

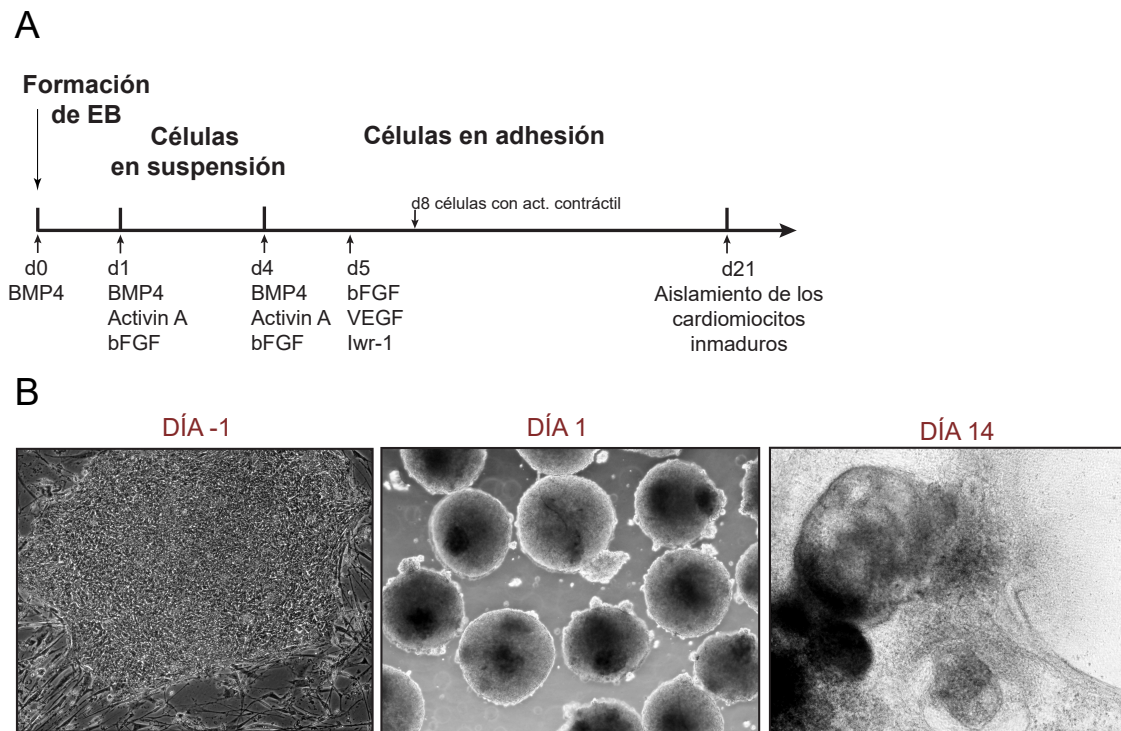


Figura 4: Diferenciación dirigida al linaje cardíaco: (A) Esquema del protocolo de diferenciación cardíaca a partir de células pluripotentes. En el día 0 del protocolo se forman los cuerpos embrioides (EB) y se crecen en suspensión durante los primeros 4 días, en un medio suplementado con los factores BMP4, Activina A y bFGF. A partir del día 4 se cultivan los EB en adhesión y en el día 5 se agregan al medio el inhibidor Iwr-1 y el factor de crecimiento VEGF. A partir del día 8 empiezan a aparecer las primeras regiones con actividad contráctil. (B) Las imágenes fueron tomadas a distintos días del protocolo de diferenciación y en ellas se observa en el día -1 una colonia de células embrionarias crecida sobre MEFi, los EB en suspensión a las 24hs luego de iniciada la diferenciación y por último un cuerpo en adhesión con actividad contráctil en el día 14.

A lo largo del protocolo de diferenciación, se tomó muestra de ARN para evaluar por PCR en tiempo real el establecimiento del linaje cardíaco. En la figura 5 se puede observar que los niveles de expresión de los genes de pluripotencia Oct-4 y Nanog disminuyen a partir del día 4 con respecto al día 0, los niveles de expresión de los genes de mesodermo Brachyury y Mesp1 aumentan principalmente en el día 4, al igual que la expresión de los genes cardíacos MYH6, NKX2.5 y cTNT (este último aumenta significativamente recién en el día 14) (figura 5). El análisis del perfil de expresión de los genes estudiados nos permitió confirmar el establecimiento del linaje cardíaco a partir de el protocolo de diferenciación utilizado.

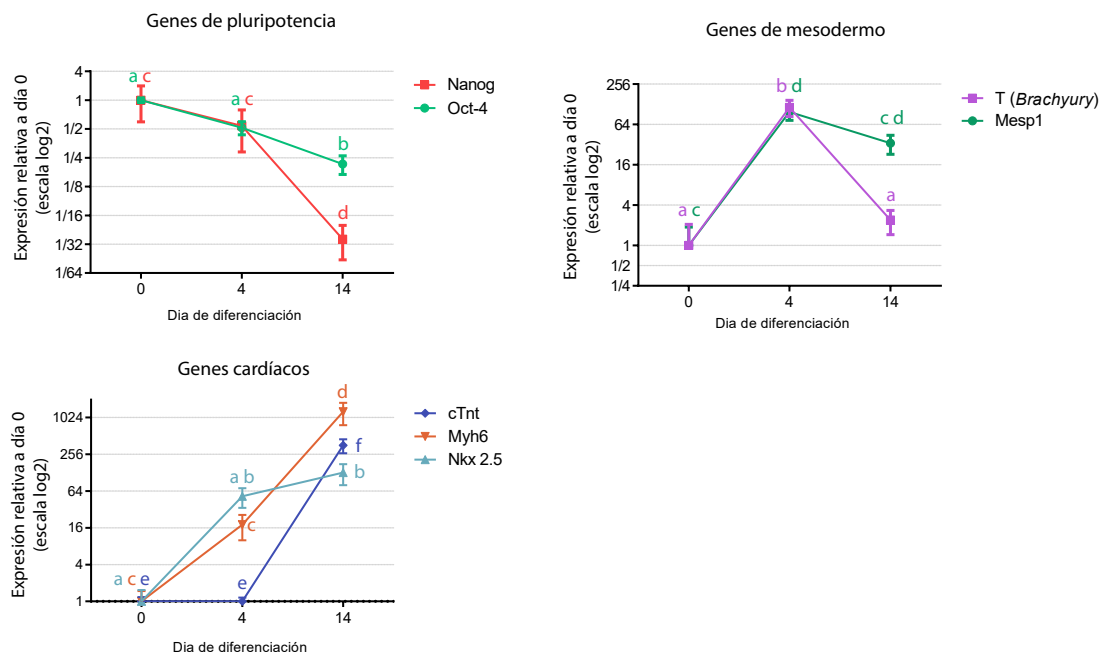


Figura 5: Niveles de expresión de los genes a lo largo de la diferenciación cardíaca: Se determinó por PCR en tiempo real los niveles de expresión de los genes de pluripotencia (Nanog y Oct-4), genes de mesodermo (Brachyury y Mesp1) y genes cardíacos (cTnt, Myh6 y Nkx2.5) en los días 0, 4 y 14 de iniciada la diferenciación. En los gráficos se muestra la media $\pm$ el SEM para al menos tres réplicas biológicas. Las letras representan diferencias significativas entre los grupos con un p-valor <0.05 .

A su vez, analizamos los niveles de expresión de las proteínas cardíacas en el día 14 de la diferenciación mediante inmunofluorescencia (figura 6). Tal como se puede observar en las imágenes, los niveles de expresión de las proteínas Nkx2.5, cTNT y MYH6 son principalmente altos en las región con estructura tridimensional que presentaban actividad contráctil.

Por otro lado, analizamos los perfiles de expresión de determinados microARNs a lo largo de la diferenciación cardíaca. Para esto, fue necesario poner a punto el análisis de expresión de los microARNs maduros por PCR en tiempo real. Es importante destacar que al ser moléculas de ARN pequeñas no fue posible utilizar los *primers* clásicos de retro-transcripción de ARNm (*random primers* y oligodT) sino que fue necesario diseñar *primers* específicos para la retro-transcripción de cada uno de ellos. Estos *primers* se denominan *Stem Loop primers* (SLO) y consisten en una región complementaria específica para un microARN (6 nucleótidos) y una región universal que forma la estructura de *loop* (ver esquema en la figura 7A). De esta manera, analizamos a lo largo de la diferenciación cardíaca la expresión de la familia miR-302,

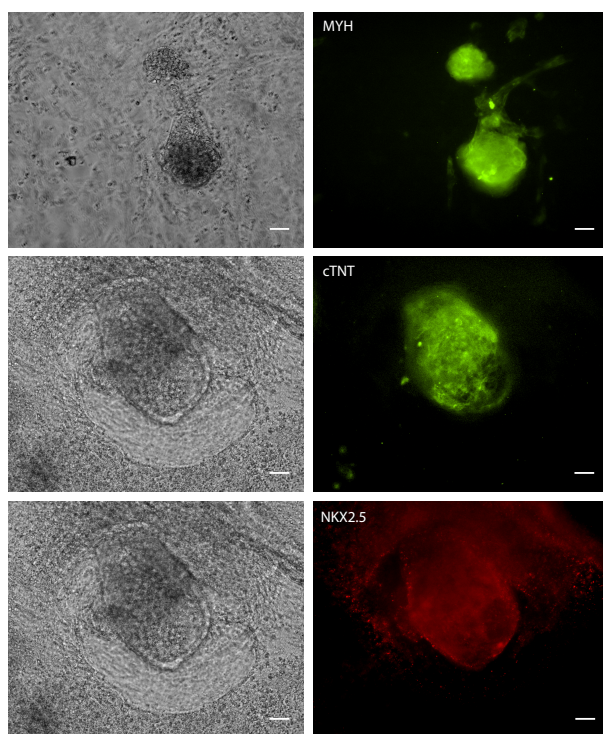


Figura 6: Caracterización de marcadores cardíacos durante la diferenciación cardíaca: Mediante inmunofluorescencia se analizó la expresión de las proteínas del linaje cardíaco MYH6, cTNT y NKX2.5 en el día 14 de iniciada la diferenciación. La barra representa 500 μ m.

del miR-145, del miR-296 y de determinados microARNs específicos de células cardíacas (miR-1, miR-133, miR-206 y miR-208) (figura 7C). Además, para confirmar que el producto de la PCR en tiempo real era efectivamente el microARN estudiado, corrimos el producto de la PCR en un gel de agarosa 3 %, aislamos la banda, la clonamos en un vector p-GEM-T y analizamos su identidad por secuenciación. Esto fue realizado solo para uno de los microARNs, el miR-1, a modo de validación de la retro-transcripción (figura 7B).

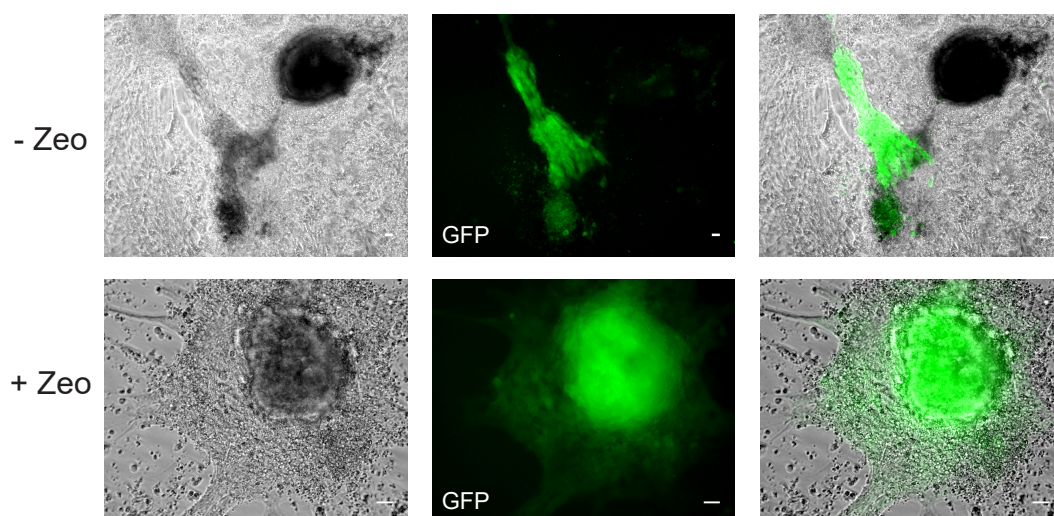


Figura 8: Selección de células cardíacas: En las imágenes se puede observar el enriquecimiento en células que expresan GFP bajo el promotor de cTNT, seleccionadas mediante dos pulsos de 48hs con zeocina. La barra representa una distancia de 500 μ m.

A partir del protocolo de diferenciación obtuvimos un cultivo muy heterogéneo y complejo en cuanto a la composición celular, por lo cual fue necesario enriquecer el cultivo en células cardíacas previo a la extracción de los ARN pequeños. Para esto, la línea que utilizamos expresa bajo el promotor de c-Tnt la resistencia a puromicina y el gen de GFP (del inglés, *Green Fluorescent Protein*). En la figura 8 se observa la expresión de GFP en dos protocolos, uno sin tratamiento con zeocina (-Zeo) y otro post enriquecimiento con el antibiótico (+Zeo). En las imágenes se puede observar la disminución de la cantidad de células negativas para GFP que rodean los cuerpos con actividad contráctil.

En el día 21 de iniciada la diferenciación y luego del tratamiento con zeocina, se analizó el porcentaje de células cardíacas mediante citometría de flujo. Mediante la selección con antibiótico se logró obtener un cultivo con aproximadamente un 90% de células GFP + post enriquecimiento (figura 9). En general, sin el enriquecimiento con antibiótico se obtiene una eficiencia entre 20-30% de células cardíacas. Una vez logrado enriquecer el cultivo en la población de cardiomiocitos, se realizó la extracción de los ARN pequeños para la secuenciación masiva.

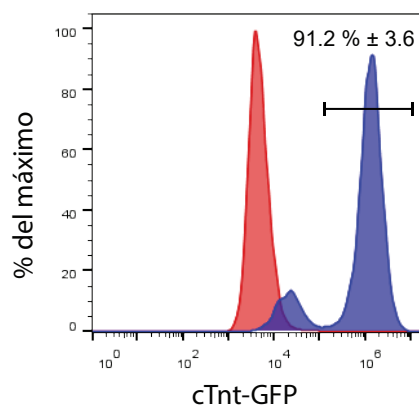


Figura 9: Diferenciación al linaje cardíaco: En el día 21 de iniciada la diferenciación cardíaca, se analizó por citometría de flujo el % de células GFP + (bajo el control del promotor de c-Tnt). El histograma rosa corresponde a las células en día 0 y el violeta al día 21. El porcentaje corresponde al promedio de tres réplicas biológicas $\pm$ el SEM.

Conclusiones parciales del capítulo:

En este primer capítulo, describimos las distintas poblaciones celulares a partir de las cuales realizamos posteriormente el análisis de expresión de los microARNs. En primer lugar, describimos la población de células pluripotentes, las cuales presentaron un aspecto morfológico característico de las células indiferenciadas así como también un nivel alto de expresión de los marcadores de pluripotencia Oct-4 y Nanog.

En segundo lugar, diferenciamos las células pluripotentes hacia un progenitor temprano de mesodermo, el cual se caracterizó por un aumento en la expresión de la molécula de membrana N-CAM (CD56) y una disminución de la molécula Ep-CAM (CD326). A su vez, determinamos mediante el análisis de expresión génica que el progenitor temprano de mesodermo presentó niveles altos de expresión de genes característicos de mesodermo, así como también de marcadores de la TEM. Estos resultados nos permitieron confirmar el compromiso hacia el linaje mesodérmico del progenitor obtenido mediante el protocolo de diferenciación y aislado por *cell sorting*.

Por último, diferenciamos las células hacia el linaje cardíaco a partir de un protocolo de diferenciación dirigido y obtuvimos aproximadamente un 90 % de células positivas para cTNT, luego de seleccionar los cardiomiocitos con antibiótico. Para caracterizar el compromiso de las

células hacia el linaje cardíaco, analizamos a lo largo del proceso de diferenciación la expresión de genes de pluripotencia (Oct-4 y Nanog), genes de mesodermo (Mesp1 y Brachyury) y genes cardíacos (cTnt, Myh6 y Nkx2.5). De esta manera caracterizamos el proceso de diferenciación cardíaca utilizado en el presente trabajo y determinamos que durante el proceso de diferenciación los genes marcadores del linaje cardíaco presentan un patrón de expresión característico de ese linaje.

Por otro lado, analizamos a lo largo de la diferenciación cardíaca el perfil de expresión de determinados microARNs reportados previamente en la bibliografía, con el fin de caracterizarlos en nuestro modelo de estudio. De ésta manera, estudiamos el perfil de expresión de la familia del miR-302, el miR-145, el miR-296 y de los microARNs cardíacos miR-1, miR-133, miR-208 y miR-206, durante el proceso de diferenciación y obtuvimos resultados consistentes con lo reportado en trabajos previos.

Finalmente, aislamos los ARN pequeños de las tres poblaciones descritas y estudiamos mediante secuenciación masiva el perfil de expresión de los microARNs para así poder definir el miRNoma que caracteriza a cada una de las poblaciones.

Capítulo 2:

Identificación del miRNoma relacionado con la diferenciación al progenitor de mesodermo y al linaje cardíaco

2.1. Análisis del perfil de expresión de los microARNs en las células pluripotentes, el progenitor de mesodermo y los cardiomiocitos

Mediante secuenciación masiva de los ARN pequeños analizamos el miRNoma de las células pluripotentes, del progenitor temprano de mesodermo y de los cardiomiocitos, realizandolo en tres replicas biológicas de cada población celular. A partir de ésta parte del trabajo haremos referencia a cada una de las poblaciones en base a las siglas de sus nombres en ingles, células pluripotentes (PSC, *pluripotent stem cells*), el progenitor de mesodermo temprano (MPC, *mesoderm progenitor cells*) y los cardiomiocitos (CM, *cardiomyocytes*).

Los resultados de la secuenciación fueron analizados mediante el *pipeline* CAP-miRSeq (Sun et al., 2014). Brevemente, a partir de los datos crudos de la secuenciación se analizó la calidad de los datos mediante el programa de uso libre *FastQC* antes y después de eliminar los adaptadores (utilizando el programa *Cutadapt*). Una vez controlada la calidad de los datos y haber superado los distintos umbrales, se "mapearon" las secuencias ("reads") con un genoma de referencia (hg19) para poder determinar así la identidad de los microARNs secuenciados y su abundancia, mediante el programa miRDeep2. Este programa, permite además, identificar posibles microARNs nuevos en base a la región del genoma en la que "mapearon" las *reads* y la posibilidad de formar la típica estructura de *hairpin*. Finalmente, se determinó la expresión diferencial entre las poblaciones mediante el paquete de R *DESeq2* (análogo al paquete *edgeR*). Para identificar otras moléculas de ARN y variantes de microARNs, se utilizó el programa *Bowtie* para el "mapeo" y se cuantificó la abundancia mediante el programa *HTSeq* y *GATK* (figura 10).

A partir del análisis de los resultados, determinamos que el 42 % de las *reads* corresponden

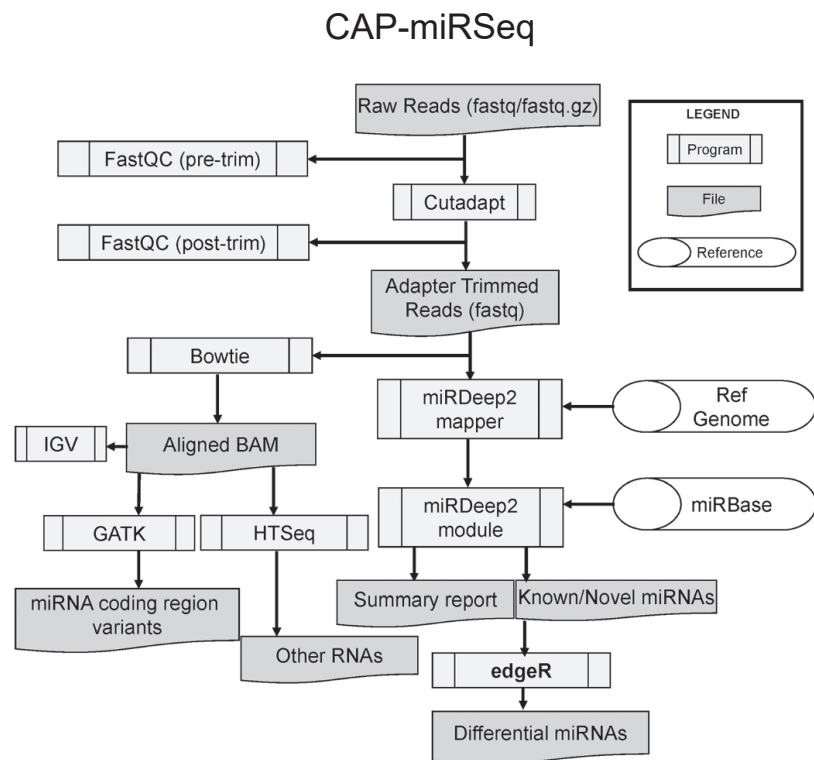


Figura 10: Análisis de los resultados de secuenciación masiva mediante el CAP-miRSeq pipe-line: En el esquema se detallan los pasos y programas utilizados en el análisis de los resultados crudos de la secuenciación masiva de los pequeños ARN. Figura adaptada de (Sun et al., 2014).

a microARNs, y el porcentaje restante corresponde a otras moléculas de ARN como por ejemplo ARNr, ARNt, snoARN y lincARN, entre otras (figura 11).

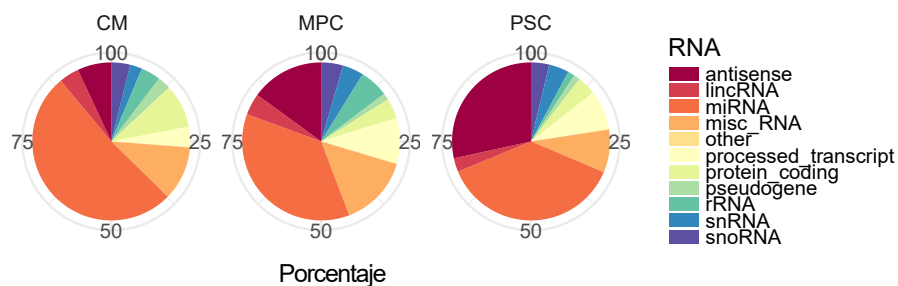


Figura 11: Abundancia de los distintos tipos de ARN: Mediante secuenciación masiva identificamos distintos tipos de moléculas de ARN, en los gráficos de torta de muestran los porcentajes correspondientes a cada tipo de ARN para las tres poblaciones CM, MPC y PSC.

Identificamos 694 microARNs expresados en al menos una de las poblaciones, teniendo en cuenta el criterio arbitrario de que el microARN tenga una expresión de al menos 5 *reads* en cada una de las replicas y en al menos una de las poblaciones celulares estudiadas (la lista completa de microARNs se detalla en la tabla suplementaria 1, ver tabla) (figura 12A). Cabe destacar que en la actualidad ya se han reportado al rededor de 2800 microARNs en el genoma humano. Los microARNs identificados presentaron un perfil de expresión específico en cada una de las poblaciones celulares, lo cual nos permitió agruparlos en 5 "clusters" con un perfil de expresión determinado (figura 12B). Al observar dichos "clusters", identificamos que hay determinados microARNs que son exclusivos de las PSC ("cluster" 1), otros que aumentan sus niveles de expresión conforme las células se diferencian ("cluster" 2), otros que aumentan su expresión en el MPC y se mantienen *up-regulados* en los CM ("cluster" 3), otros que aumentan en el MPC y luego disminuyen en los CM ("cluster" 4) y por último, otros microARNs aumentan exclusivamente en los CM ("cluster" 5) (en la tabla suplementaria 2 se detallan los microARNs de cada "cluster", ver tabla).

A partir de un análisis de expresión diferencial entre las poblaciones de a pares, pudimos identificar los microARNs que aumentan y los que disminuyen sus niveles de expresión en cada caso. En la figura 13 en cada *volcano plot* se muestran los nombres de los microARNs que presentaron la mayor diferencia en sus niveles de expresión entre las dos poblaciones analizadas en cada caso. Tal como se puede observar, existe un mayor número de microARNs diferencialmente expresados entre las PSC y MPC con respecto a los CM, que entre las PSC y el MPC entre sí. Este resultado podría deberse a que los cardiomiocitos son células que aparecen más tardíamente en el desarrollo celular, es decir, son células más maduras que el progenitor de mesoderma, mientras que éste último está temporalmente más próximo en el desarrollo a las células pluripotentes. En las tablas 1 y 2 se detallan los microARNs con mayor nivel de expresión en las PSC con respecto a CM y en los CM con respecto a las PSC, respectivamente (en las tablas suplementarias 3, 4 y 5 se detalla la lista completa de los microARNs de cada *volcano plot*, ver tablas).

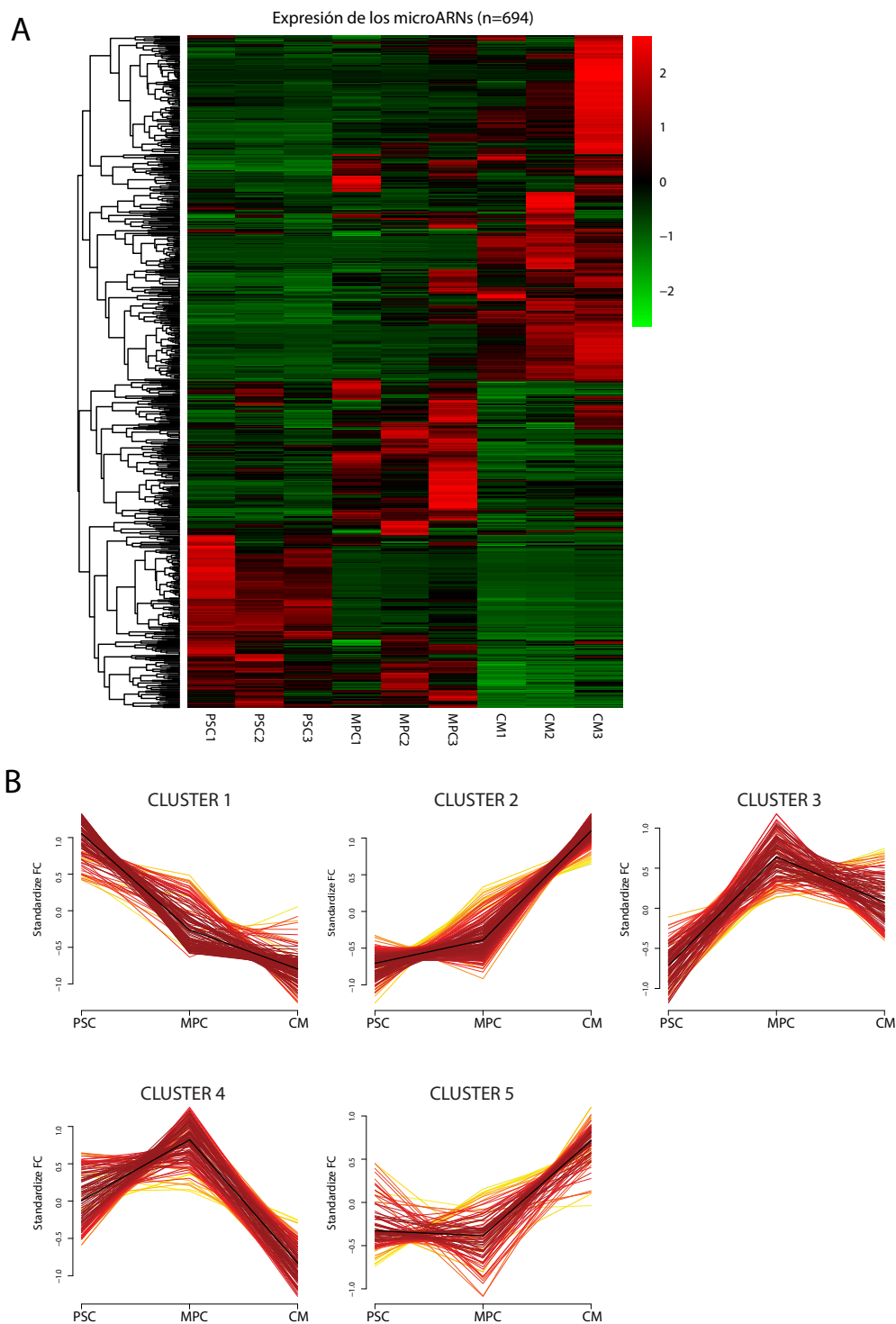


Figura 12: Niveles de expresión de los microARNs en las tres poblaciones celulares: (A) En el *heatmap* se puede observar los niveles de expresión de los 694 microARNs identificados por secuenciación masiva. En rojo están representados los niveles de expresión altos y en verde los bajos. (B) En los *fuzzy plots* se pueden observar los 5 "clusters" que representan los 5 perfiles de expresión de los microARN, según sus niveles de expresión en cada población celular.

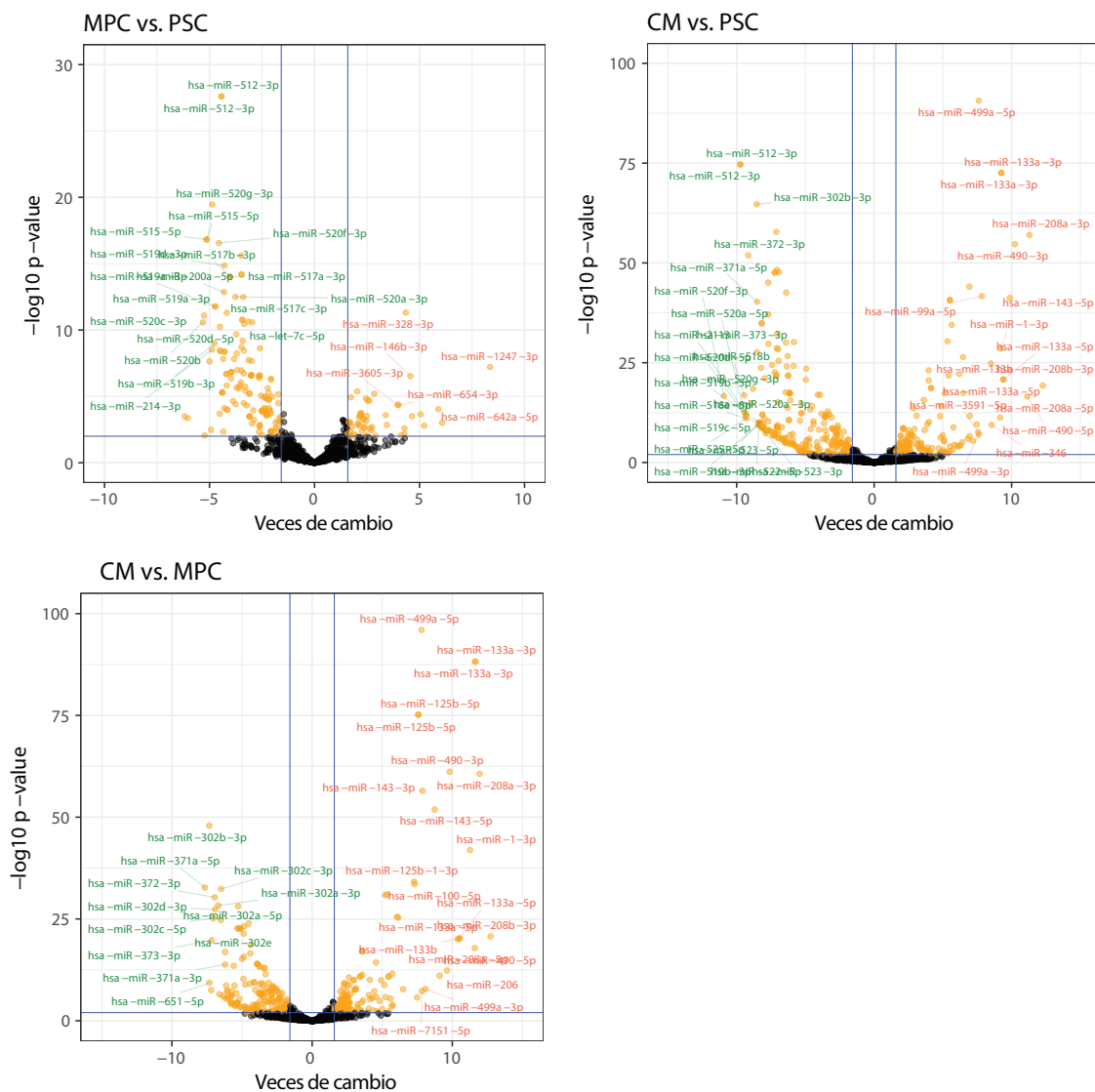


Figura 13: Análisis diferencial de la expresión de los microARNs: En los *volcano plots* se muestra el análisis diferencial de los niveles de expresión de los microARNs al comparar las poblaciones de a pares. En rojo y verde están representados los microARNs que aumentan y disminuyen sus niveles de expresión respectivamente, que presentaron un $\log_2$ de las veces de cambio mayor a 3 y un p-valor <0.01 .

microRNA	PSC1	PSC2	PSC3	CM1	CM2	CM3	log2FC
hsa-miR-302b-3p	314847	376199	427597	1969	1901	1363	-8.55
hsa-miR-302a-5p	163784	213498	141767	3357	2231	835	-6.98
hsa-miR-372-3p	109988	17394	43904	126	209	224	-9.15
hsa-miR-302d-3p	22718	33794	18820	376	353	223	-7.08
hsa-miR-302a-3p	17769	15461	20310	211	166	131	-7.70
hsa-miR-302c-3p	17042	15771	16136	196	118	102	-7.70
hsa-miR-92a-3p	10865	16838	8793	5889	4798	4973	-1.96
hsa-miR-92a-3p	10523	16060	8442	5446	4487	4575	-1.96
hsa-miR-20a-5p	9663	9099	10467	1153	1791	2088	-3.44
hsa-miR-363-3p	10751	8047	9329	1227	674	1023	-4.09

Tabla 1: microARNs diferencialmente más expresados en PSC con respecto a CM: En cada columna se muestran las replicas de PSC y CM con los valores de expresión de los microARNs (*reads*). En la última columna se muestran las veces de cambio en escala log2 (log2FC).

microRNA	PSC1	PSC2	PSC3	CM1	CM2	CM3	log2FC
hsa-miR-143-3p	966	524	457	158524	134668	113107	6.93
hsa-miR-100-5p	655	276	584	87971	98908	37292	6.47
hsa-miR-30d-5p	4215	3747	3322	26154	30870	30846	2.13
hsa-miR-99a-5p	43	69	30	20866	20682	10214	7.84
hsa-miR-22-3p	2084	3294	1339	8564	13724	22444	1.83
hsa-miR-125a-5p	603	1108	831	6504	13054	22746	3.08
hsa-miR-27b-3p	1191	861	906	12334	9900	10577	2.67
hsa-miR-125b-5p	125	168	91	5410	12738	13794	5.51
hsa-miR-125b-5p	121	162	88	5282	12550	13631	5.51
hsa-miR-499a-5p	25	29	27	7328	10437	11278	7.62

Tabla 2: microARNs diferencialmente más expresados en CM con respecto a PSC: En cada columna se muestran las replicas de PSC y CM con los valores de expresión de los microARNs (*reads*). En la última columna se muestran las veces de cambio en escala log2 (log2FC).

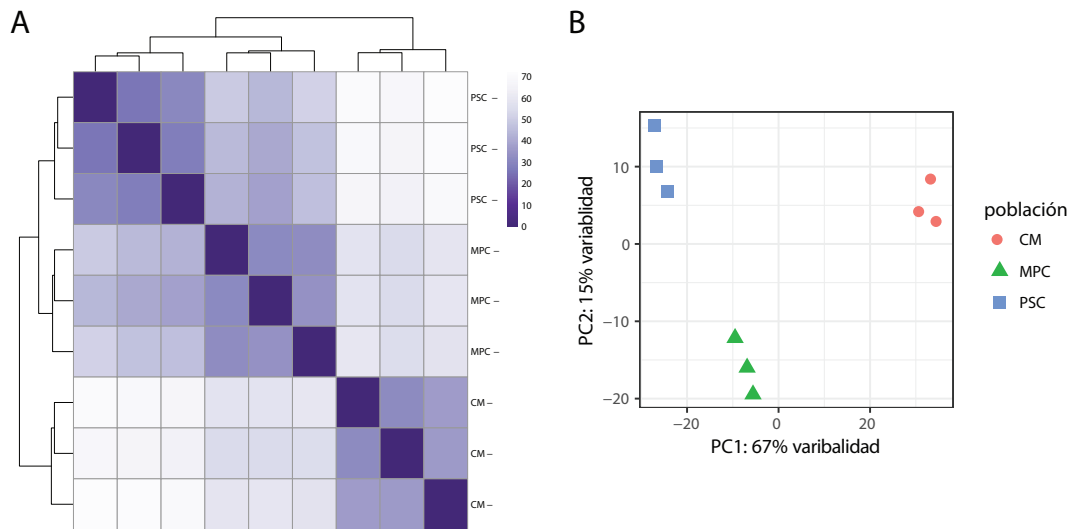


Figura 14: Heatmap de distancia y PCA plot: (A) En el *heatmap* está representada la varianza que existe en la expresión de los microARNs entre las tres poblaciones y sus replicas biológicas. Una mayor intensidad de color se corresponde con menor variabilidad entre las muestras. (B) En el gráfico de PCA se muestra cómo se agrupan las poblaciones y sus respectivas réplicas, según la similitud que existe entre ellas.

Las similitudes y diferencias, en cuanto a la expresión de los microARNs entre las tres poblaciones celulares, también se puede apreciar en el *heatmap* de distancias y en el gráfico de componentes principales (PCA) de la figura 14. Tal como se observa en la figura las PSC se agrupan más próximas a el MPC que a los CM, en coherencia con lo observado en los *volcano plots*. Además, en el gráfico de PCA se puede visualizar que las tres replicas de cada población se agrupan entre sí, validando de ésta manera los resultados y las replicas biológicas. Por último, se puede observar en el gráfico de PCA el porcentaje alto de discriminación (67% y 15 %) con dos variables y que permite una alta discriminación entre las poblaciones celulares.

2.2. Análisis de los isomiRs

Tal como se explicó en la introducción, en los últimos años ha aumentado la evidencia de que existe un gran número de variantes de microARNs, denominadas isomiRs, que difieren tanto en la composición como en el largo de las secuencias con respecto a las secuencias canónicas reportadas en la base de datos miRbase. Sin embargo, el programa utilizado para determinar la identidad de las *reads* luego de la secuenciación de ARN pequeños, miRDeep2, no contempla aquellas secuencias que poseen un número de *mismatches* mayor a uno y una extensión mayor a dos nucleótidos en el extremo 5' y mayor a cinco en el extremo 3'. Toda secuencia que logre superar dichos parámetros son alineados y anotados como un mismo microARN, coincidente con las secuencias canónicas de miRBase. Teniendo esto en cuenta, decidimos re-analizar los datos con el programa *miraligner/IsomiR*, el cual nos permitió identificar los isomiRs y estudiar así su presencia en la diferenciación celular. A partir del análisis, determinamos el número de *reads* para los microARNs canónicos ("exactos") y para los distintos tipos de isomiRs. Tal como se puede observar en la figura 15, los isomiRs son muy abundantes, hasta incluso más abundantes que los microARNs canónicos. En la figura 15A se muestra la abundancia de *reads* sin filtrar, la cual fue de 6359 secuencias en total contemplando los microARN exactos y las variantes (luego de filtrar con un punto de corte mayor a 5 *reads* tal como en el análisis anterior, el número se redujo a 2809 reads en total, la lista completa se detalla en la tabla suplementaria 6, ver tabla). A su vez, determinamos que la abundancia de los isomiRs no varía significativamente entre las tres poblaciones, sino que se mantiene entre ellas ($p=0.18$, análisis de ANOVA) (figura 15B).

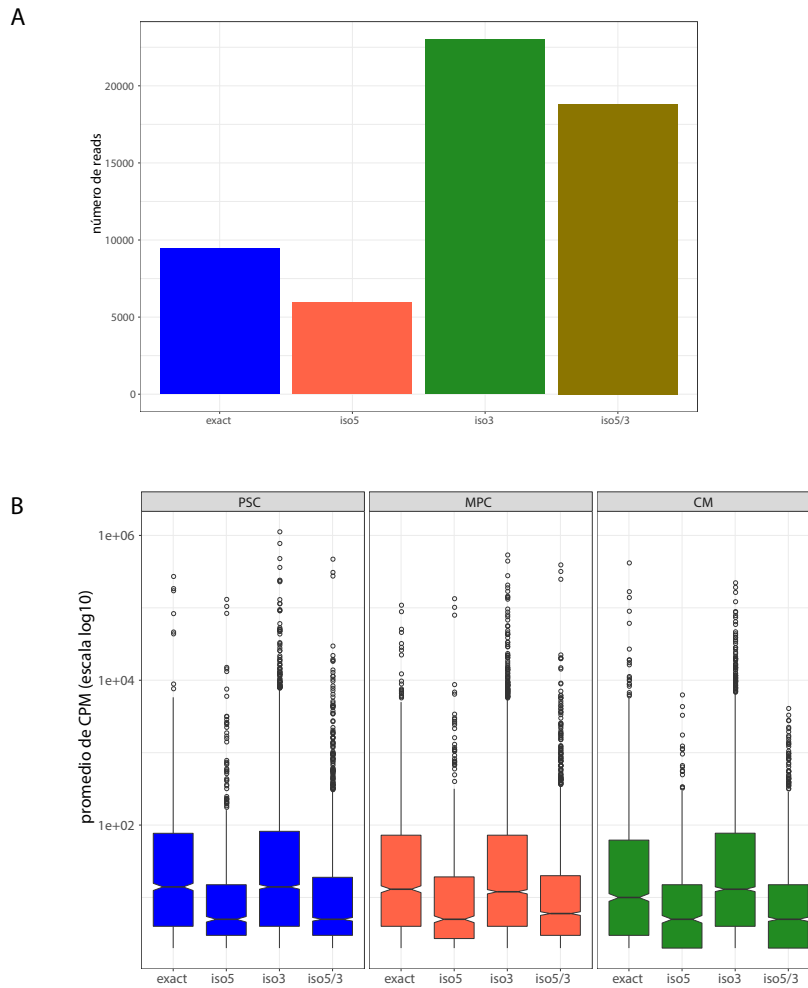


Figura 15: Abundancia y heterogeneidad de los isomiRs en la diferenciación cardíaca: (A) Número de *reads* determinado para los microARNs canónicos ("exacto"), isomiRs-5' (iso5), isomiRs-3' (iso3) e isomiRs mixtos (iso5/3), sin filtrar por el punto de corte (promedio de *reads* entre las tres replicas mayor a 5). (B) Abundancia de los distintos tipos de isoformas discriminado para cada población celular (PSC, MPC y CM). El número de *reads* esta expresado como "counts per millon" y en escala log10.

A continuación analizamos el número de isoformas en función de sus abundancias y la posición de los cambios a lo largo de la secuencia. En la figura 16A se observa que para los isomiRs-5', las isoformas más abundantes tienen cambios en la posición -3 con respecto a la secuencia "exacta" o canónica (es decir, tienen una delección de 3 nucleótidos), pero los cambios más frecuentes se dan en la posición -1. Por el contrario, para los isomiRs-3' el cambio más frecuente y más abundante se da en la posición -1 (es decir, delección de el último nucleótido).

Por otro lado, existen también sustituciones a lo largo de la secuencia ("SNPs"), los cuales son menos frecuentes y abundantes en comparación con las modificaciones en los extremos 5' y 3'. Al analizar la distribución de SNPs, se puede observar que la mayor cantidad de cambios se dan en los primeros 4 nucleótidos de la secuencia (figura 16B). Es importante destacar que estos cambios podrían afectar la secuencia semilla y generar así un efecto en la funcionalidad de los isomiR, ya que a través de la secuencia semilla los microARNs reconocen a sus genes blanco.

Dado que las modificaciones en el extremo 5' pueden alterar el sitio de reconocimiento de un microARN por sus genes *target*, decidimos analizar de manera *in silico* el efecto de dichas modificaciones en la secuencia semilla en un caso en particular. Para esto, elegimos analizar a modo de ejemplo el miR-302a y sus isomiRs-5', ya que la abundancia de los isomiRs era equivalente o incluso superior al microARN canónico. En la figura 17A se puede observar la variabilidad en la secuencia semilla del miR-302a-5p (alineamiento de arriba) y del miR-302a-3p (alineamiento de abajo). En ambos casos el microARN canónico (la primer secuencia del alineamiento) presentó un número de *reads* menor que varios de los isomiRs. Para analizar el efecto de las modificaciones en el reconocimiento de genes blanco, agrupamos las secuencias según las secuencias semillas y las nombramos como "exacta" a la semilla sin modificaciones con respecto al microARN canónico e iso-1, iso-2 e iso-3 para las isoformas tanto del miR-302a-5p y del miR-302a-3p. Tal como se puede observar en los diagramas de Venn (figura 17B), el miR-302a-5p y sus isomiRs comparten un determinado número de genes blanco predichos pero a su vez cada uno regula genes blanco específicos (328 para miR-302a-5p-"exacto", 167 para el iso2 y 15 para los iso1 e iso3). Un resultado similar obtuvimos en el análisis del miR-302a-3p y sus isomiRs (331 genes específicos para el miR-302a-3p "exacto", 81 para el iso2, 181 para el iso3 y 94 para el iso1). Luego, a partir de los genes blanco predichos exploramos los términos de *Gene Ontology* (GO) relacionados con los mismos. En la figura 18 se muestran los términos GO para los genes blanco del miR-302a-5p canónico ("exacto") (figura 18A) y los GO para los genes blanco del miR-302a-5p canónico más los genes de los isomiRs (18B). Tal como se puede observar, al agregar al análisis los genes blanco predichos para los isomiRs, aparecen nuevos términos GO que no aparecían en el análisis únicamente con el microARN "exacto", como por ejemplo la vía de Wnt. En la figura 19 se muestra un análisis equivalente para el miR-302a-3p. En este caso también aparecen algunos términos GO diferentes, tal como el desarrollo de dendritas y axones. Como conclusión, si bien el análisis fue exploratorio e *in*

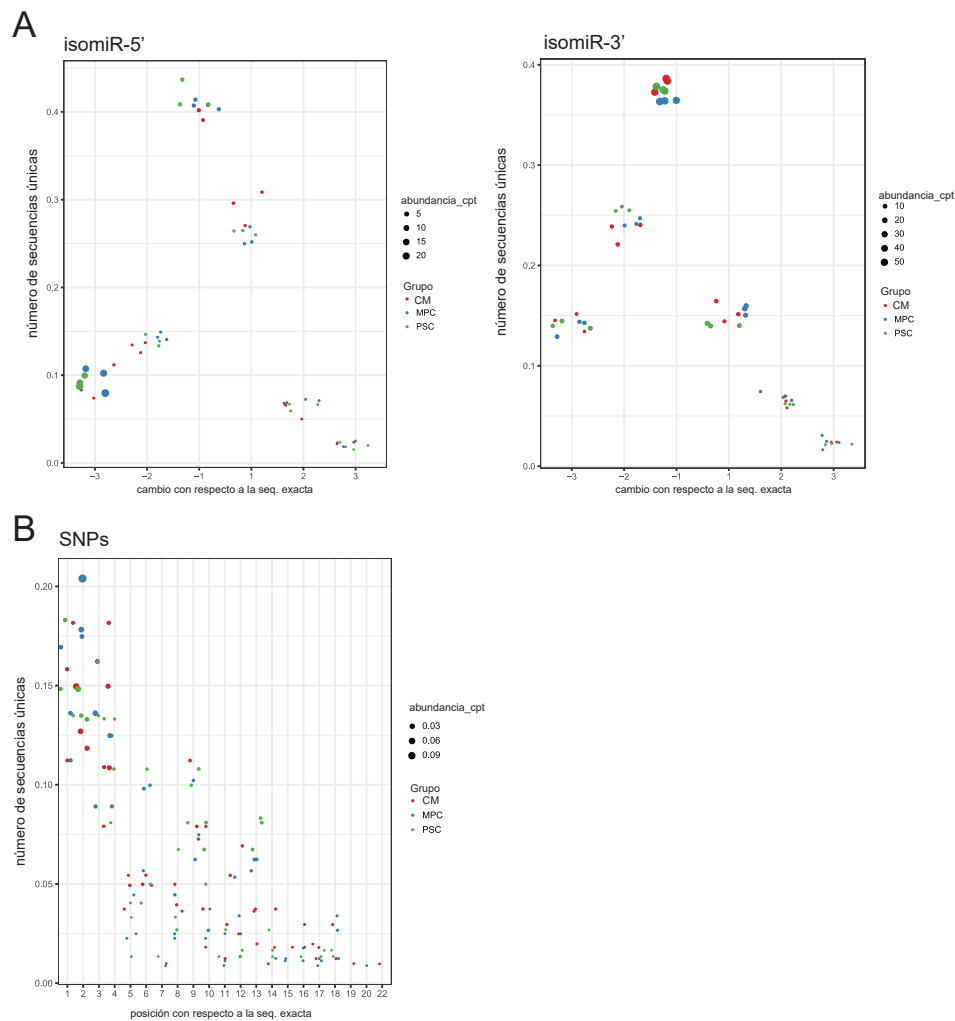


Figura 16: Análisis de los isomiRs: Posición y abundancia de las modificaciones. (A) En los *dot plots* se muestra la abundancia de las distintas isoformas según la adición o eliminación de tres bases con respecto al microARN canónico ("exacta"). El tamaño del punto representa la abundancia. El gráfico de la derecha corresponde a los isomiRs con modificaciones en el extremo 5' y el de la izquierda a las modificaciones en el extremo 3'. (B) Distribución y abundancia de los SNPs a lo largo de la secuencia. Cada color corresponde a una de las 3 poblaciones celulares.

silico, refuerza la idea de que los isomiRs no serían un error en la biogénesis de los microARNs, sino que podrían tener un rol funcional en las células y la regulación de diferentes procesos biológicos. A su vez, su función expandiría la red de regulación de la expresión de genes *target*.

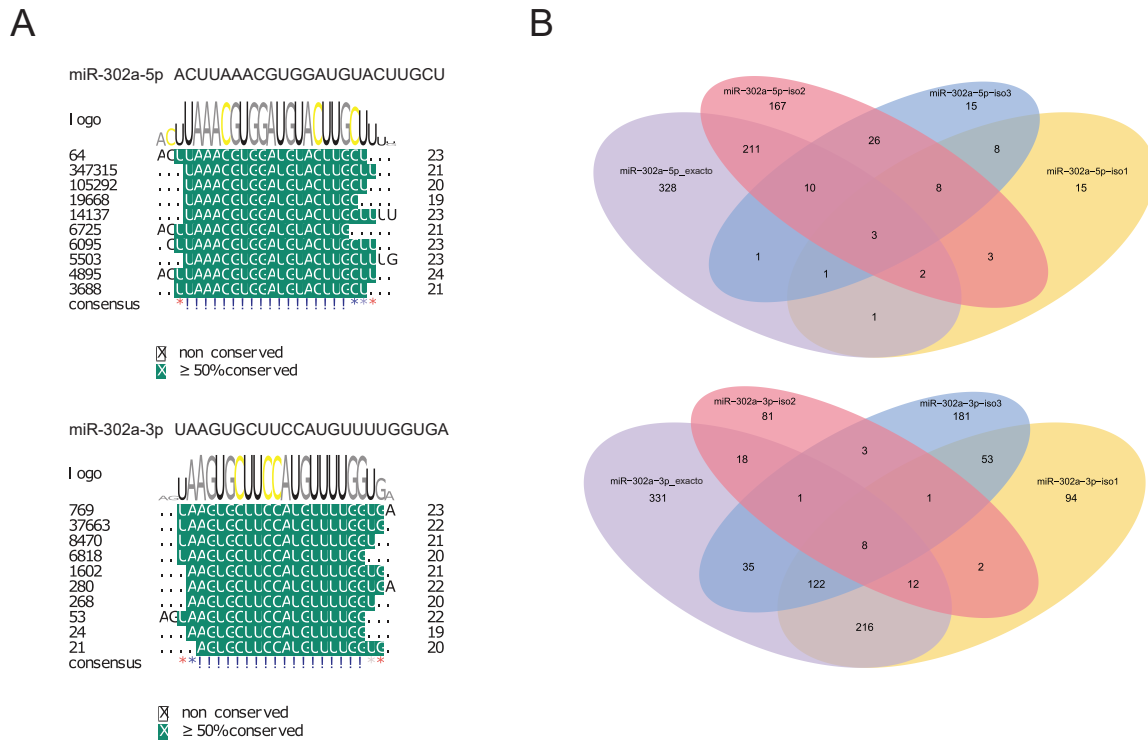


Figura 17: Análisis del miR-302a-5p y el miR-302a-3p y sus respectivos isomiRs: (A) Alineamiento múltiple de las secuencias del miR-302a-5p (arriba) y miR-302a-3p (abajo). A la izquierda de los alineamientos se muestra el número de *reads* para cada secuencia y arriba de los mismos está la secuencia canónica. En color verde se muestran las bases mas conservadas y sin color, las menos conservadas. La secuencia consenso ("logo") está representada con las bases en distintos tamaños según el grado de conservación de las mismas. (B) En los diagramas de Venn se muestra el número de genes blanco predichos para cada microARN y sus respectivos isomiRs. Los genes blanco fueron predichos mediante el *software* MR-microT.

Además de analizar las modificaciones en los extremos 5', 3' y los SNPs, analizamos mediante el programa *Chimira* las sustituciones del nucleótido adenosina por inosina (A-I) llevado a cabo por las enzimas ADARs (*Adenosine Deaminasa Acting on RNA*) (figura 20). A partir de los resultados, identificamos isomiRs con este tipo de edición a lo largo de toda la secuencia y de manera indistinta (figura 20A). Además, estas modificaciones fueron observadas en las tres poblaciones celulares, entre las cuales los cardiomiocitos (CM) presentaron un número de isomiRs con modificaciones A-I significativamente menor al de las células pluripotentes (p-valor=0.014) (figura 20B). A su vez, la abundancia de las modificaciones A-I fue similar en las tres poblaciones, aunque menor en comparación con la abundancia de los microARNs canónicos, tal como se ha descrito en trabajos anteriores (Nishikura, 2016) (fi-

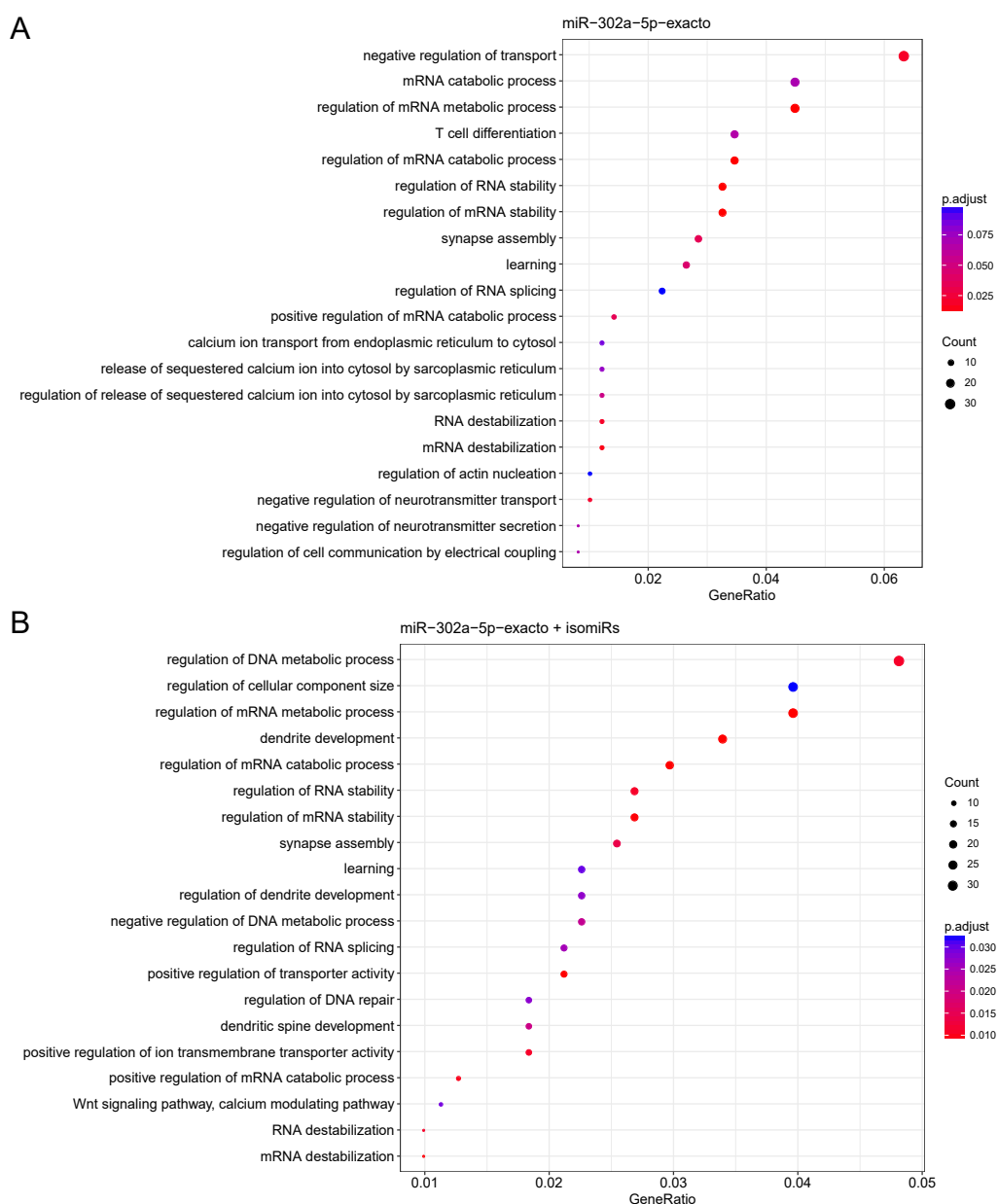


Figura 18: Análisis de Gene Ontology del miR-302a-5p e isomiRs: En gráfico están representados los primeros 20 términos de GO (con un p-valor <0.01) determinados para los genes blanco del miR-302a-5p "exacto" (A) y del miR "exacto" y sus isomiRs (B). El tamaño del *dot* corresponde a el número de genes relacionados con cada término GO, el color a el p-valor y en el eje x el GeneRatio hace referencia a el número de genes observados sobre los esperados.

gura 20C). Esto se ve claramente cuando se expresa el porcentaje de *reads* con respecto a la secuencia canónica. Sólo en algunos casos particulares estas isoformas superan el 1 % de las secuencias canónicas (figura 20D).

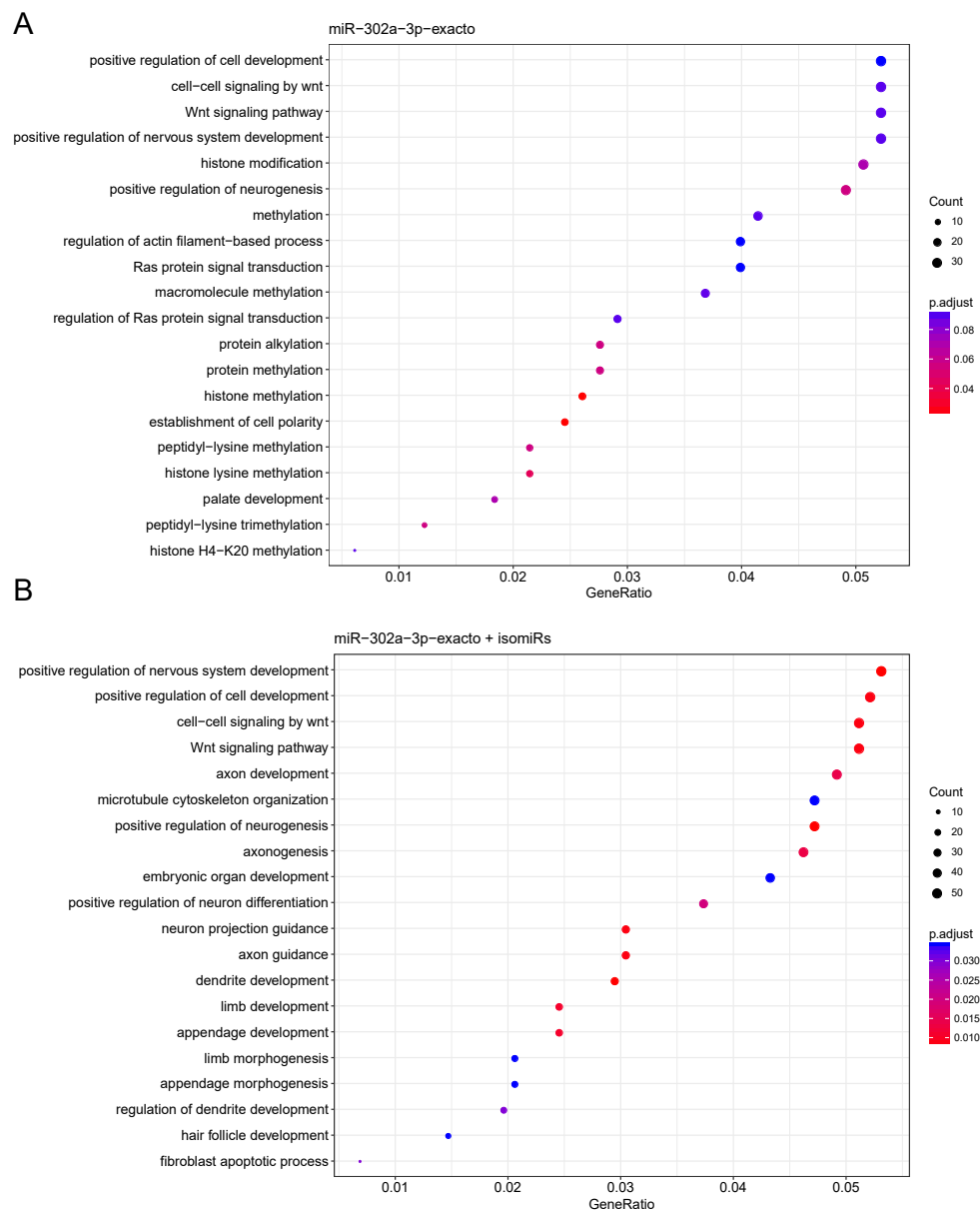


Figura 19: Análisis de Gene Ontology del miR-302a-3p e isomiRs: En gráfico están representados los primeros 20 términos de GO (con un p-valor <0.01) detmrinados para los genes blanco del miR-302a-3p "exacto" (A) y del miR "exacto" y sus isomiRs (B). El tamaño del dot corresponde a el número de genes relacionados con cada término GO, el color a el p-valor y en el eje x el GeneRatio hace referencia a el número de genes observados sobre los esperados.

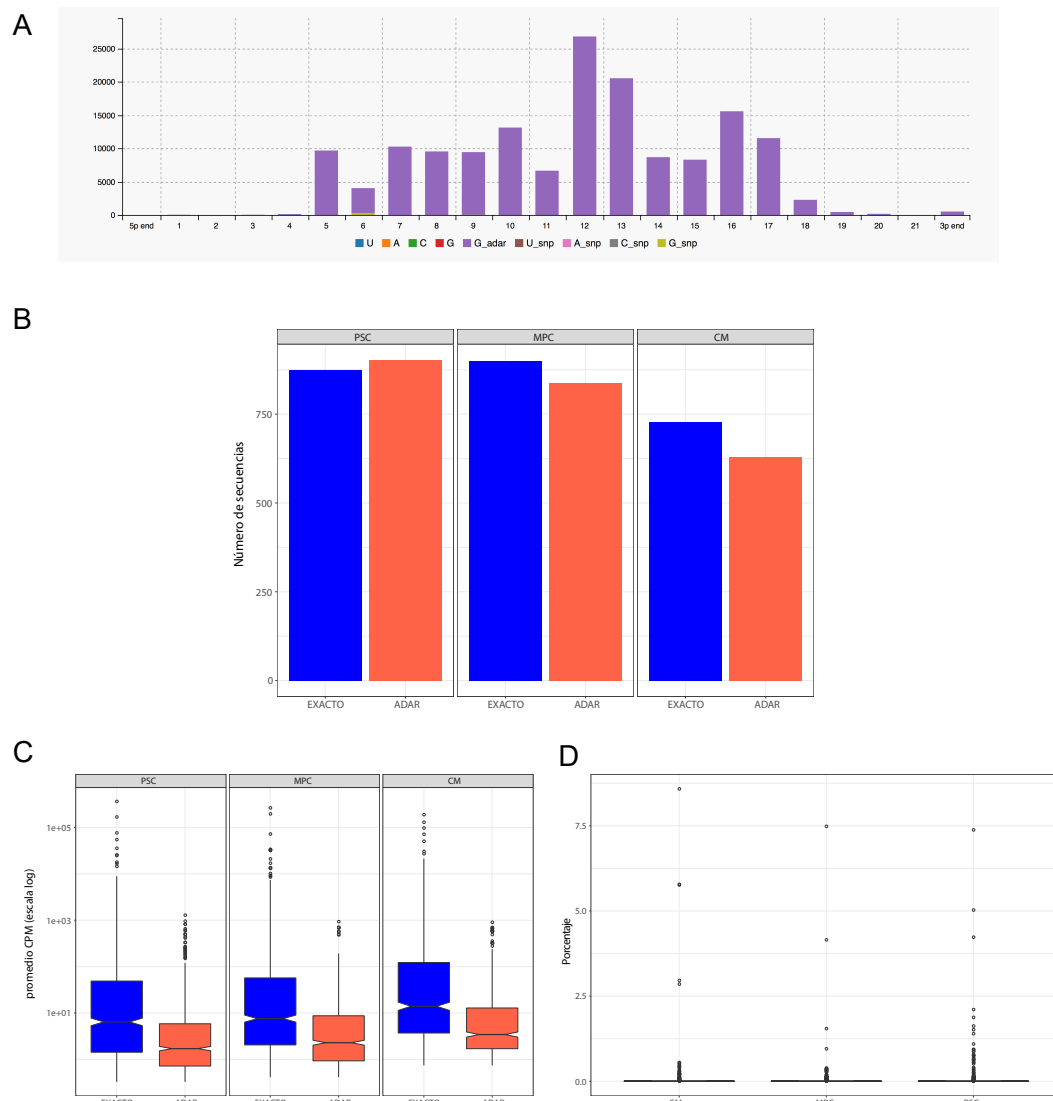


Figura 20: Análisis de los isomiRs con ediciones A-I: (A) Abundancia y posición de las distintas bases, SNPs y modificaciones A-I ("G_adar") a lo largo de la secuencia de isomiRs determinados con el programa Chimira. Cada base o modificación está representada con un color diferente. (B) Número de microARNs canónicos ("exactos") e isomiRs con ediciones A-I en cada población celular. (C) Abundancia de los microARNs canónicos e isomiRs en las tres poblaciones expresada como el promedio de las reads por millón de reads (CPM, *counts per millón*) en escala log₁₀. (D) Abundancia de isomiRs con modificaciones A-I en cada población celular expresada como porcentaje de reads con respecto a la forma canónica.

2.3. Análisis de las familias y clusters de microARNs en la diferenciación cardíaca

En un determinado contexto celular existe un número extenso de microARNs a lo que se conoce habitualmente como miRNoma. Los microARNs presentes actúan generalmente en conjunto con otros microARNs y no de manera individual, formando parte de familias o *clusters*, lo cual esta directamente relacionado con los genes y las vías de señalización que regulan. En este sentido, estudiamos el miRNoma de las células a lo largo de la diferenciación cardíaca, según la dinámica de expresión de las familias de microARNs (comparten la secuencia semilla) y de los *clusters* de microARNs (arbitrariamente definido como los microARNs localizados en el genoma a una distancia igual o menor a 10.000 pb). En la figura 21 se muestra en un *circos plot* la expresión de los *clusters* de microARNs en las tres poblaciones celulares, en función de su localización en el genoma (representada en el cariotipo del círculo más externo) y la relación de estos con las familias de microARNs. En este gráfico se puede observar por ejemplo, el perfil de expresión del *cluster* de la familia del miR-302 (denominado por su ubicación como "Chr4.1"), el cual presentó niveles altos de expresión en las PSC y MPC y expresión baja en los CM. Asimismo, se puede observar que los miembros de ésta familia están localizados todos en un mismo *cluster* representado por la línea roja. Por el contrario, las líneas azules unen a los microARNs de una misma familia pero que no comparten su localización en el genoma, es decir forman parte de distintos *clusters*, tal como el caso de las familias del miR-17, miR-19, miR-25 y miR-363, los cuales están localizados en tres *clusters* diferentes, Chr7.1, Chr13.2 y ChrX.8.

En la figura 22 se muestran los perfiles de expresión de las familias y *clusters* de microARNs en las tres poblaciones celulares. En los *heatmaps* se puede observar la dinámica de expresión en la diferenciación cardíaca para los 63 *clusters* de microARNs y las 141 familias de microARNs identificados en los tres grupos de células. Por otro lado, analizamos la similitud que existe entre las poblaciones celulares en base a la expresión de los *clusters* y las familias de microARNs, lo cual se puede observar en los *heatmaps* de distancia y en los gráficos de componentes principales (PCA). Estos resultados sugieren que los microARNs de las familias y de los *clusters* comparten sus perfiles de expresión a lo largo del proceso de diferenciación, dado que se observan patrones definidos de expresión en cada población celular. Por otro lado, tal como esperábamos, se observa una mayor similitud entre las PSC y MPC, tanto en el

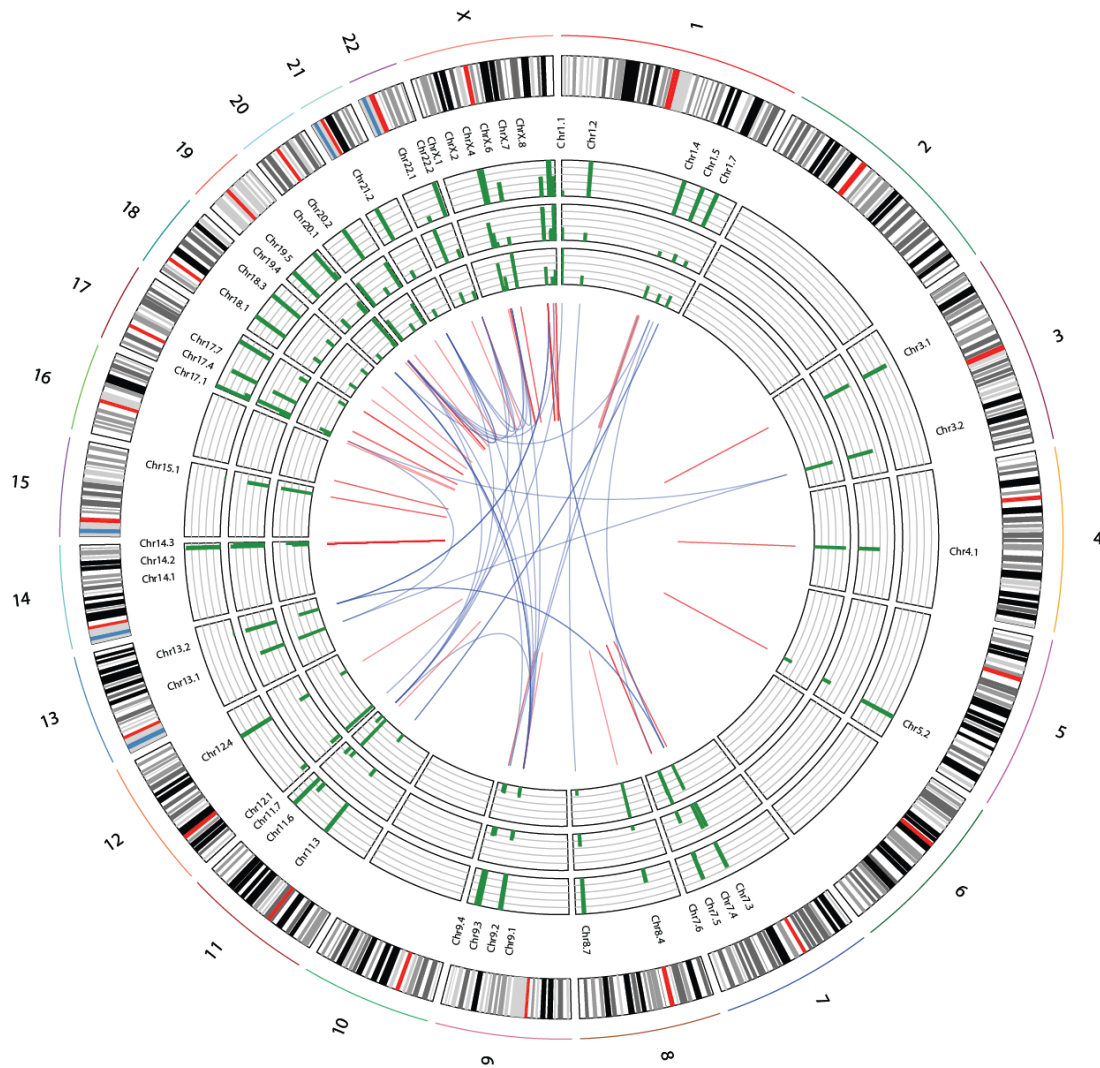


Figura 21: Perfiles de expresión de los *clusters* de microARNs y su relación con las familias de microARNs: En el *circos plot* se observa en el círculo más externo los 23 cromosomas y en los gráficos de barra color verde están representados los niveles de expresión de los *clusters* en las células PSC, MPC y CM (de adentro hacia afuera). En el centro del círculo, las líneas rojas representan las familias de los microARNs que comparten la localización en el genoma, y en azul aquellas cuyos miembros están localizados en distintas regiones del genoma.

análisis de *clusters* como en el de familias.

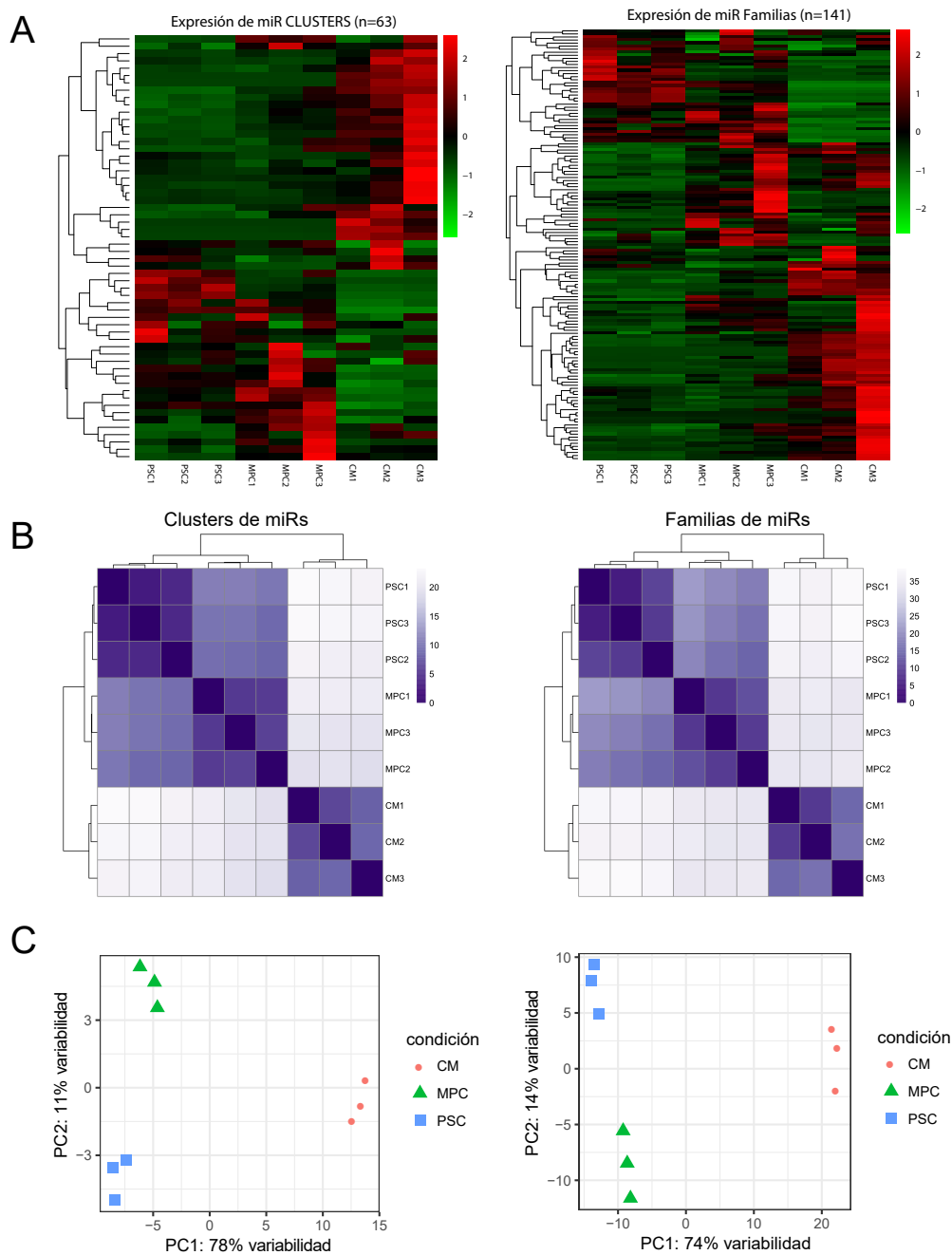


Figura 22: Análisis de expresión de las familias y *clusters* de microARNs: (A) En el *heatmap* de la izquierda están representados los 63 *clusters* de microARNs y sus niveles de expresión en las tres poblaciones celulares PSC, MPC y CM y sus correspondientes replicas y a la derecha están representadas las 141 familias de microARNs y sus niveles de expresión. En rojo y verde están representados los niveles de expresión altos y bajos, respectivamente. (B) Los *heatmaps* representan la variabilidad o diferencias que existen entre las tres replicas de las tres poblaciones, PSC, MPC y CM en base a los niveles de expresión de los *clusters* (izquierda) y familias (derecha) de microARNs. La escala de color representa menor a mayor variabilidad en tonos más y menos intensos respectivamente. (C) Los gráficos de PCA permiten agrupar las poblaciones y sus respectivas réplicas, según las similitudes de los *clusters* (izquierda) y familias (derecha) de microARNs.

A continuación, decidimos estudiar mediante el análisis de los términos GO los procesos biológicos que regularían las familias/clusters de microARNs, en base a la predicción *in silico* de genes blanco. Para esto, elegimos tres familias/clusters representativas de cada población celular. La familia del miR-302 que es característica de las células pluripotentes, la familia/cluster del miR-17/92 que presentó niveles altos de expresión tanto en las PSC como en las MPC y los microARNs cardíacos que incluyen la familia del miR-1 (miR-1 y miR-206), el miR-133 (miR-133a y miR-133b), el miR-208 (miR-208a y miR-208b), el miR-499, el miR-490 y el miR-143.

En la figura 24A se detallan los términos GO asociados con los genes blanco predichos para la familia del miR-302 (en la tabla suplementaria 9 se detalla la lista completa de términos GO, ver tabla). Tal como se puede observar, esta familia estaría regulando genes relacionados con el desarrollo embrionario tales como el desarrollo del corazón, desarrollo de músculo y el desarrollo del encéfalo anterior, entre otros. Al analizar el número de genes blanco en función del número de microARNs que los regulan, obtuvimos que la mayoría de los *targets* predichos serían regulados por entre uno a cinco microARNs de la familia. Esto coincide con el concepto de que los microARNs actúan de manera colaborativa y no individualmente en la regulación de los ARNm *target*. Luego, analizamos entre los microARNs de la familia la correlación existente entre los *scores* obtenidos para cada interacción microARN/ARNm *target*. Este tipo de análisis nos permitió explorar de manera indirecta si la afinidad que tienen los microARNs por los ARNm blanco en común, es similar o no. Es decir, solamente se representa la correlación entre los genes *targets* en común. Cabe aclarar que cada interacción predicha con las herramientas bioinformáticas definen un *score* de interacción en base a la probabilidad *in silico* de interacción. Al analizar la correlación entre los *scores*, obtuvimos que la mayor correlación se da entre los miR-302a, miR-302b, miR-302c y miR-302d. El resto de los microARNs de la familia compartirían ciertos genes blanco pero la afinidad por ellos es diferente. Este resultado habla de especificidades relativas que pueden tener los microARNs por un mismo ARNm *target*.

Realizamos el mismo análisis para la familia/cluster miR-17/92 y determinamos que estos microARNs regularían distintos procesos biológicos relacionados también con el desarrollo hacia el linaje cardíaco, el desarrollo neuronal y el desarrollo del músculo (figura 25A) (la lista completa de términos GO se detalla en la tabla suplementaria 10, ver tabla). Al analizar el número de *targets* en función del número de microARNs, determinamos que la mayoría de los genes serían regulados por uno, dos u ocho de los microARNs de la familia/cluster

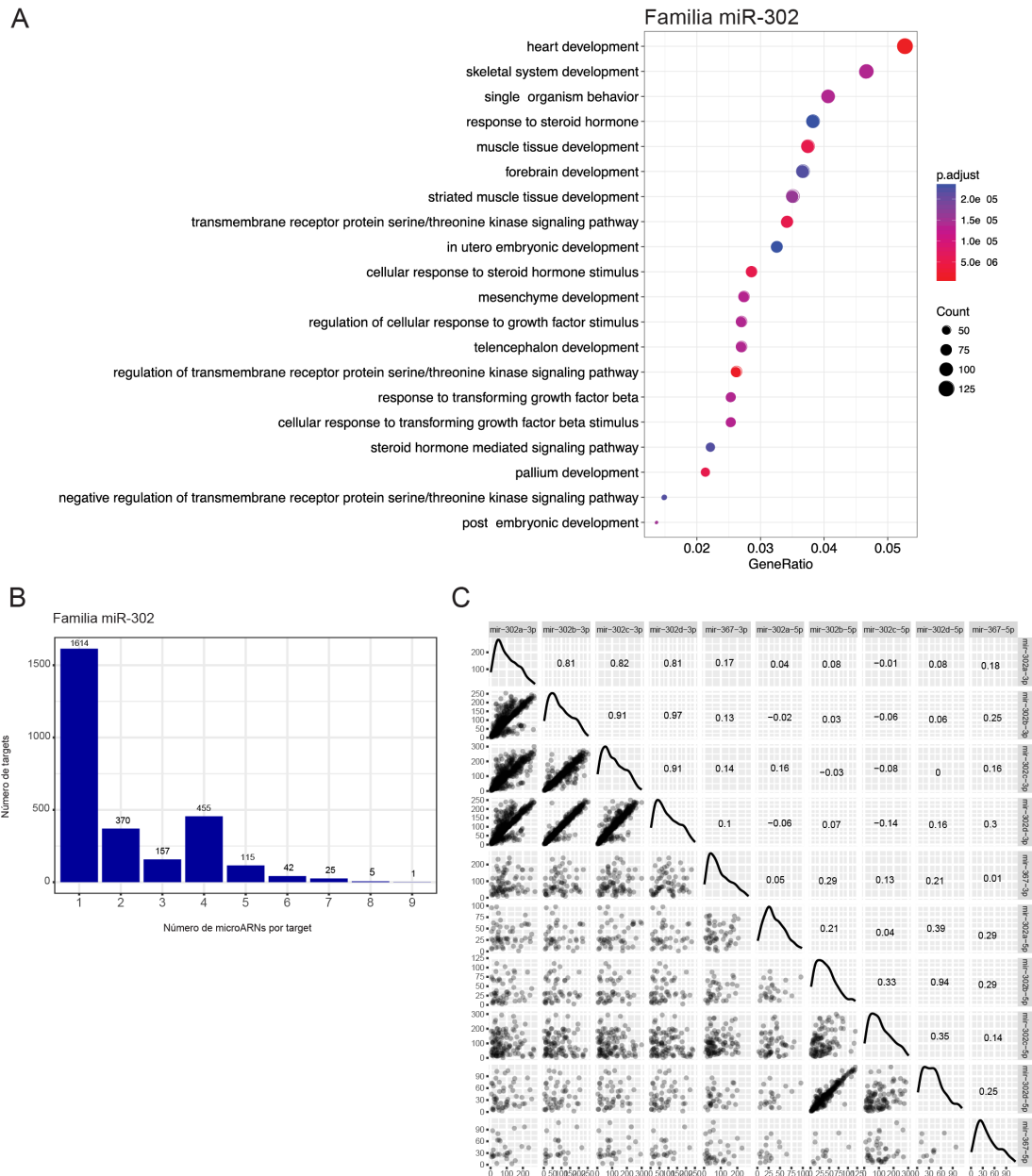


Figura 24: Análisis de los genes *target* de la familia miR-302: (A) Términos GO asociados con los genes blanco de la familia del miR-302, en función del *Gene Ratio*, p-valor (escala de color) y el número de genes por término GO (tamaño del *dot*). (B) Número de genes blanco en función del número de microARNs que los regulan. (C) Matriz de correlación de los *scores* obtenidos para la interacción microARN/ARNm *target* entre los microARNs de la familia miR-302, analizados de a pares. Cada punto en el *scatter plot* representa un gen blanco en común en los dos microARNs. Las curvas de la diagonal representan el ajuste al modelo de correlación y por encima de la diagonal se muestran los coeficientes de correlación.

(figura 25B). Al igual que con el miR-302 esto da idea de la colaboración que existe entre los microARNs en la regulación de sus *targets*. En la matriz de correlación, se puede observar que existe una gran correlación entre varios de los microARNs analizados, tales como el miR-106b-5p con el miR-93-5p, el miR-25-3p y el miR-92b-3p (figura 25C).

Por último, analizamos los procesos biológicos relacionados con los genes blanco predichos para los microARNs cardíacos seleccionados. En este caso, tal como se puede observar en la figura 26A, a partir del análisis de los términos GO identificamos distintos procesos relacionados con el desarrollo de distintos tipos celulares que también derivan del mesodermo, tales como la diferenciación al linaje mieloide, desarrollo del sistema renal y desarrollo del músculo (ver lista completa de los términos GO en la tabla suplementaria 11, ver tabla). Es decir, los microARNs cardíacos analizados reprimen genes relacionados con la diferenciación a linajes mesodérmicos no cardíacos. Al analizar el número de genes blanco en función del número de microARNs, identificamos que la mayoría de los *targets* predichos serían regulados por uno a cuatro de los microARNs cardíacos. Sin embargo, en la matriz de correlaciones, se puede observar que la afinidad por los genes blanco en común es diferente para cada microARN (excepto para los miR-208a y miR-208b). Nuevamente, estos resultados refuerzan la idea de que sí bien los microARNs forman redes complejas de regulación en donde un gen *target* es regulado por varios microARNs, en muchos casos la interacción microARN-ARNm sería específica y diferente al resto. Cabe aclarar que en este caso a diferencia de los otros dos grupos de microARNs estudiados, el número de genes blanco en común es menor, dado que los microARNs tienen un perfil de expresión similar pero forman parte de distintas familias/*clusters*.

Finalmente, con el fin de comparar la regulación que efectuarían los microARNs de una familia/cluster de manera individual en comparación con la familia/cluster entera, decidimos estudiar a modo de ejemplo los términos GO asociados con los genes blanco del miR-302a-3p, miR-302a-5p y el miR-302d-3p (figura 27). Tal como se puede observar en los tres gráficos de *dot plots*, los términos GO son muy diferentes a los identificados para todos los miembros de la familia del miR-302. Este resultado nos pareció interesante, ya que sugiere que determinados microARNs tienen un efecto en la regulación de importantes vías de señalización cuando actúan en conjunto y no de manera individual. Es importante destacar que estas conclusiones son únicamente conjeturas que necesitan ser validadas, ya que el ensayo se basa en datos predictivos y no experimentales.

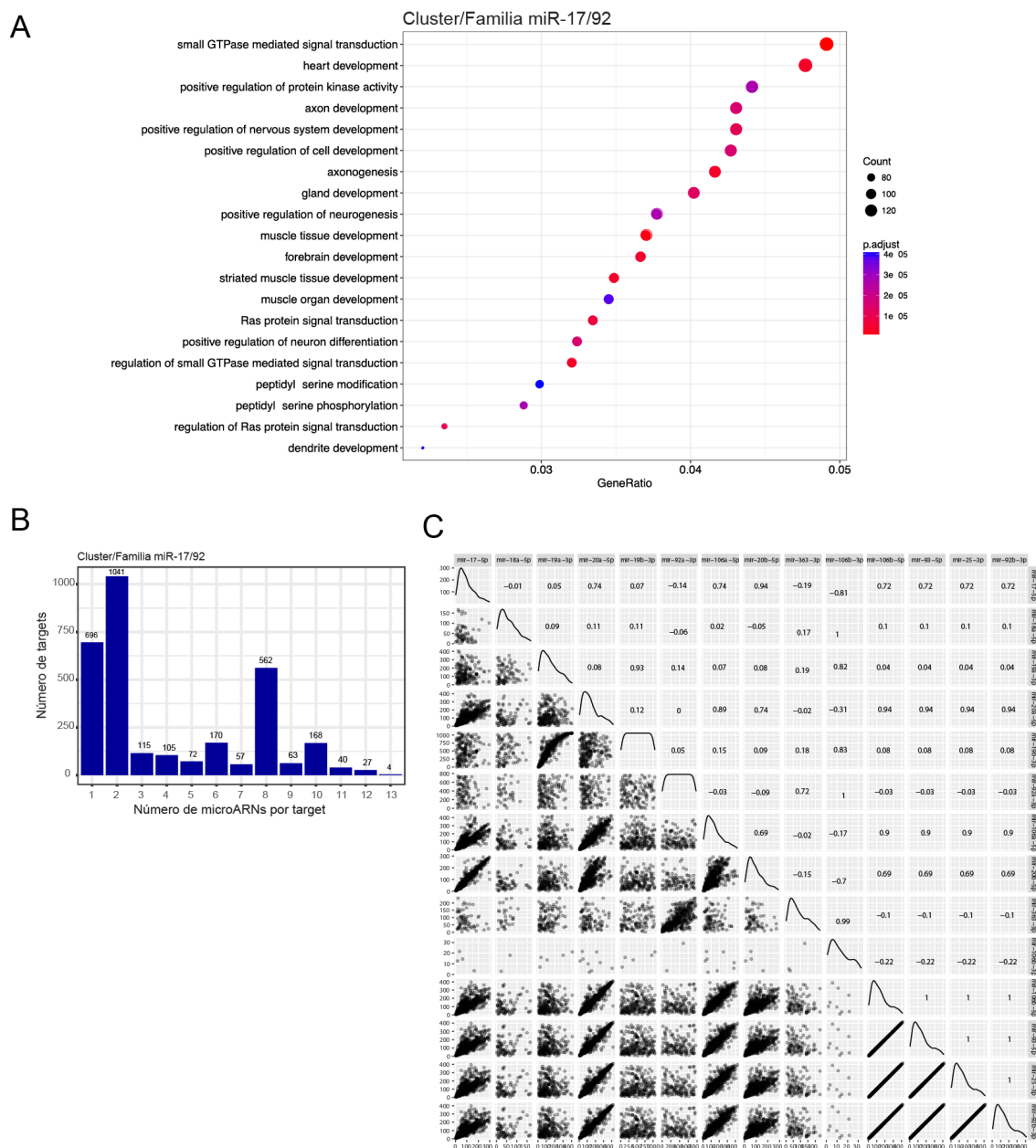


Figura 25: Análisis de los genes *target* de la familia/cluster miR-17/92: (A) Términos GO asociados con los genes blanco de la familia/cluster del miR-17/92, en función del *Gene Ratio*, p-valor (escala de color) y el número de genes por término GO (tamaño del dot). (B) Número de genes blanco en función del número de microARNs que los regulan. (C) Matriz de correlación de los *scores* obtenidos para la interacción microARN/ARNm *target* entre los microARNs de la familia/cluster del miR-17/92, analizados de a pares.

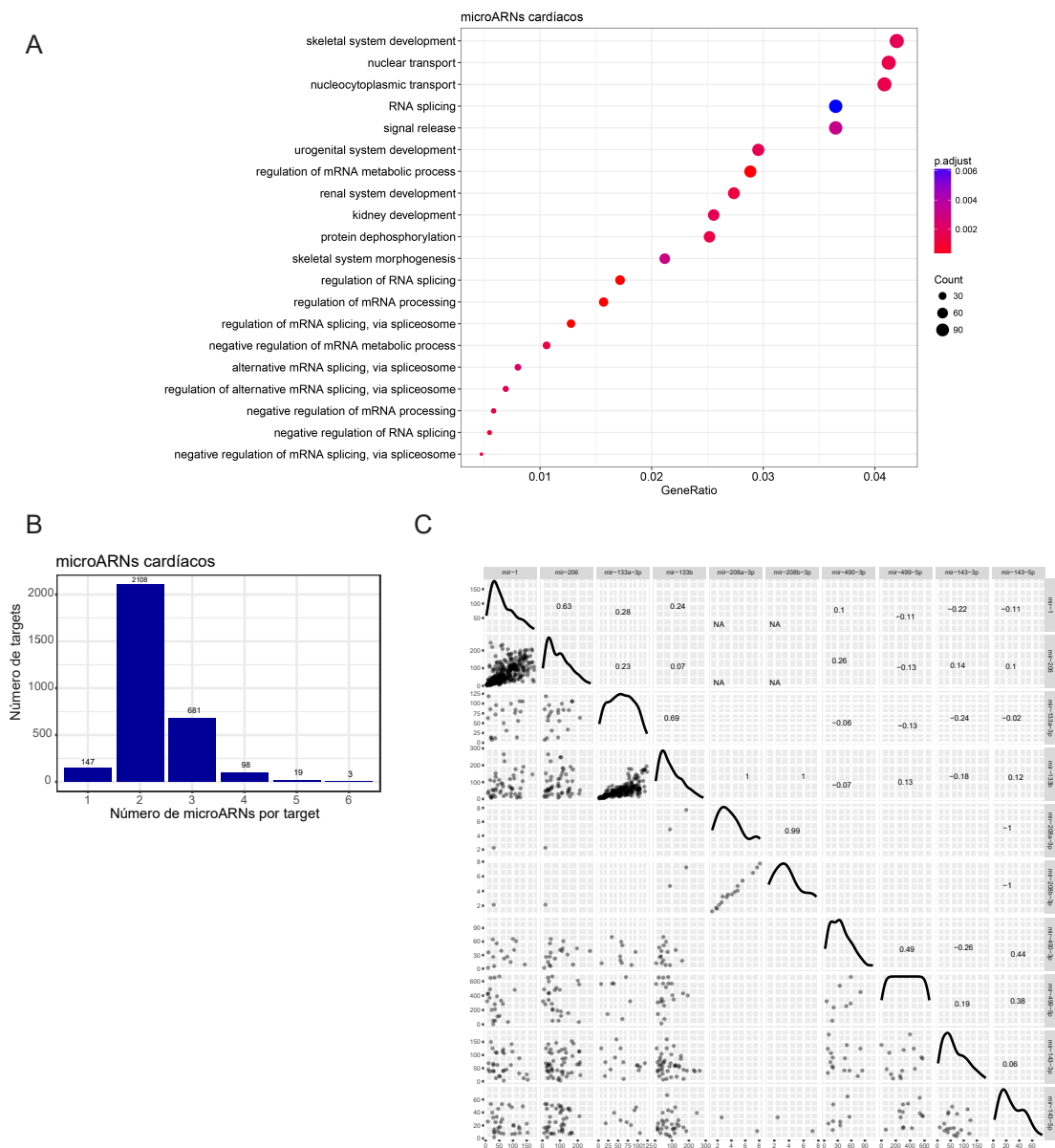


Figura 26: Análisis de los genes *target* de los microARNs cardíacos: (A) Términos GO asociados con los genes blanco de los microARNs cardíacos miR-1, miR-206, miR-133a, miR-133b, miR-208a, miR-208b, miR-499, miR-490 y miR-143, en función del *Gene Ratio*, el p-valor (escala de color) y el número de genes por término GO (tamaño del *dot*). (B) Número de genes blanco en función del número de microARNs que los regulan. (C) Matriz de correlación de los *scores* obtenidos para la interacción microARN/ARNm *target* entre los microARNs cardíacos, analizados de a pares.

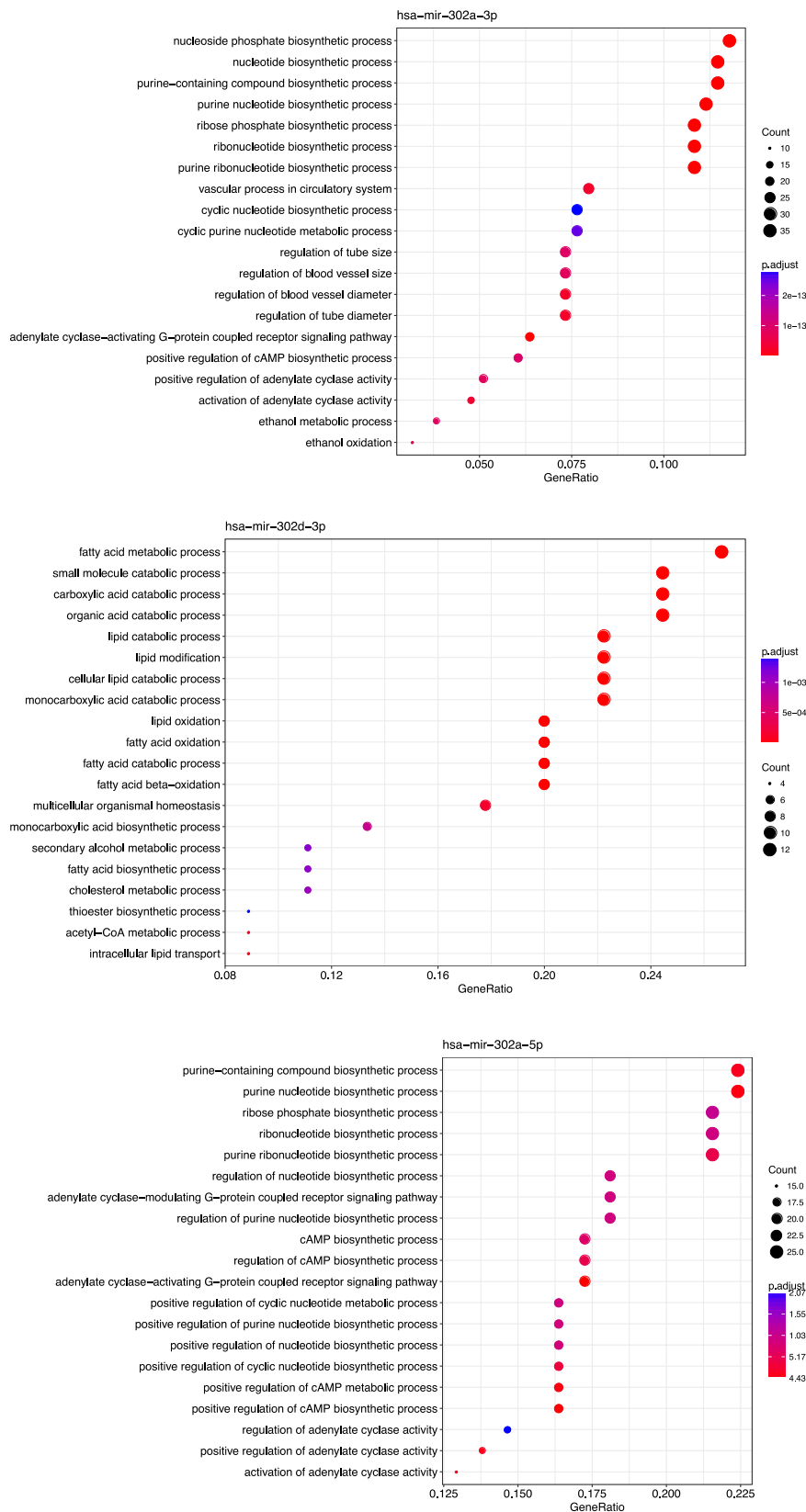


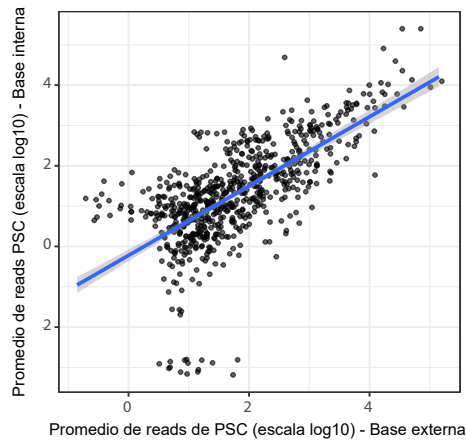
Figura 27: Análisis de los genes blanco de los microARNs individuales: Cada gráfico de *dot plot* representa los términos GO identificados para el miR-302a-3p (arriba), el miR-302d-3p (medio) y el miR-302a-5p (abajo), en función del *gene ratio*, el valor-p (escala de colores) y el número de genes asociado a cada término GO (tamaño del *dot*).

2.4. MicroARNs asociados con el linaje cardíaco

Con el fin de identificar los microARNs que se expresan específicamente en las células cardíacas, decidimos realizar un análisis de expresión diferencial entre los niveles de expresión de los microARNs en cardiomiocitos (CM) y la expresión de los microARN en otros tipos de células completamente diferenciadas. Para ello, descargamos los datos de *RNA-Seq* publicados en GEO con el *accession number*: GSE68189, los cuales habían sido obtenidos también a partir de células embrionarias humanas (H9). En la figura 28A se muestra en un *scatter plot* la correlación, en cuanto a la expresión de los microARNs, en las células pluripotentes de ambas bases de datos (la propia de nuestro trabajo y la externa). Tal como se puede observar, la correlación de los niveles de expresión de microARNs entre ambas bases de datos fue alta, otorgándole de esta manera robustez a nuestros resultados. En la figura 28B se muestran los niveles de expresión de los microARNs en los cardiomiocitos (CM) y los cinco tipos celulares, los cuales consisten en células mesenquimales (MPC_1), fibroblastos embrionarios de pulmón (HLF) y tres tipos de células obtenidas de una diferenciación neuronal, *fore*-, *mid*- y *hind-brain* (F_NPC, M_NPC y H_NPC). Como se puede observar en el *heatmap* existe un patrón bien definido de los microARNs que se *up-regulan* y *down-regulan* en cada población celular.

Luego, realizamos el análisis de expresión diferencial entre las poblaciones de a pares para identificar cuáles son los microARNs que aumentan/disminuyen sus niveles de expresión de manera significativa entre los CM y el resto de las poblaciones celulares estudiadas. En la figura 29 se pueden observar en cada *volcano plot* la expresión diferencial entre los CM y los distintos tipos celulares (en las tablas suplementarias 12-16 se detallan las listas completas de los microARNs de cada *volcano plot*, ver tablas). En la tabla 3 se detallan los microARNs más abundantes en los CM ("todos"), los microARNs que se expresan diferencialmente más en los CM en relación con el resto de los tipos celulares ("DE") y los microARNs que se expresan únicamente en los CM y no en las otras células analizadas ("únicos"). En el último grupo identificamos los microARNs extensamente estudiados en el desarrollo cardíaco como miR-1, miR-133a y b, miR-208 y miR-499, así como también otros microARNs relacionados con el fenotipo cardíaco pero hasta ahora poco estudiados en la diferenciación cardíaca, como el miR-675, el miR-585, el miR-1248 y el miR-509-5p.

A



B

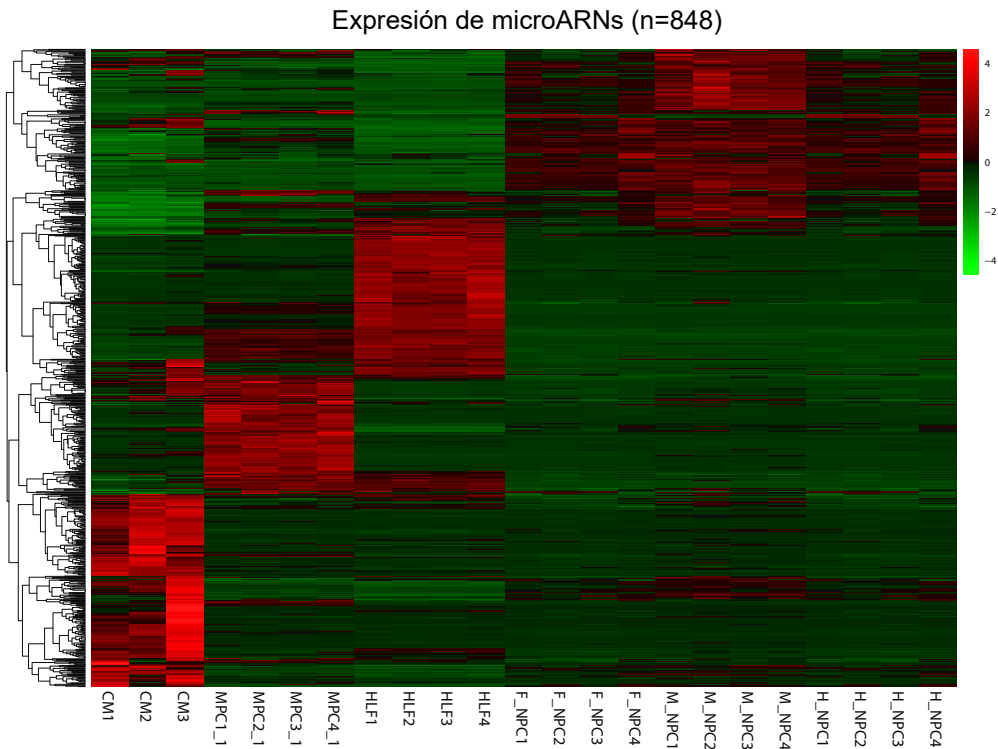


Figura 28: Expresión de los microARNs en los CM en relación con otros tipos de células: (A) Análisis de la correlación entre la expresión de los microARNs en las células PSC (base de datos interna) y las células pluripotentes de la base de datos externa. En el gráfico están representados los niveles de expresión en promedio entre las replicas y en escala log10. (B) En el *heatmap* se muestran los niveles de expresión de los microARNs en los cardiomiocitos (CM), células mesenquimales (MPC_1), fibroblastos embrionarios de pulmón (HLF), neuronas de cerebro anterior (F_NPC), neuronas de mesencéfalo (M_NPC) y neuronas del romboencéfalo (H_NPC). Los números corresponden al número de replica de cada tipo celular.

microARN (todos)	<i>reads</i>	microARN - DE	<i>reads</i>	microARN - único	<i>reads</i>
hsa-miR-21-5p	199059	hsa-miR-143-3p	135432	hsa-miR-1-3p	48933
hsa-miR-143-3p	135432	hsa-miR-100-5p	74723	hsa-miR-499a-5p	9598
hsa-miR-100-5p	74723	hsa-miR-1-3p	48933	hsa-miR-208a-3p	4940
hsa-miR-1-3p	48933	hsa-miR-30d-5p	29289	hsa-miR-675-5p	686
hsa-miR-30d-5p	29289	hsa-miR-99a-5p	17254	hsa-miR-208b-3p	621
hsa-miR-148a-3p	28649	hsa-miR-22-3p	14910	hsa-miR-133b	348
hsa-miR-26a-5p	24070	hsa-miR-125a-5p	14101	hsa-miR-133a-5p	199
hsa-miR-26a-5p	24021	hsa-miR-27b-3p	10936	hsa-miR-585-3p	175
hsa-miR-99b-5p	21955	hsa-miR-125b-5p	10647	hsa-miR-1248	104
hsa-miR-99a-5p	17254	hsa-miR-499a-5p	9680	hsa-miR-509-5p	88

Tabla 3: miRNoma de los cardiomiocitos: En esta tabla se detallan los diez microARNs más destacados según su expresión en las células cardíacas. En las primeras dos columnas se muestran los microARNs más abundantes en los CM y sus respectivos niveles de expresión (en *reads*). La tercera y cuarta columna (microARNs - DE) corresponden a los microARNs significativamente más expresados en los CM, con respecto al resto de los tipos celulares. Por último, la quinta y sexta columnas corresponden a los microARNs que se expresan exclusivamente en los CM. Los valores de *reads* corresponden al promedio de las tres replicas de CM.

Conclusiones parciales del capítulo:

En base a los resultados de la secuenciación masiva determinamos que al rededor de 700 microARNs presentan un perfil de expresión dinámico a lo largo de la diferenciación cardíaca y específico para cada una de las tres poblaciones celulares estudiadas, las células pluripotentes humanas (PSC), el progenitor temprano de mesodermo (MPC) y los cardiomiocitos (CM). De esta manera logramos determinar el miRNoma característico de cada población celular, clave en la regulación de los procesos celulares que permiten el establecimiento de la identidad celular en la diferenciación cardíaca.

Al comparar la expresión de los microARNs pudimos distinguir que existen cinco perfiles de expresión bien definidos, en los cuales se agrupan los microARNs que comparten la dinámica de expresión. A su vez, a partir del análisis de expresión diferencial entre las poblaciones de a pares identificamos aquellos microARNs que presentaron cambios significativos en sus niveles de expresión en el proceso de diferenciación celular. Al considerar de manera global la

expresión de los microARNs determinamos que existe una mayor similitud entre las PSC y MPC que al comparar las PSC con los CM. Creemos que estas diferencias podrían explicarse debido a que los CM son células más especializadas y maduras que el progenitor de mesodermo, lo cual se ve reflejado en la expresión de los microARNs.

A su vez, analizamos las variantes de los microARNs denominadas isomiRs y determinamos que las isoformas más abundantes son aquellas con modificaciones en el extremo 3'. Al estudiar la abundancia de los microARNs canónicos y los isomiRs, identificamos que los isomiRs-3' presentan niveles de expresión equivalentes a los microARNs canónicos y mayores que los isomiR-5'. A lo largo de la diferenciación cardíaca no encontramos diferencias en los niveles de expresión a nivel general de los microARNs y sus isoformas, lo cual sugiere que no hay un sesgo en los mecanismos de regulación de dichas moléculas entre los tres grupos de células. Por otro lado, exploramos la posición de las variantes de los isomiR-5' e isomiR-3' y determinamos que la modificación más frecuente es el "podado" (en inglés *trimming*) de tres bases en el extremo 5' y una base en el extremo 3'. Asimismo, la frecuencia de modificaciones dentro de la secuencia (SNPs) fue más alta en el extremo 5' que en el 3'. Si bien las modificaciones en los extremos 5' fueron menos abundantes que las del extremo 3', el impacto que pueden tener en la funcionalidad de los microARNs puede ser muy alto, ya que podría modificarse la región responsable del reconocimiento de los genes *target*. En este sentido al analizar en profundidad el miR-302a y el impacto que podrían tener sus isomiR-5' en el reconocimiento de los genes *target* predichos, determinamos que las isoformas con modificaciones en la región semilla (*seed sequence*) reconocerían genes distintos a los reconocidos por el microARN canónico, lo cual se ve reflejado en los procesos celulares asociados a dichos genes (determinados en el análisis de *Gene Ontology*).

Por otro lado, estudiamos la dinámica de expresión de los microARNs en base a los grupos funcionales de familias y *clusters*, y determinamos que dichos grupos de microARNs comparten una dinámica de expresión bien definida en cada una de las tres poblaciones de células. El análisis de similitud entre las poblaciones en base a la expresión de las familias y *clusters* de microARNs dio resultados similares al análisis global de los microARNs sin agrupar, lo cual sugiere que cada población celular presenta una identidad de familias y *clusters* que permite distinguirlas del resto de las poblaciones estudiadas. A su vez, analizamos tres grupos de familias/*clusters* característicos de cada una de las tres poblaciones e identificamos que regulan genes en común y que su acción en conjunto es la que permitiría establecer la identidad

celular. Estos resultados sugieren que el peso de un microARN único en su rol regulatorio en una célula puede ser muy acotado y que puede ser mucho más importante el peso de una familia o un *cluster* de microARNs.

Por último, describimos el miRNoma de las células cardíacas comparando la expresión de los microARNs con distintos tipos celulares provenientes de la misma línea celular, tales como células mesenquimales, fibroblastos embrionarios de pulmón y células neuronales de distintas regiones del encéfalo. Identificamos mediante este análisis cuáles son los microARNs más abundantes en los cardiomiocitos en comparación con el resto de las células, y aquellos microARNs con una expresión no ubicua, sino que su expresión fue exclusiva en los cardiomiocitos y no en las otras poblaciones.

Los resultados de esta parte del trabajo fueron publicados recientemente en Scientific Reports (Garate et al., 2018).

Capítulo 3:

Estudio del *cluster* C19MC y su rol en las células pluripotentes humanas

A partir del análisis de los resultados de secuenciación masiva de los ARN pequeños identificamos un *cluster* de microARNs, conocido como C19MC, que está exclusivamente *up-regulado* en las células pluripotentes y disminuye su expresión hasta niveles indetectables en el progenitor de mesodermo y en los cardiomiocitos. Dado que se conoce poco acerca de la función de los microARNs de este *cluster* en las células pluripotentes humanas y en el proceso de diferenciación, decidimos estudiar y analizar el rol que éstos tienen en nuestro modelo de estudio. El *cluster* C19MC está localizado en el cromosoma 19 y su nombre proviene de las siglas en inglés *Chromosome 19 MicroRNAs Cluster*. Este *cluster* de microARNs se encuentra únicamente en primates y es uno de los *clusters* más grande presente en el genoma, formado por 46 genes que codifican para 56 microARNs maduros, agrupados en una región de aproximadamente 100 kb. En esta región del genoma no hay genes codificantes para proteínas, pero sí hay un gran número de secuencias *Alu* ubicadas entre los genes de los microARNs. Los microARNs del C19MC pertenecen a la familia del miR-506 (únicamente el miR-512) y del miR-515 (todo el resto de los microARNs del *cluster*) y poseen una gran similitud en su secuencia. Dado la naturaleza móvil de las secuencias *Alu* se cree que su presencia dentro del *cluster* podría estar relacionado con la formación de estos genes en *tandem* y la similitud de sus secuencias. Río abajo del *cluster* C19MC se encuentra otro *cluster* formado por la familia del miR-373 (miR-371, miR-372 y miR-373), muy expresado también en las células madre pluripotentes. En la figura 30 se muestra un esquema del *locus* del C19MC en el genoma.

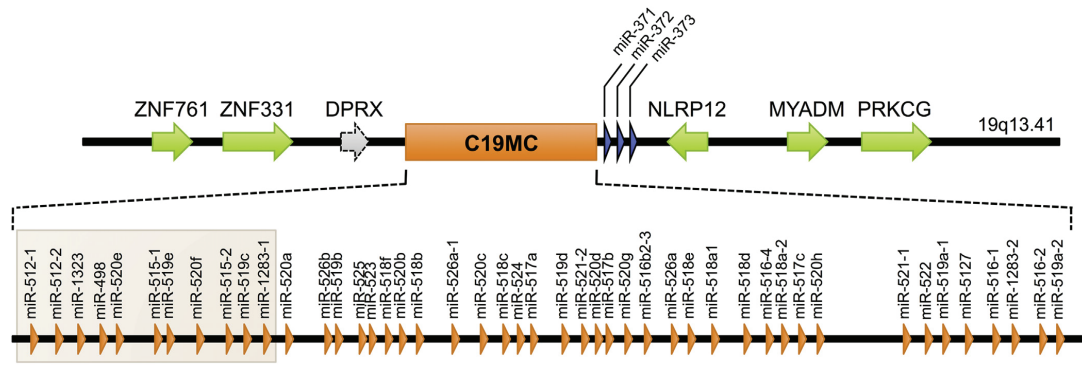


Figura 30: Locus del *cluster* C19MC: En este esquema se pueden observar los genes de los microARNs que forman el *cluster* C19MC, el *cluster* de los microARNs miR-371-73 y los genes próximos al *cluster*. La figura fue adaptada de (Mouillet et al., 2015).

Inicialmente, los microARNs del C19MC han sido identificados en placenta (Dance and Weber, 2009; Noguer-dance et al., 2010) y se ha propuesto que regularían la capacidad migratoria de las células del trofoblasto en la invasión de la decidua (Xie et al., 2015). Por otro lado, se ha determinado que algunos microARNs del *cluster* C19MC se expresan en células cancerígenas (Spence et al., 2014) y en células madre embrionarias humanas (Stadler et al., 2010), pero se desconoce el rol que tendrían en éstas últimas.

3.1. Análisis de la expresión del *cluster* C19MC en la diferenciación celular

Al analizar en detalle los niveles de expresión de los microARNs del *cluster* C19MC obtenidos en la secuenciación, identificamos que 41 microARNs de los 56 presentan niveles de expresión por arriba del punto de corte (un promedio del número de *reads* entre las tres replicas mayor o igual a 10) y sus niveles de expresión bajan drásticamente en el progenitor de mesodermo y en los cardiomiocitos (figura 31).

A continuación, validamos los niveles de expresión de algunos de los microARNs del *cluster* mediante PCR en tiempo real durante el proceso de diferenciación celular. En la figura 32 se muestran los niveles de expresión para los microARNs miR-512, miR-517a,b, miR-517c, miR-519d, miR-520a, miR-520f, miR-520g y miR-526. Como se puede observar, existe una disminución rápida de los niveles de expresión en el proceso de diferenciación.

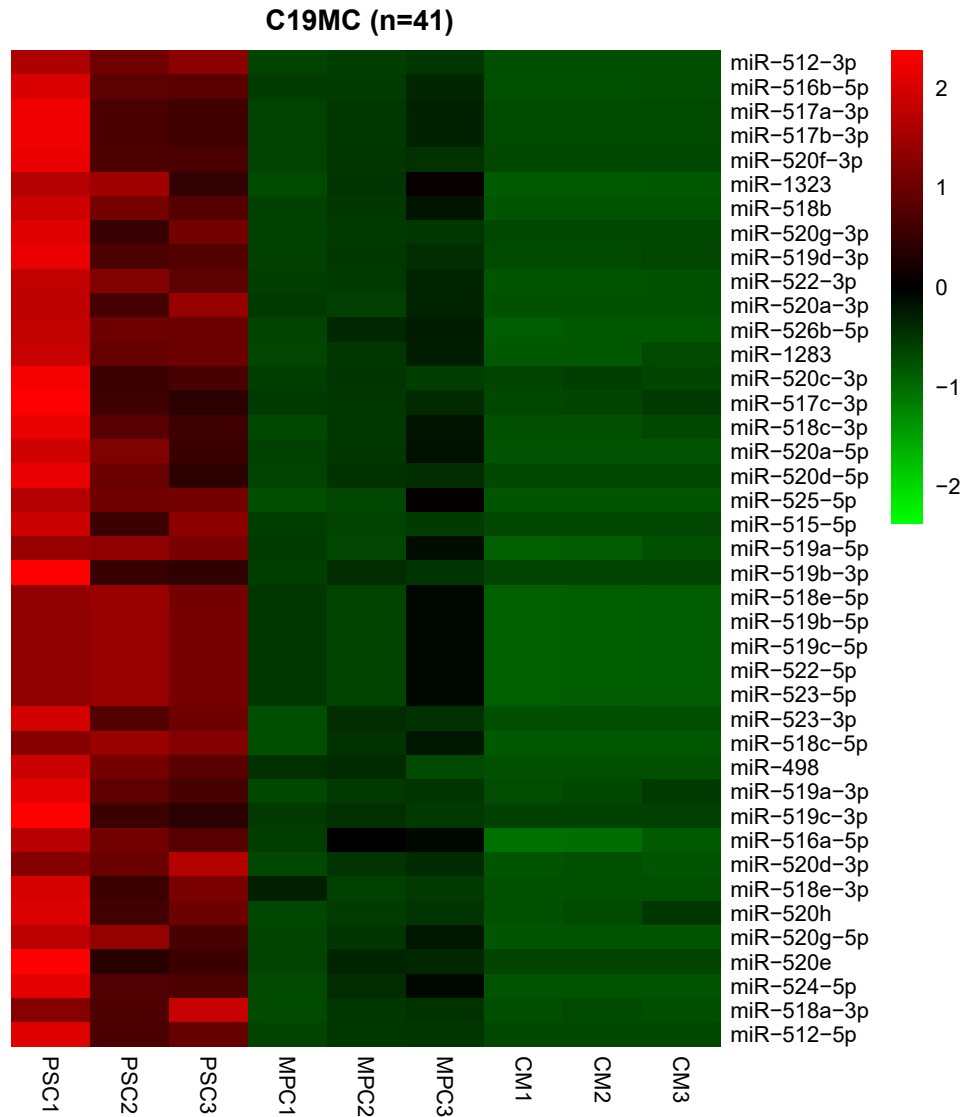


Figura 31: Niveles de expresión de los microARNs del *cluster* C19MC: En el *heatmap* se pueden observar los niveles de expresión del C19MC en las tres poblaciones celulares. En rojo se representan los niveles de expresión alto y en verde los bajos (los microARNs están ordenados según su nivel de expresión de mayor a menor). Cada columna del *heatmap* corresponde a una réplica biológica de las células PSC, MPC y CM.

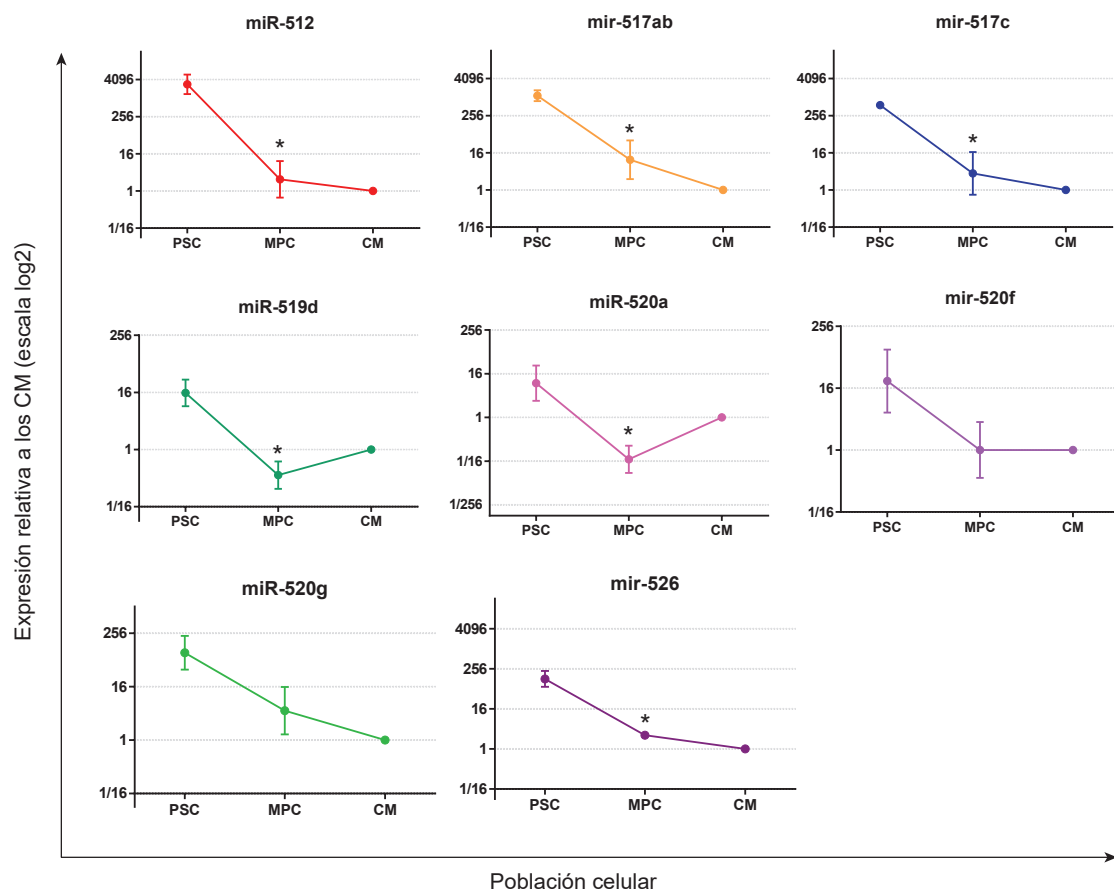


Figura 32: Validación de la expresión de los microARNs del C19MC: Se analizó los niveles de expresión de determinados microARNs del *cluster* C19MC mediante PCR en tiempo real en las células PSC, MPC y CM. En el gráfico se muestra la media $\pm$ el SEM de tres réplicas biológicas para las poblaciones PSC y MPC. * p-valor <0.05 entre las PSC y MPC.

3.2. Análisis *in silico* del *cluster* C19MC y sus genes blanco

En una primera etapa estudiamos el *cluster* C19MC mediante herramientas informáticas. Por un lado, analizamos a partir de datos cargados en Ensembl los factores de transcripción y marcadores de transcripción activa que se unen a las regiones próximas del C19MC, determinados mediante experimentos de ChIP-Seq (*Chromatin Immunoprecipitation sequencing*) (figura 33). A partir de éste análisis identificamos que en las CMEh (H1) existen sitios de unión de factores tales como Oct-4 (POU5F1) y Nanog río arriba y hacia el final del *cluster*, respectivamente. A su vez, existen sitios de unión de CTCF y p300. CTCF es un factor de transcripción

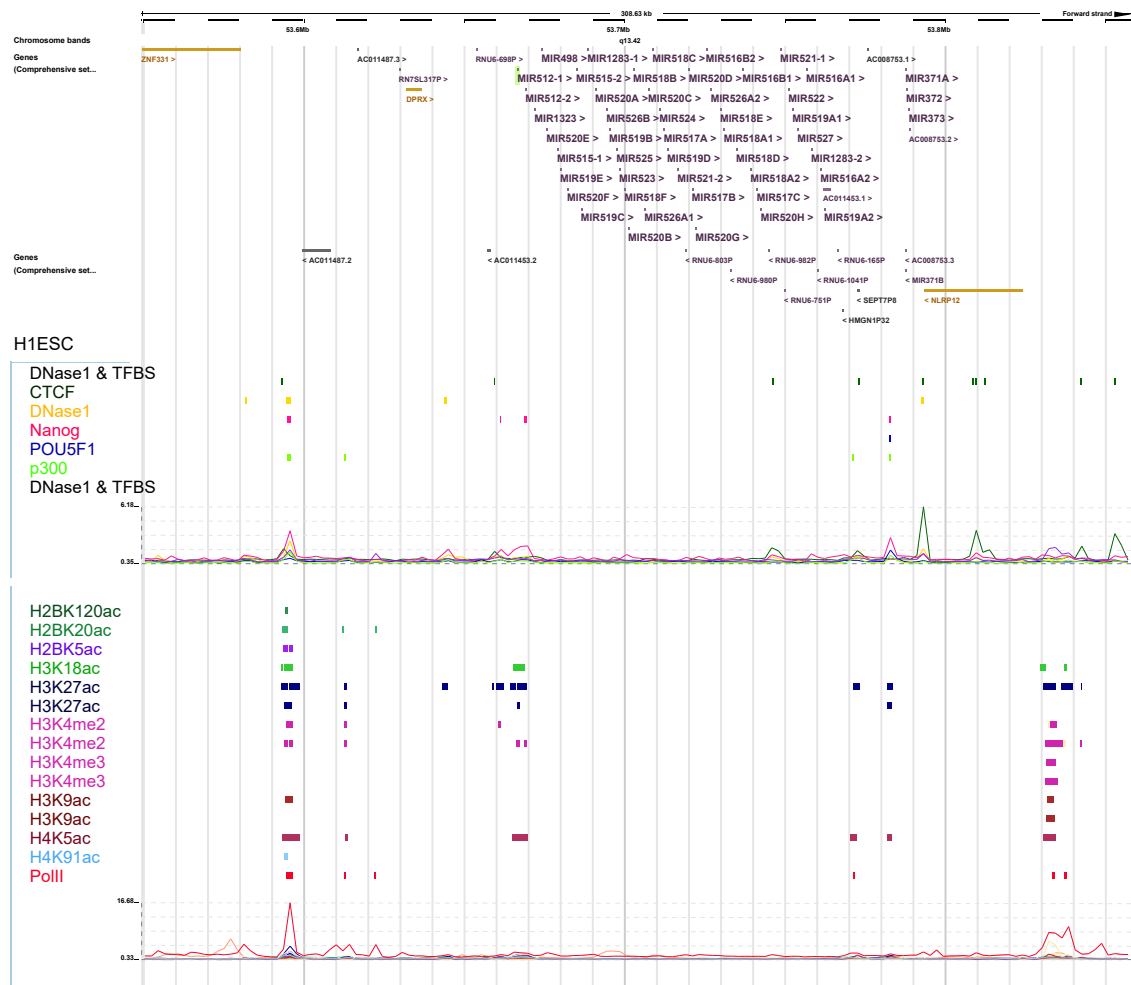


Figura 33: Análisis de marcas de transcripción activa/inactiva en las regiones próximas al C19MC: En la figura se muestran los sitios de unión al ADN de los distintos factores y de las histonas con sus respectivas modificaciones, a lo largo del *cluster* C19MC. Las barras de colores representan la ubicación de los sitios de unión al ADN de cada factor, según su respectivo color. En la parte inferior se puede observar la frecuencia de unión de éstos factores en la altura de cada pico. Los datos que se muestran en la figura están publicados en la base de datos *Ensembl* y aquellos que están repetidos se debe a que provienen de resultados experimentales diferentes.

que se une al ADN y colabora en la formación de estructuras 3D y las berreras que separan la cromatina activa de la inactiva y p300 es un coactivador de la transcripción que entre otras cosas se une a la ARN polimerasa II. Además, en la región río arriba próxima al inicio del *cluster* existen sitios de unión de histonas con modificaciones de transcripción activa. Estos datos sugieren que en las células pluripotentes la región donde se encuentra el *cluster* C19MC está transcripcionalmente activa, lo cual es consistente con nuestros datos de secuenciación y le brinda una mayor robustez a los mismos.

Debido a que la mayoría de los microARNs del *cluster* pertenecen a la misma familia, decidimos analizar la similitud de sus secuencias mediante alineamiento múltiple (figura 34). Tal como se puede observar en el alineamiento, varios de los microARNs del C19MC comparten parte de sus secuencias, presentando una gran similitud en la región 5' como por ejemplo los miR-520-3p. Dado que la complementariedad con sus genes blanco se da por la secuencia semilla (*seed sequence*) ubicada en el extremo 5', consideramos que estos microARNs podrían estar regulando genes en común. Además, la acumulación de su expresión puede brindarles una importancia extra al afectar todos a genes *target* similares.

A continuación, exploramos los posibles genes blanco de los microARNs del C19MC mediante un análisis *in silico* y analizamos la superposición entre los genes blanco predichos para los microARNs del *cluster*. Para esto, seleccionamos los microARNs que presentaron los mayores niveles de expresión para poder trabajar así con un número más acotado de genes. Por lo tanto, seleccionamos 9 microARNs que presentaron en las PSC un nivel de expresión por arriba de los 100 *reads* en promedio entre las tres replicas. Para la predicción de los genes blanco combinamos las listas de genes obtenidas mediante cinco algoritmos diferentes (*DIANA tools*, *Miranda*, *PicTar*, *TargetScan* y *mirDB*) y seleccionamos los genes predichos por los cinco *softwares* para aumentar la certeza de la predicción. En la figura 35 se observan círculos que corresponden a los microARNs y sus genes *target*, y las líneas grises que representan la interacción de cada microARN con su gen *target*. Tal como se puede observar existe una gran superposición entre los genes *target* predichos para los 9 microARNs, lo cual se ilustra en la alta densidad de líneas grises que cruzan y unen los círculos. A pesar de que este análisis está basado en una predicción, sugiere que los microARNs del *cluster* C19MC actúan de manera conjunta en la regulación de sus genes blanco y presentarían un cierto nivel de redundancia entre ellos.

Como se explicó anteriormente, cada interacción microARN-gen blanco determinada por los algoritmos de predicción presenta un *score* asociado, el cual se relaciona con la afinidad que tiene el microARN en la interacción con dicho gen. Teniendo en cuenta esto, estudiamos el nivel de correlación de dichos *scores* comparando los microARNs de a pares (figura 36). El análisis de los coeficientes de correlación determinó que la mayoría son bajos o incluso negativos. Una posible interpretación de estos resultados es que a pesar de que los microARNs del *cluster* comparten varios de sus genes *target*, la afinidad por los mismos es diferente para cada microARN. La única excepción determinada fue para el par miR-517a y el miR-517b que al

```

hsa-miR-1283      1 U-CUACAAAGG---A-AAGCGCU-UUC-----U
hsa-miR-1323      1 U-C--AAAACU---G-AGGGGA-UU---UUCU-
hsa-miR-498       1 --UUUCA-AGCCAGG-GGGCGUU-UUU-----C
hsa-miR-512-3p    1 -----AAGU---G-CUGUCAUAGCUGAGGU-C
hsa-miR-512-5p    1 --CACU-CAGCCUUGAGGGCACU-UU-----C
hsa-miR-515-5p    1 UUCUCCAAAG---A-AAGCACU-UUC---U--G
hsa-miR-516a-5p   1 --UUCUCGAGGAAAG-AAGCACU-UU-----C
hsa-miR-516b-5p   1 --AUCUGGAGGUAAG-AAGCACU-U-----U
hsa-miR-517a-3p   1 -----AUCGU---G-CAUCCCU-UUAGAGUG-U
hsa-miR-517b-3p   1 -----AUCGU---G-CAUCCCU-UUAGAGUG-U
hsa-miR-517c-3p   1 -----AUCGU---G-CAUCCCU-UUAGAGUG-U
hsa-miR-518a-3p   1 --G---AAAGC---G-CUUCUU-UUGCUGG--A
hsa-miR-518b      1 --C---AAAGC---G-CUCCCU-UUAGAG-G-U
hsa-miR-518c-3p   1 --C---AAAGC---G-CUUCUU-UUAGAGUG-U
hsa-miR-518c-5p   1 U-CUCUGGAGG---G-AAGCACU-UUC---U--G
hsa-miR-518e-3p   1 -----AAAGC---G-CUUCUU-UCAGAGU--G
hsa-miR-518e-5p   1 --CUCUAGAGG---G-AAGCGCU-UUC---U--G
hsa-miR-519a-3p   1 -----AAAGU---G-CAUCCUU-UUAGAGUG-U
hsa-miR-519a-5p   1 --CUCUAGAGG---G-AAGCGCU-UUC---U--G
hsa-miR-519b-3p   1 -----AAAGU---G-CAUCCUU-UUAGAGGU-U
hsa-miR-519b-5p   1 --CUCUAGAGG---G-AAGCGCU-UUC---U--G
hsa-miR-519c-3p   1 -----AAAGU---G-CAUCCUU-UUAGAGGA-U
hsa-miR-519c-5p   1 --CUCUAGAGG---G-AAGCGCU-UUC---U--G
hsa-miR-519d-3p   1 --C---AAAGU---G-CCUCCU-UUAGAGU--G
hsa-miR-520a-3p   1 -----AAAGU---G-CUUCUU-UUGGACUG-U
hsa-miR-520a-5p   1 --CUCCAGAGG---G-AAGUACU-UUC-----U
hsa-miR-520c-3p   1 -----AAAGU---G-CUUCUU-UUAGAGGG-U
hsa-miR-520d-3p   1 -----AAAGU---G-CUUCUU-UUGGUGGG-U
hsa-miR-520d-5p   1 --CUACAAAGG---G-AAGCGCU-UU-----C
hsa-miR-520e      1 -----AAAGU---G-CUUCUU-UUUGAGG--G
hsa-miR-520f-3p   1 -----AAGU---G-CUUCUU-UUAGAGGGUU
hsa-miR-520g-3p   1 A-C---AAAGU---G-CUUCUU-UUAGAGUG-U
hsa-miR-520g-5p   1 ---UCUAGAG---G-AAGCACU-UUC-UGUU-U
hsa-miR-520h      1 A-C---AAAGU---G-CUUCUU-UUAGA--G-U
hsa-miR-522-3p    1 -----AAAU---G-CUUCUU-UUAGAGUG-U
hsa-miR-522-5p    1 --CUCUAGAGG---G-AAGCGCU-UUC---U--G
hsa-miR-523-3p    1 --G---AAGC---G-CUUCUU-AUAGAGGG-U
hsa-miR-523-5p    1 --CUCUAGAGG---G-AAGCGCU-UUC---U--G
hsa-miR-524-5p    1 --CUACAAAGG---G-AAGCACU-UUC---U--C
hsa-miR-525-5p    1 --CUCCAGAGG---G-AUGCACU-UUC-----U
hsa-miR-526b-5p   1 --CUCUUGAGG---G-AAGCACU-UUC---UG-U

```

Figura 34: Alineamiento múltiple de las secuencias de los microARNs del *cluster* C19MC: En la figura se puede observar el alineamiento múltiple de las secuencias correspondientes a los 41 microARNs presentes en las células pluripotentes humanas. En color negro y gris se representan las posiciones que están más conservadas entre las secuencias alineadas.

tener secuencias muy similares, tanto los genes blanco predichos como el *score* de interacción, son los mismos. Un hecho a destacar es que para muchos pares de microARNs los genes *target* compartidos son escasos (representado por la presencia de muy pocos puntos). Esto sugiere que a pesar de ser muy parecidos a nivel de secuencia, tienen especificidad distinta. Este hecho se observa claramente en la figura 35.

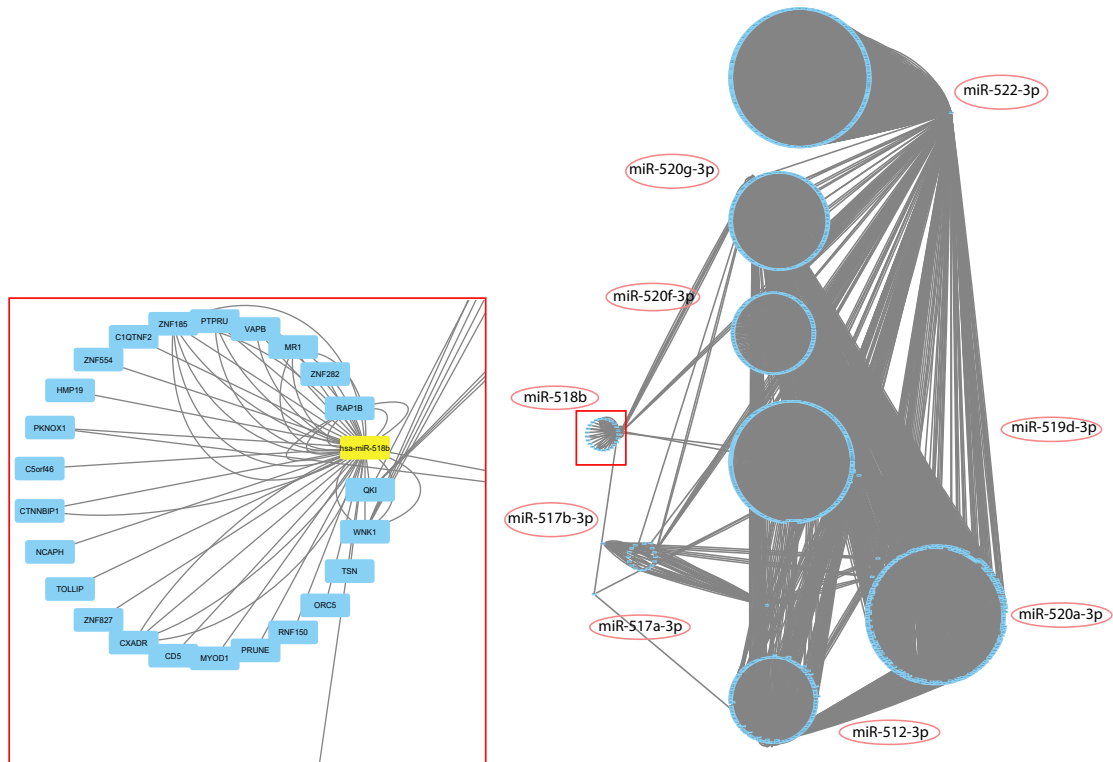


Figura 35: El *cluster* C19MC y la red de sus genes blanco predichos: Cada círculo de la figura representa un microARN y todos sus genes blanco y las líneas grises corresponden a dichas interacciones. En el recuadro de la izquierda se muestra un aumento del miR-518b y sus genes blanco. Las líneas que salen de los círculos representan los genes compartidos entre más de un microARN.

Luego analizamos los procesos biológicos asociados a los genes blanco predichos para comprender cuáles serían los posibles procesos regulados por el *cluster* C19MC. En la figura 37 se pueden observar los términos de *Gene Ontology* determinados a partir de los genes blanco predichos. Tal como se puede observar en el gráfico, los genes regulados por el *cluster* C19MC estarían relacionados con diversos procesos del desarrollo embrionario tales como el desarrollo neuronal, el desarrollo de músculo y desarrollo del sistema urogenital, así como también vías de señalización relacionadas con el estado de pluripotencia, tal como la vía de Wnt. Si bien este análisis es exploratorio, estos resultados sugieren que el C19MC podría regular genes relacionados con el mantenimiento del estado de indiferenciación de las células pluripotentes, lo cual es consistente con el nivel de expresión determinado en estas células.

Dado que el *cluster* C19MC se expresa específicamente en las células pluripotentes, decidimos alinear y comparar las secuencias de los microARNs de este *cluster* con los microARNs del *cluster* miR-371-373 y *cluster* del miR-302, ambos grupos de microARNs muy expresados

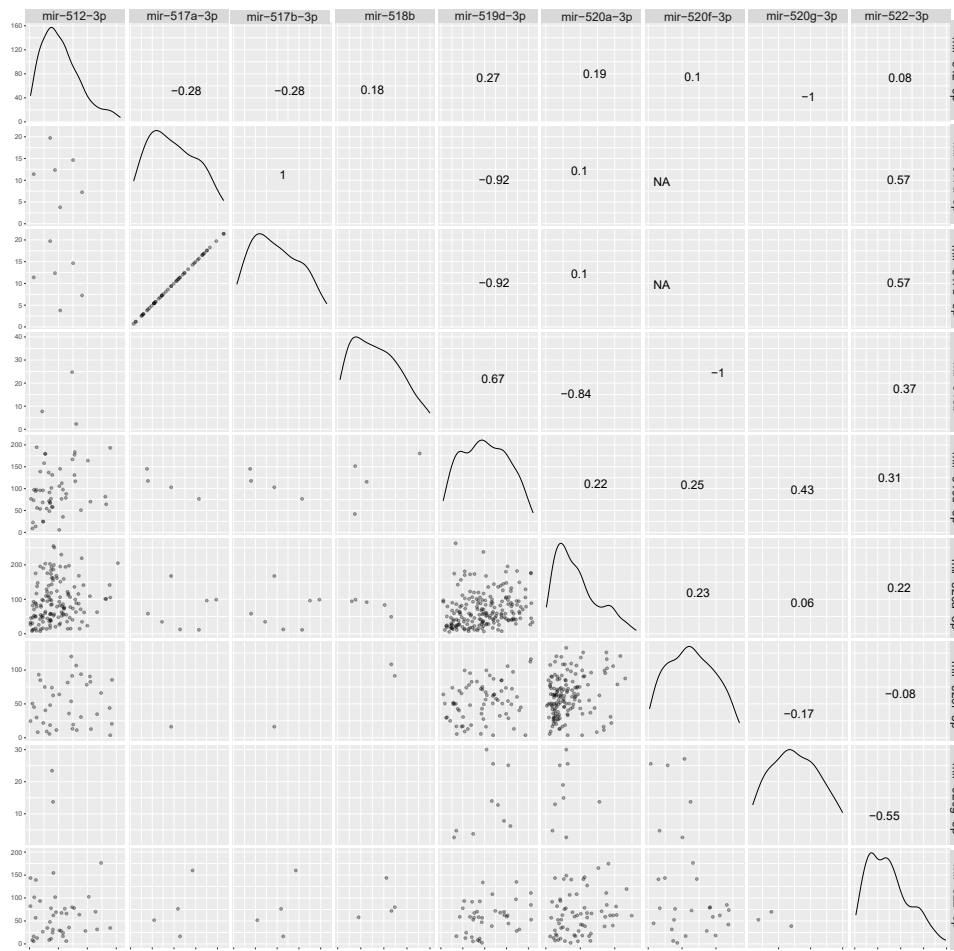


Figura 36: Correlación de la afinidad de los microARN del C19MC por sus genes blanco: En los gráficos de dispersión se muestran los distintos scores para cada gen blanco compartido por el par de microARNs representados en los ejes. Los gráficos de líneas en la diagonal representan el ajuste al modelo de correlación y por encima de la diagonal se muestran los coeficientes de correlación de cada comparación.

también en las células pluripotentes (figura 38A). Como se puede observar en el alineamiento múltiple de sus secuencias, existen varias posiciones idénticas a lo largo de las secuencias de los microARNs de los tres grupos, lo cual sugiere que estos microARNs podrían regular genes en común (en el análisis sólo consideramos la variante -3p de los microARN más expresados del C19MC).

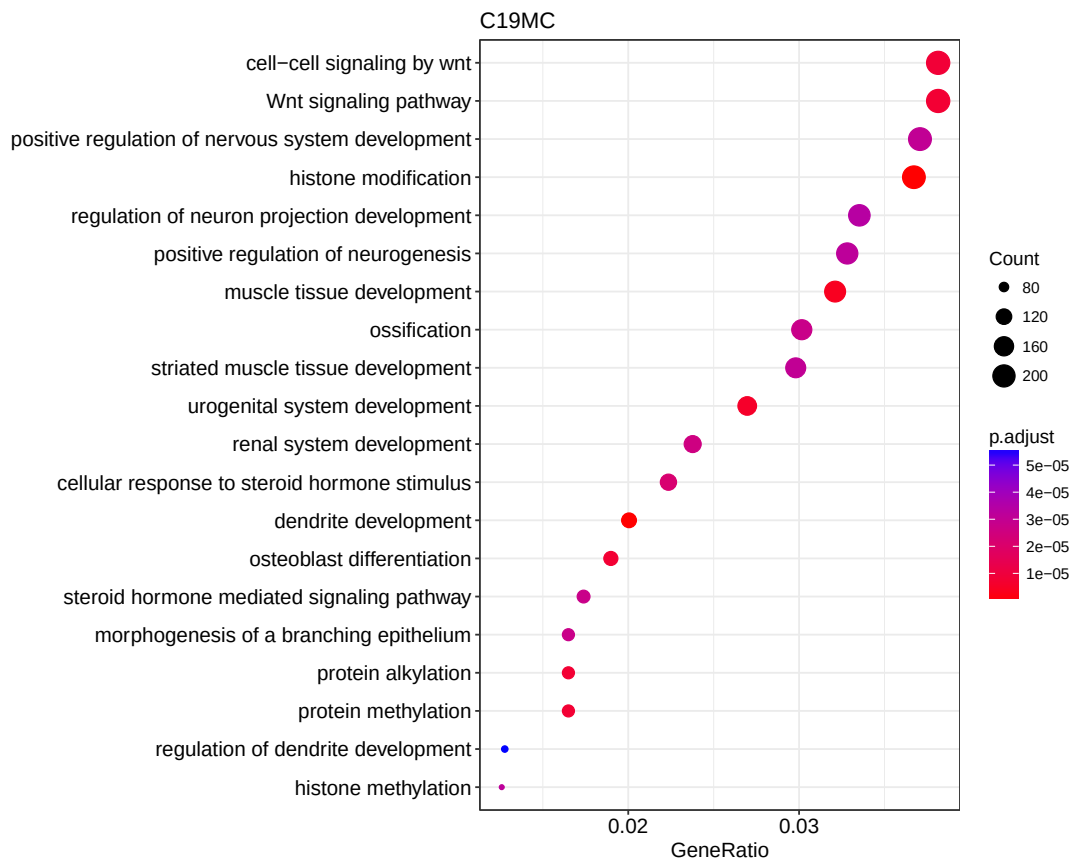


Figura 37: Análisis de *Gene Ontology* de los genes blanco predichos para los microARNs del C19MC: En el gráfico de *dot plot* se muestran los términos de GO asociados a los genes blanco del C19MC. El tamaño de cada *dot* representa el número de genes asociados a cada término, el *Gene ratio* representa la relación entre los genes observados y esperados y el color corresponde al p-valor del *test de Bonferroni*.

Por lo tanto, analizamos a continuación los genes blanco predichos para cada *cluster* mediante herramientas bioinformáticas y determinamos que los microARNs de los 3 *clusters* tienen en común 487 genes blanco predichos (figura 38B). Al analizar los términos de *Gene Ontology* de los 487 genes identificamos diversos procesos biológicos relacionados con el desarrollo embrionario, el ciclo celular, la apoptosis y la migración celular (figura 39). Este análisis reveló que varios microARNs del *cluster* C19MC podrían regular importantes procesos para el mantenimiento del estado de indiferenciación de las células pluripotentes, junto con los microARNs del *cluster* miR-302 y miR-373. Estos resultados son interesantes porque a pesar de que estrictamente se los clasifique dentro de familias distintas, su similitud en comportamiento de expresión y la similitud en sus secuencias nos hace pensar de una relación mayor entre las familias que la que ha sido descripta.

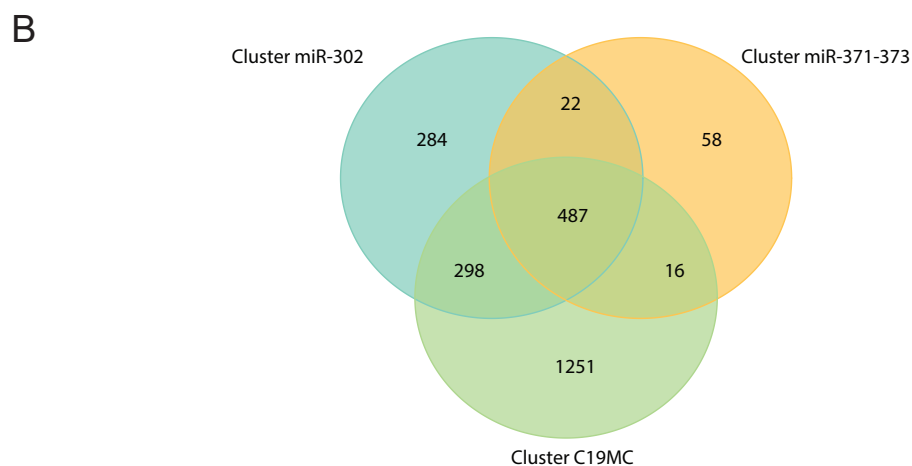
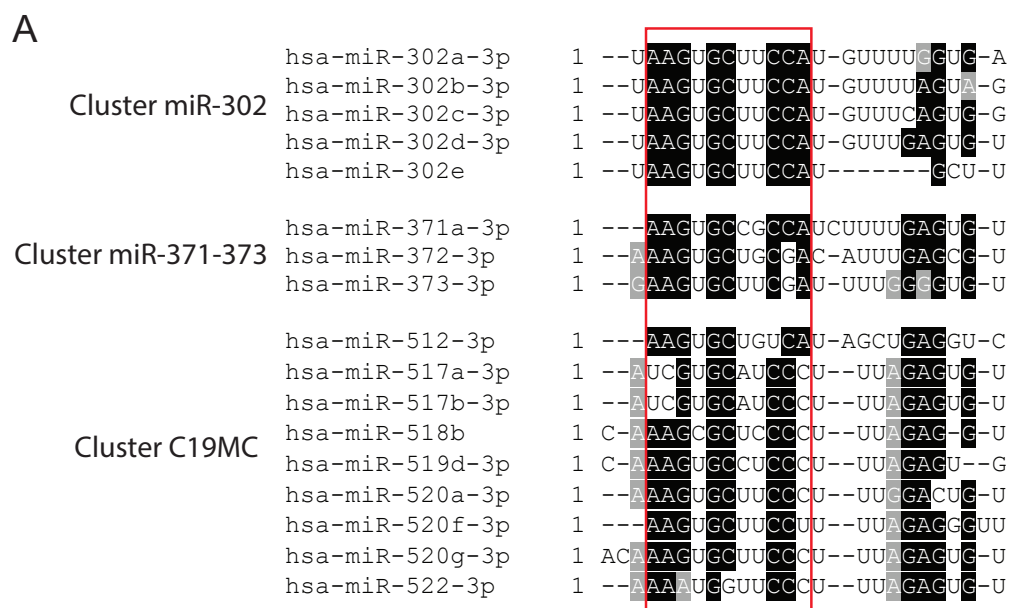


Figura 38: Comparación entre los cluster C19MC, miR-302 y miR-371-373: (A) Alineamiento múltiple las secuencias de los microARNs de los tres clusters (únicamente las variantes -3p). En el recuadro rojo se destaca la posibel región semilla de los microARNs. (B) En el diagrama de Venn se muestra el solapamiento que existe entre los genes blanco predichos para cada uno de los *clusters*.

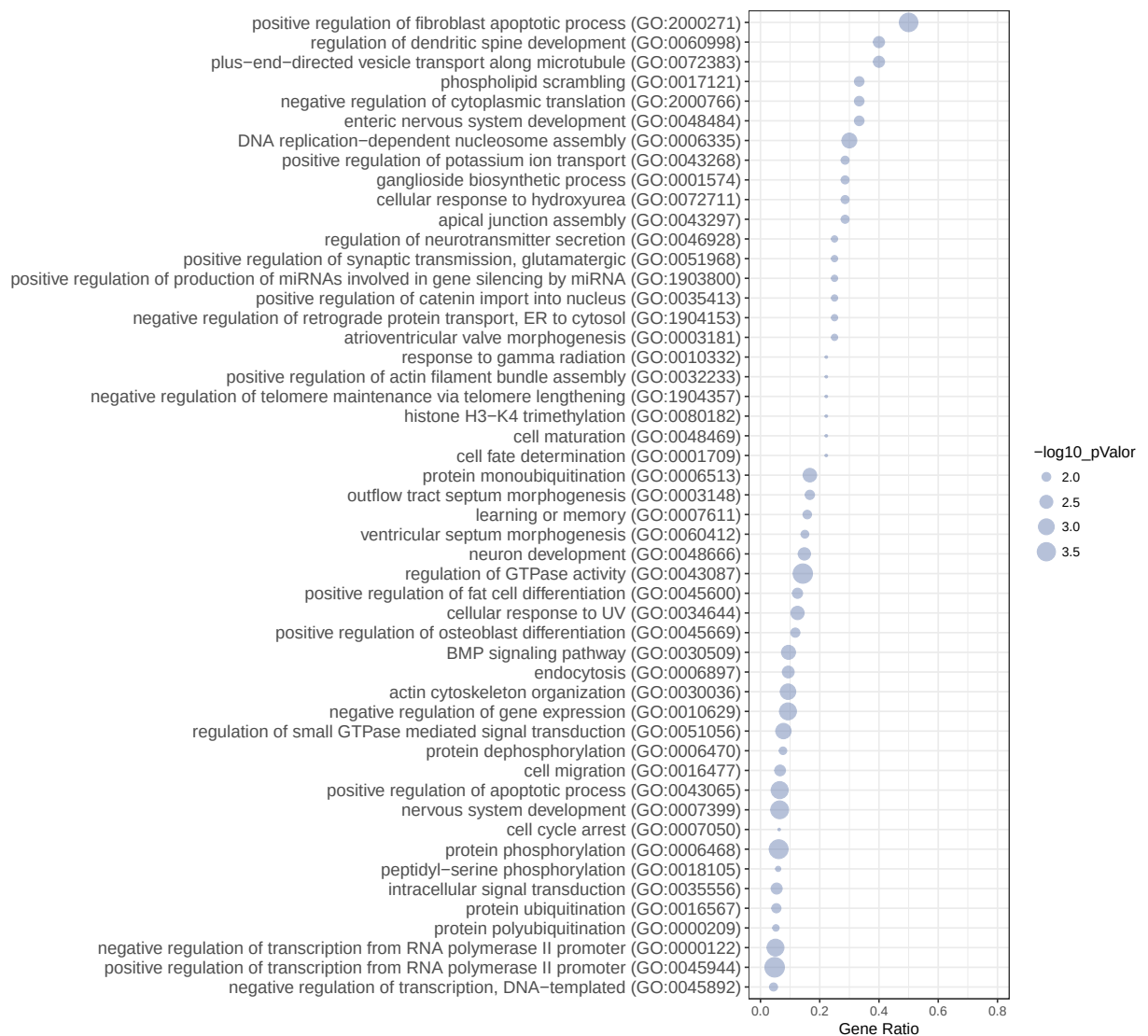


Figura 39: Términos de Gene Ontology asociados con los genes blanco en común a los cluster C19MC, miR-302 y miR-371-373: A partir de los 487 genes blanco predichos, analizamos los términos de GO de los procesos biológicos asociados a dichos genes. En el gráfico se muestran los primeros 50 términos en función del *gene ratio* y el tamaño de la burbuja representa el p-valor (*test de Bonferroni*).

3.3. Estudio del rol del *cluster* C19MC en las células pluripotentes

Una vez realizado el análisis del *cluster* C19MC mediante herramientas bioinformáticas, decidimos analizar de manera experimental cuáles son los genes regulados por los microARNs del *cluster* y estudiar de ésta manera el rol que estos tienen en las células pluripotentes humanas. Para ello, y gracias a la colaboración con el grupo de investigación del Dr. Mercola en la Universidad de Stanford, realizamos un ensayo de co-inmunoprecipitación de los ARN

mensajeros unidos a la proteína del complejo RISC *Argonauta* (Ago) seguido de *RNA-Seq* (*Ago-IP-Seq*). Ésta técnica se basa en sobreexpresar un microARN en particular con un oligonucleótido mimic, luego co-inmunoprecipitar los ARN mensajeros que están interactuando con Ago en el complejo RISC, y finalmente determinar su identidad por secuenciación masiva de ARN. El fundamento de la técnica se basa en que al sobreexpresar un microARN, el complejo de silenciamiento se enriquece en los genes blanco reconocidos por el microARN, permitiendo de esta manera identificar principalmente los genes que este regula. Es importante destacar que es necesario realizar los controles adecuados para evitar falsos positivos, los cuales consisten en co-inmunoprecipitar en paralelo con un anticuerpo IgG para determinar uniones inespecíficas y sobreexpresar en paralelo un mimic control negativo que permita identificar los genes que interactúan con la proteína Ago en condiciones normales. En base al análisis de los genes blanco predichos, elegimos estudiar los ARN mensajeros *target* del miR-520a. En particular, este microARN forma parte de una subfamilia dentro del *cluster* C19MC denominada miR-520, de los cuales varios han sido identificados por el grupo del Dr. Mercola como reguladores de la proliferación celular en cardiomiocitos derivados de CMPi (datos aún no publicados). Por ende, estudiamos los genes blanco del miR-520a tanto en células pluripotentes como en cardiomiocitos derivados de CMPi. Los CM fueron obtenidos mediante un protocolo de diferenciación en monocapa basado en la modulación de la vía de Wnt (ver detalle en materiales y métodos). En la figura 40 se muestra un esquema del experimento de Ago-IP.

Dado que el grupo del Dr. Mercola no contaba con la misma línea celular (H9), los experimentos fueron realizados en una línea de células pluripotentes inducidas humanas (CMPi). Por lo tanto, validamos la expresión de los microARNs en las CMPi y en los CM derivados de las CMPi mediante PCR en tiempo real (figura 41). Por otro lado, controlamos la eficiencia de transfección de las células mediante la transfección en paralelo de un pre-miRNA unido a un fluoróforo (Pre-miR-Cy3) (figura 42A), y confirmamos el aumento en la expresión del miR-520a mediante PCR en tiempo real (figura 42B).

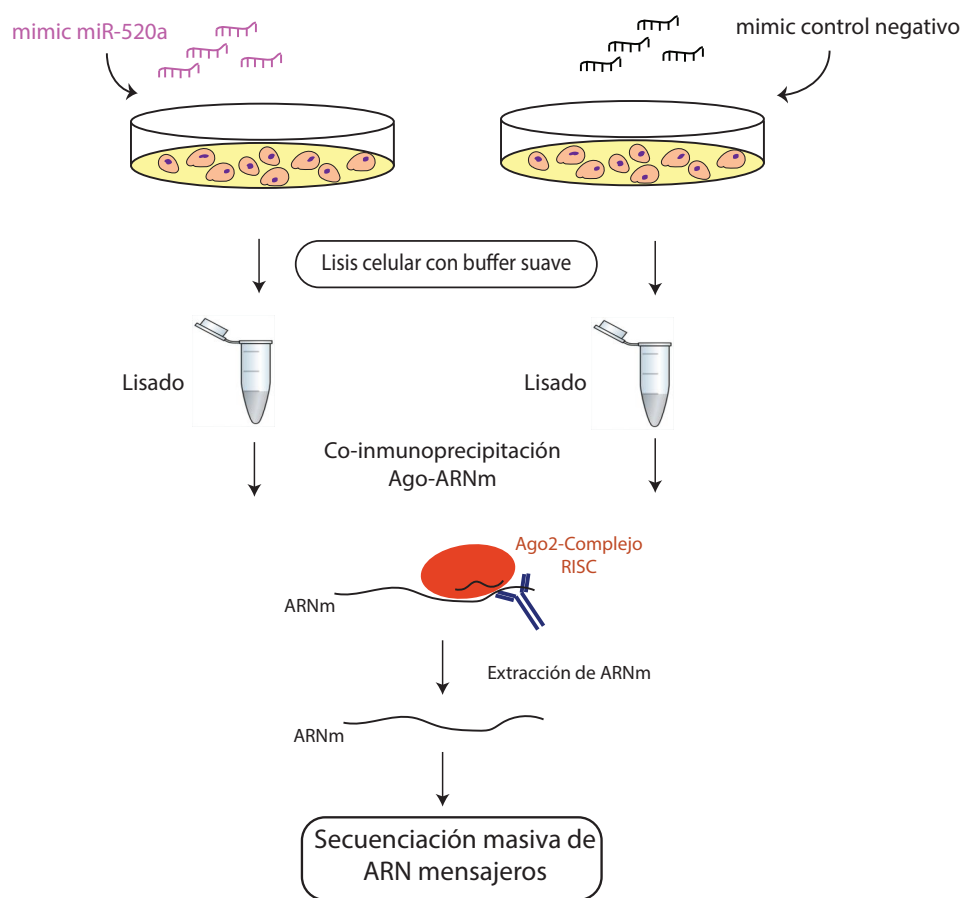


Figura 40: Esquema del protocolo de Ago-IP-Seq: Las células fueron transfectadas con el mimic del miR-520a o el mimic control negativo y a las 12hs post-transfección fueron lisadas con un buffer suave que permitió preservar las interacciones ARN-proteína. Luego, se co-inmunoprecipitó la proteína *Argonauto* (Ago) junto con los ARNm por medio de un anticuerpo específico para Ago que interactúa con la proteína G unida a *beads* magnéticas. A continuación se realizó la extracción de ARN mensajero y se determinó la identidad de los mismos mediante secuenciación masiva.

Una vez recibidos los datos de la secuenciación masiva, los resultados se analizaron siguiendo el *pipeline* que se muestra en la figura 43. El primer paso fue analizar las lecturas crudas (*reads*) mediante el programa FastQC para confirmar la calidad de los datos. En la figura 44 se muestra como ejemplo el informe de calidad para una de las muestras secuenciadas. Como se puede observar en el margen derecho de la imagen, cada uno de los controles fueron superados sin problemas (tilde verde) o alguna advertencia (signo de exclamación). Los dos controles que arrojaron un error (cruz roja) están relacionados con un bajo nivel de calidad

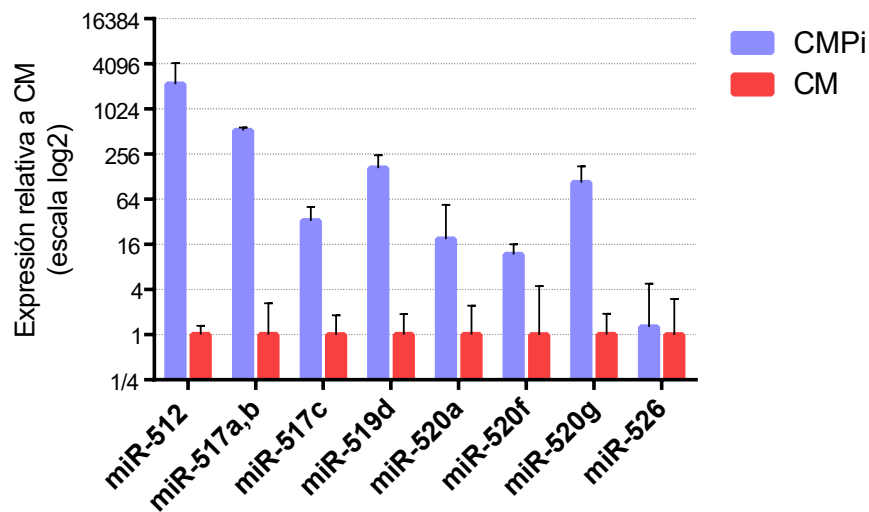


Figura 41: Análisis de los niveles de expresión de los microARNs del C19MC: Se analizó por PCR en tiempo real los niveles de expresión de los microARNs del C19MC en las células CMPi y los CM derivados de CMPi. Los niveles de expresión están relativizados los niveles de expresión en CM, las barras representan la media $\pm$ el SEM para tres y dos replicas de las CMPi y CM respectivamente.

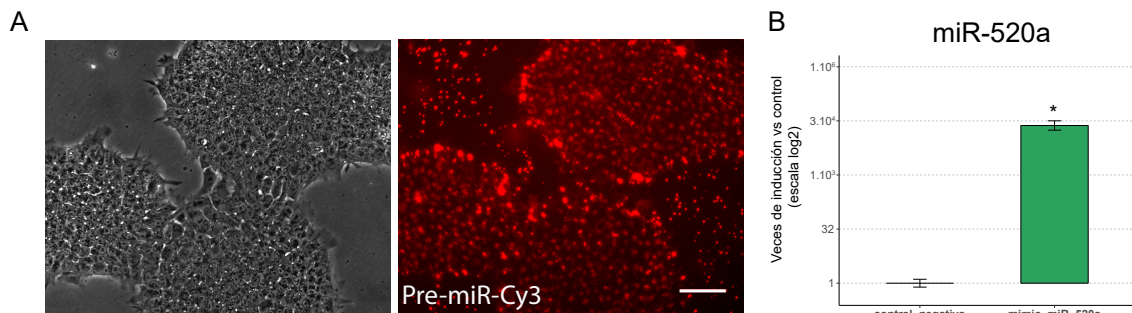


Figura 42: Sobreexpresión del miR-520a en las células pluripotentes inducidas y cardiomiocitos: Las CMPi y CM fueron transfectadas con el mimic del miR-520a 10nM o un mimic control negativo 10nM. (A) La eficiencia de transfección fue controlada mediante la transfección con un pre-miRNA control que esta unido a un fluoróforo (Pre-miR-Cy3) y se analizó por microscopia de fluorescencia su presencia dentro de las células a las 48hs post-transfección (la barra representa 500 μ m). En la figura se muestra a modo de ejemplo la transfección en las CMPi. (B) El aumento en los niveles del miR-520a se determinó mediante PCR en tiempo real a las 48hs post-transfección. Los valores de inducción fueron relativizados al control negativo y se muestra el valor de la media $\pm$ el SEM para un total de 3 replicas biológicas (* p=0,001).

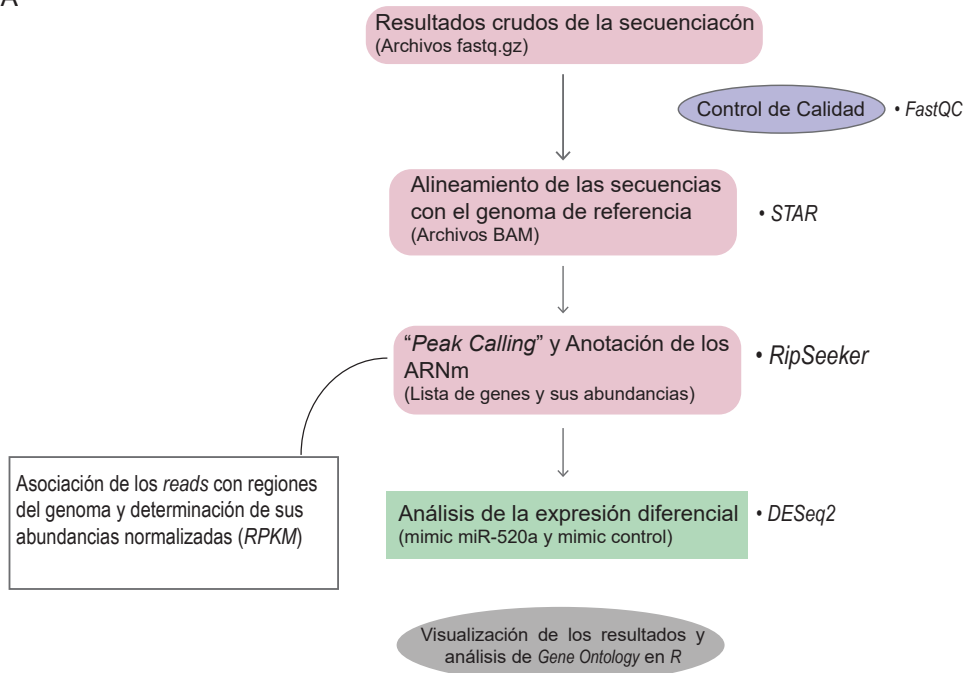
únicamente en las primeras 5 bases de las secuencias. Luego, una vez validada la calidad de los datos, se realizó el mapeo de las *reads* contra el genoma de referencia (hg19) utilizando el programa *STAR*. De esta manera identificamos la posición en el genoma de cada una de las *reads*. A continuación, mediante el programa *RIPSeeker* realizamos el "*peak calling*" e identificamos la identidad de las secuencias y sus abundancias absolutas y relativas al control (IgG). El "*Peak calling*" se basa en determinar las regiones donde existe un apilamiento de *reads* y diferenciarlo de la señal de ruido, que en este caso corresponde a las *reads* obtenidas en el control IgG (el análisis realizado es análogo al análisis de un experimento de ChIP-Seq). En la figura 43B se muestra a modo de ejemplo los "picos" de reads identificados para el gen FZD5 en una de las muestras de células pluripotentes. Finalmente, identificamos aquellos mensajeros que se unieron diferencialmente a Argonauta al sobreexpresar el miR-520a versus el mimic control negativo, mediante el uso del paquete *DESeq2* de Bioconductor.

El resultado de la secuenciación masiva de ARN reveló que, en las células pluripotentes hay 489 genes que se unen diferencialmente a *Argonauta* cuando se sobreexpresa el miR-520a con respecto al control negativo. En la tabla 4 se muestran algunos de los genes junto con el cambio relativo con respecto al control negativo ($\log_2FC$) y los estadísticos asociados (la lista completa está en la tabla suplementaria 17, ver tabla). A partir de esta lista, analizamos los términos de *Gene Ontology* con el fin de identificar los procesos biológicos en los que participa el miR-520a. En la figura 45 se observan los términos GO identificados, el número de genes asociados a los mismos y el p-valor para cada término.

A partir del análisis de los términos de Gene Ontology identificamos varios procesos biológicos relacionados con el desarrollo embrionario y en particular varios procesos relacionados con la diferenciación hacia el linaje cardíaco, como la diferenciación a mioblasto, la morfogénesis del pericardio, de los tabiques cardíacos y de la válvula pulmonar, así como también vías de señalización como la vía de Wnt y la vía de Notch, ambas muy relacionadas con los mecanismos moleculares que regulan la diferenciación cardíaca. En la tabla 5 detallamos los genes para los primeros 20 términos GO que presentaron un p-valor menor. Estos resultados coinciden con el análisis *in silico* del C19MC y sugieren que el miR-520a participaría de la regulación de genes necesarios para la diferenciación hacia el linaje cardíaco, lo cual es consistente con el perfil de expresión del miR-520a y del *cluster* C19MC, cuyos niveles disminuyen drásticamente al iniciar la diferenciación a mesodermo.

Por otro lado, a partir de la sobreexpresión del miR-520a en los cardiomiocitos derivados

A



B

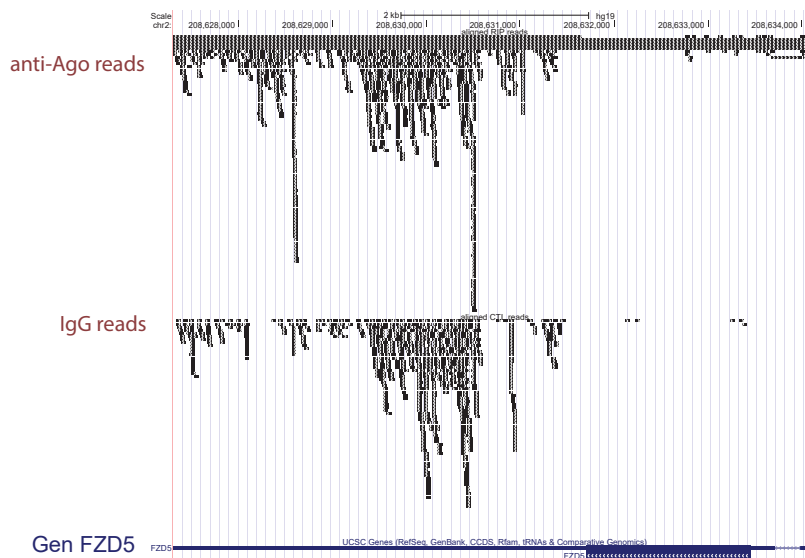


Figura 43: Esquema del *pipeline* del análisis de los resultados de AgoIP-Seq: (A) En primer lugar se determinó la calidad de los datos mediante el programa FASTQC. Luego se alinearon las secuencias (*reads*) con el genoma de referencia (Hg19) con el programa STAR, y se realizó el "*peak calling*", anotación y cuantificación de los genes con el programa RipSeeker. Por último, se realizó el análisis diferencial de los genes co-inmunoprecipitados al sobreexpresar el miR-520a vs el control, mediante el programa DESeq2. El análisis posterior de los términos de ontología (GO) se realizó en R. (B) Visualización en el Genome Browser de UCSC del locus de Fzd5 y el "pico" o apilamiento de reads identificado para la muestra inmunoprecipitada con un anti-Argonauta (anti-Ago) y con el anticuerpo IgG. El gen FZD5 es uno de los genes que interactuó diferencialmente más con Argonauta al sobreexpresar el miR-520a.

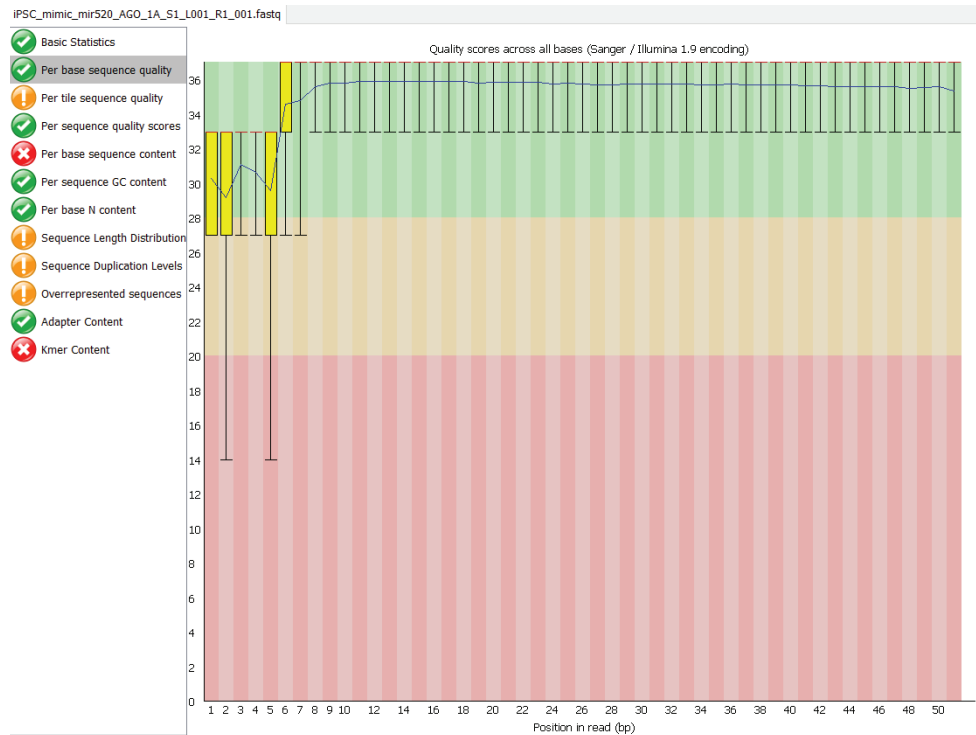


Figura 44: Control de la calidad de los resultados de secuenciación: Se analizó la calidad de las *reads* mediante el programa FastQC. A modo de ejemplo se muestra en la figura el informe del análisis para una de las muestras del experimento de AgoIP. En la figura se puede observar el *score* de calidad en función del tamaño de los *reads*, el cual fue muy bueno ya que la mayoría de las secuencias caen dentro de la región verde del gráfico.

de las células pluripotentes, encontramos que 152 genes interactúan con Argonauta de manera diferencial. Es importante destacar que en este experimento no contábamos con replicas biológicas, por lo cual los resultados son preliminares. A pesar de esto, realizamos un contraste adicional dado que contábamos con los ARNm que interactúan con Argonauta en las células sin transfectar, es decir comparamos los genes identificados al sobreexpresar el miR-520a con el mimic control negativo y con las células sin transfectar y nos quedamos con aquellos genes que presentaron diferencias mayores a un log2FC de 1 en ambos contrastes. En la tabla suplementaria 18 se puede ver la lista completa de genes (ver tabla). Al analizar los términos GO relacionados con los 152 genes identificamos distintos procesos biológicos relacionados con el metabolismo, la división celular y algunos procesos de diferenciación asociados con desarrollo del linaje hematopoyético (figura 46). En la tabla 6 se detallan los genes asociados con algunos de los términos GO que presentaron un p-valor menor.

Gen (Official symbol)	log2FC	p-valor	p-adj
PXDN	10,77282	1,61E-12	1,58E-09
PAK2	10,78332	2,27E-12	1,69E-09
XPO5	10,6498	4,99E-12	3,11E-09
GGA2	10,34927	3,98E-11	1,94E-08
TCF7L1	10,05417	3,53E-10	1,48E-07
TOPBP1	10,24769	4,76E-10	1,86E-07
C6orf62	10,0127	5,28E-10	1,93E-07
TOB2	9,928333	8,13E-10	2,81E-07
URB1	9,603721	7,9E-09	2,21E-06
SALL4	9,475611	1,54E-08	3,76E-06
EPC2	9,393411	3,97E-08	8,02E-06
FIGN	9,289783	4,53E-08	8,27E-06
PRR12	9,271083	4,66E-08	8,27E-06
RPAP1	9,272595	4,82E-08	8,31E-06
LRP6	9,013139	1,98E-07	3,04E-05
AK000154	9,008687	2,31E-07	3,38E-05
MAML1	9,007904	2,39E-07	3,42E-05
AK025726	8,854773	4,59E-07	5,98E-05
WDFY2	8,964469	4,83E-07	6,15E-05
ITPR3	8,782918	6,36E-07	7,61E-05
EMP2	8,563903	1,67E-06	0,000178
ACO2	8,448661	2,76E-06	0,000276
PELO	8,454844	2,81E-06	0,000276
RUSC1-AS1	8,441655	2,82E-06	0,000276
FXR2	8,457451	4,31E-06	0,000375
LOC100652999	8,081365	1,24E-05	0,000935
TMEM109	8,078488	1,32E-05	0,000982
RB1	8,068379	1,61E-05	0,001167
PARP4	8,144304	1,70E-05	0,0012
ANKRD33B	7,993896	1,74E-05	0,001201

Tabla 4: Lista de genes *target* del miR-520a en las células pluripotentes: En la tabla se muestran los primeros 30 genes que presentaron un cambio relativo mayor con respecto al control (representado por el log2FC.)

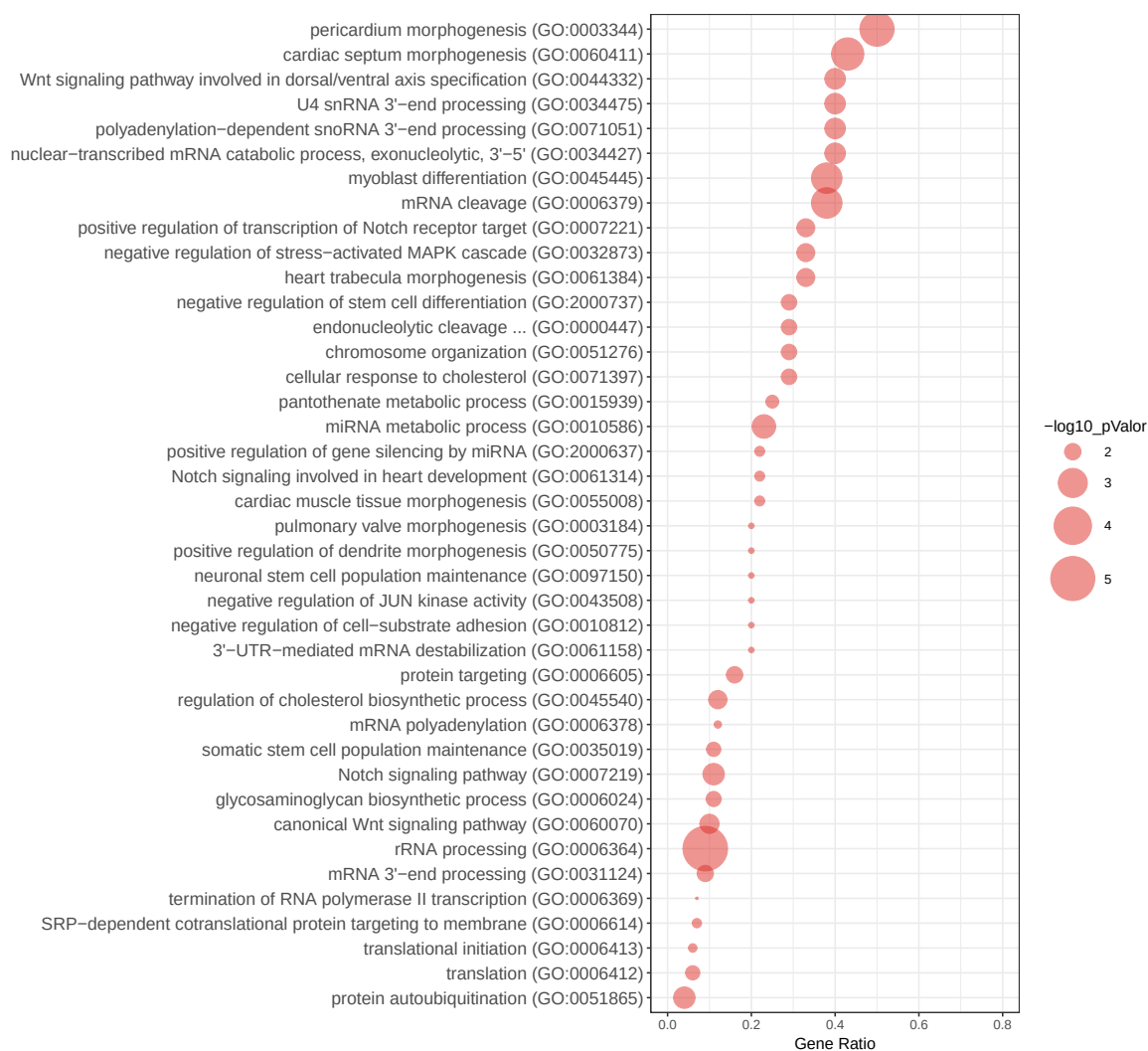


Figura 45: Términos de Gene Ontology asociados con los genes blanco del miR-520a en las PSC: En el dot plot se muestran algunos de los términos GO relacionados con los 489 genes blanco, en función del número de genes relacionados con cada término y los respectivos valores p (en escala $-\log_{10}$). El análisis se realizó mediante el paquete *Enrichr* (Kuleshov et al., 2016).

Término de Gene Ontology	Valor-P	Genes (Official symbol)
rRNA processing (GO:0006364)	8,2E-6	RPS5, NOP58, POP4, RPL11, RPL13A, RRP1B, PWP2, RPS26, EXOSC5, PDCCD11, RPS20, UTP20, RIOK2, EXOSC3, MDN1, RPL19
pericardium morphogenesis (GO:0003344)	0,0002	BMP7, NOTCH1, LRP6
cardiac septum morphogenesis (GO:0060411)	0,0004	JAG1, NOTCH1, BMP7
myoblast differentiation (GO:0045445)	0,0007	DDX17, RB1, JAG1
mRNA cleavage (GO:0006379)	0,0007	PCF11, CPSF1, WDR33
miRNA metabolic process (GO:0010586)	0,0033	TRIM71, DDX17, XPO5
Notch signaling pathway (GO:0007219)	0,0047	NOTCH1, JAG1, MAML1, MYC, EP300
protein autoubiquitination (GO:0051865)	0,0047	AMFR, TRIM71, BFR, RNF4, RNF187
U4 snRNA 3'-end processing (GO:0034475)	0,0055	EXOSC3, EXOSC5
polyadenylation-dependent snoRNA 3'-end processing (GO:0071051)	0,0055	EXOSC3, EXOSC5
nuclear-transcribed mRNA catabolic process, exonucleolytic, 3'-5' (GO:0034427)	0,0055	EXOSC3, EXOSC5
Wnt signaling pathway involved in dorsal/ventral axis specification (GO:0044332)	0,0055	LRP6, FZD5
canonical Wnt signaling pathway (GO:0060070)	0,0068	FZD5, TCF7L1, FZD4, MYC, LRP6
regulation of cholesterol biosynthetic process (GO:0045540)	0,0077	FASN, MBTPS1, PMVK, HMGCR
heart trabecula morphogenesis (GO:0061384)	0,0081	BMP7, NOTCH1
negative regulation of stress-activated MAPK cascade (GO:0032873)	0,0081	GSTP1, MYC
positive regulation of transcription of Notch receptor target (GO:0007221)	0,0081	MAML1, NOTCH1
protein targeting (GO:0006605)	0,0100	KATNB1, RPL11, GABARAP
mRNA 3'-end processing (GO:0031124)	0,0102	CPSF1, CHTOP, PCF11, SYMPK, WDR33
chromosome organization (GO:0051276)	0,0111	RLF, MYC

Tabla 5: Términos de Gene Ontology relacionados con los genes blanco del miR-520a en las PSC: En la tabla se muestran los 20 términos GO que presentaron un p-valor menor y los respectivos genes asociados, identificados como genes blanco del miR-520a.

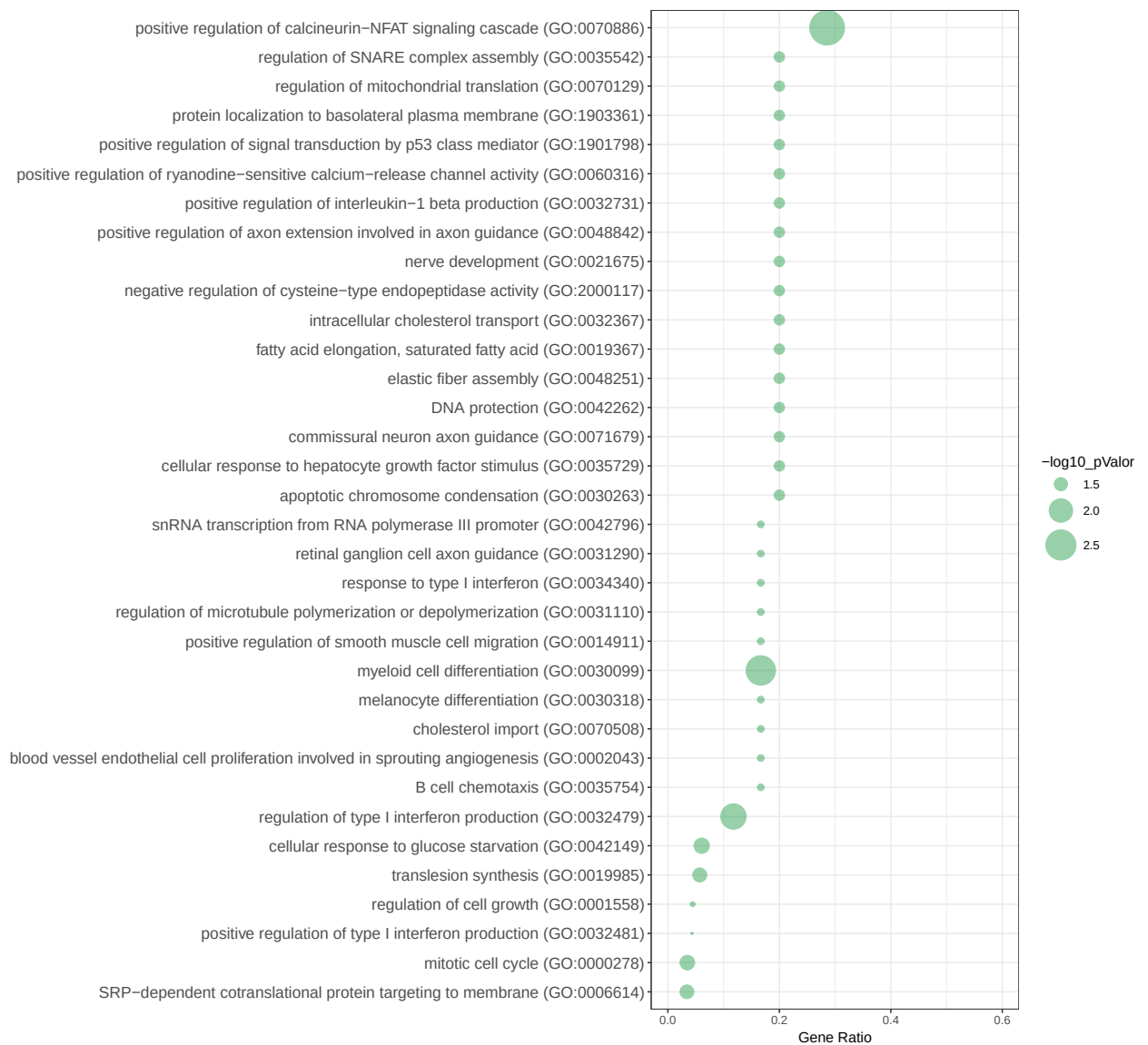


Figura 46: Términos de Gene Ontology asociados con los genes blanco del miR-520a en CM: En el dot plot se muestran algunos de los términos de GO relacionados con los 152 genes blanco, en función del Gene Ratio (número de genes observados/número de genes esperados) y los respectivos valores p (en escala $-\log_{10}$). El análisis se realizó mediante el programa *Enrichr* (Kuleshov et al., 2016).

Término de Gene Ontology	Valor-p	Genes (Official symbol)
positive regulation of calcineurin-NFAT signaling cascade (GO:0070886)	0.00119	LMCD1, AKAP6
myeloid cell differentiation (GO:0030099)	0.00365	TET2, IFI16
regulation of type I interferon production (GO:0032479)	0.00733	TRIM56, TBK1
cellular response to glucose starvation (GO:0042149)	0.0263	IFI16, NUA2
mitotic cell cycle (GO:0000278)	0.0283	NUDT15, SUGT1, SKA2
translesion synthesis (GO:0019985)	0.0293	POLK, ZBTB1
SRP-dependent cotranslational protein targeting to membrane (GO:0006614)	0.03	RPL35A, RPS3, RPL17
DNA protection (GO:0042262)	0.0377	NUDT15
negative regulation of cysteine-type endopeptidase activity (GO:2000117)	0.0377	IFI16
positive regulation of ryanodine-sensitive calcium-release channel activity (GO:0060316)	0.0377	AKAP6
nerve development (GO:0021675)	0.0377	NRP1
regulation of SNARE complex assembly (GO:0035542)	0.0377	VPS18
elastic fiber assembly (GO:0048251)	0.0377	MYH11
intracellular cholesterol transport (GO:0032367)	0.0377	STARD4
apoptotic chromosome condensation (GO:0030263)	0.0377	ACIN1
positive regulation of interleukin-1 beta production (GO:0032731)	0.0377	IFI16
positive regulation of signal transduction by p53 class mediator (GO:1901798)	0.0377	HEXIM1
regulation of mitochondrial translation (GO:0070129)	0.0377	NSUN3
positive regulation of axon extension involved in axon guidance (GO:0048842)	0.0377	NRP1
fatty acid elongation, saturated fatty acid (GO:0019367)	0.0377	ELOVL4

Tabla 6: Términos de Gene Ontology relacionados con los genes blanco del miR-520a en los CM: En la tabla se muestran los 20 términos GO que presentaron un p-valor menor y los respectivos genes asociados, identificados como genes blanco del miR-520a.

Luego, decidimos comparar los genes blanco del miR-520a identificados mediante el ensayo de Ago-IP con los genes blanco predichos *in silico*, tanto en las células pluripotentes como en los cardiomiocitos. En la figura 47 se muestra en los diagramas de Venn la superposición entre los genes identificados experimentalmente y de manera predictiva. Tal como se puede observar, en las células pluripotentes existen 114 genes en común entre los determinados por el Ago-IP y los predichos. En la tabla se detallan los genes identificados mediante ambas metodologías. Como es de esperar, el número de genes blanco predichos es mucho mayor ya que la predicción se realiza considerando todo el genoma y no solamente sobre los genes expresados en un contexto celular particular. Entre los *targets* identificados encontramos genes de la vía de Wnt (FZD4, FZD5, LRP6), genes relacionados con la regulación negativa de la transición G1/S del ciclo celular (RB1, POLR2A, NFE2L3, TRIP11, RPAP1, HCFC1), genes del desarrollo cardíaco (BMP7) y del sistema nervioso (PTPRD). Estos resultados sugieren que al menos 114 de los genes identificado por Ago-IP cumplen con los requisitos considerados por los algoritmos de predicción, es decir, no solo poseen la secuencia complementaria a la *seed* del miR-520a, y el contexto y acceso al sitio de *binding* por parte del microARN. En cuanto a los cardiomiocitos, encontramos que sólo 35 genes de los 152 son identificados como genes blanco del miR-520a mediante las herramientas bioinformáticas utilizadas. Algunos de ellos participan de la regulación del ciclo celular (CEP72, SDCCAG8), de la cascada de ERK1 y ERK2 (BMPER, NBR1), en la regulación de apoptosis (MDM4, ACIN1) y de procesos metabólicos como el catabolismo de fosfolípidos (CPT1A, PLA2G4B). Dado que el miR-520a no se expresa normalmente en las células cardíacas, la modulación de su expresión en los cardiomiocitos podría tener un efecto en la expresión de éstos genes que no se da en condiciones fisiológicas, sino que serían una consecuencia de forzar el sistema al sobreexpresar el microARN.

Es importante aclarar que dada la naturaleza exploratorio del ensayo Ago-IP, es necesario confirmar en un futuro mediante otras técnicas los genes blanco identificados. A pesar de esto, los resultados del análisis de *Gene Ontology* confirman la especificidad del experimento al encontrarse genes relacionados con la inhibición de la diferenciación celular.

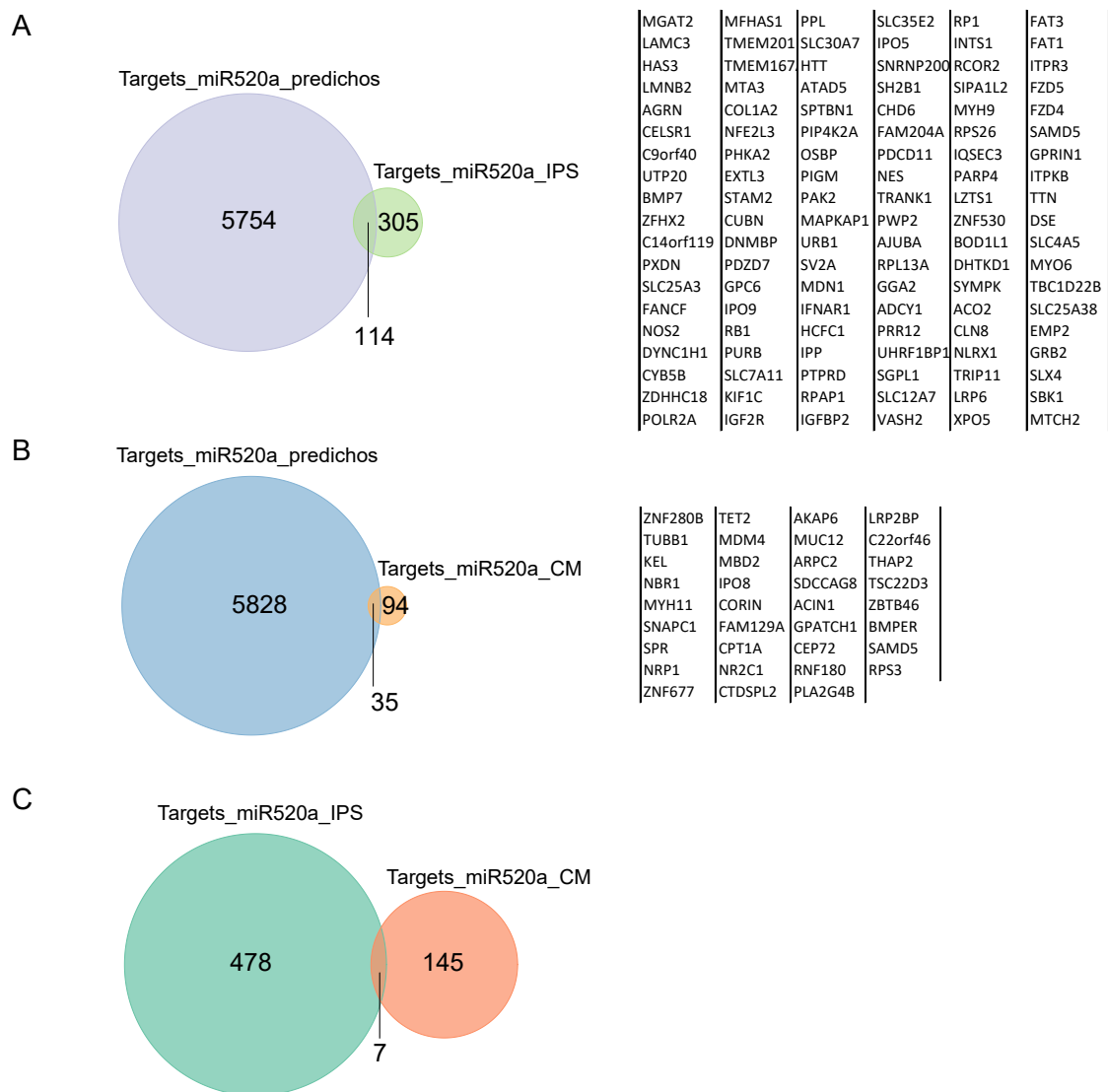


Figura 47: Comparación de los genes blanco del miR-520a identificados *in silico* y experimentalmente: En los diagramas de Venn se puede observar el número de genes blanco del miR-520a identificados mediante el ensayo de Ago-IP y los genes *target* predichos con herramientas bioinformáticas (*miRWalk*, *TargetScan*, *Pictar2* y *RNA22*). Se consideraron para la predicción sitios de regulación tanto en las regiones 3'UTR, 5'UTR y regiones codificantes (cds). En (A) se comparan los genes *target* predichos con los genes *target* identificados experimentalmente en las células CMPI, en (B) los genes *target* predichos y los genes *target* identificados experimentalmente en los CM y en (C) los genes *target* experimentales en CMPI y CM. En las tablas de genes se detallan los genes *target* en común de los diagramas de Venn en (A) y (B).

A partir de los genes *target* del miR-520a determinados en las células pluripotentes, tanto en el ensayo de Ago-IP como mediante la predicción *in silico*, encontramos que este microARN

podría regular varios genes relacionados con la muerte celular. Por consiguiente, decidimos analizar el efecto de bloquear la acción tanto del miR-520a como algunos otros microARNs del C19MC, en la apoptosis. Para esto, transfectamos las células pluripotentes con un inhibidor denominado miRCURY LNA *power family inhibitor*. Este LNA bloquea la acción de 9 microARNs del C19MC (miR-520a, miR-520b, miR-520c, miR-520d, miR-520e, miR-520f, miR-520g, miR-520h y miR-526). Nuestro objetivo fue bloquear el mayor número de microARNs posible, para lograr reducir una probable compensación generada por la redundancia que existe entre los microARNs del *cluster*. Cabe destacar que, no fue posible combinar más LNAs que bloqueen al C19MC por completo debido a que fue desaconsejado por la misma empresa que sintetiza los LNAs, a causa de la gran complementariedad que existe entre sus secuencias. Analizamos mediante el ensayo de Anexina V el efecto de transfectar las células pluripotentes con el LNA específico para la subfamilia 520 y un LNA control inespecífico, al cabo de 24, 48 y 72hs post-transfección (figura 48). Como se observa en la figura, a pesar de bloquear una gran cantidad de microARNs del C19MC, no encontramos cambios significativos en la apoptosis al bloquear la subfamilia miR-520 con respecto al control negativo.

Por otro lado, decidimos analizar el efecto de bloquear la subfamilia miR-520 en la proliferación de las células pluripotentes, ya que tanto en los resultados de Ago-IP como en la predicción de genes *target* identificamos genes relacionados con el ciclo celular tales como Rb1, Sox15 y CTGF (determinados por Ago-IP), y el factor E2F5 y las ciclinas D1 y D2 (en el análisis *in silico*). Para poder evaluar cambios en la proliferación celular transfectamos las células con el LNA 520 o el LNA control y a las 48hs analizamos el perfil del ciclo celular mediante la incorporación de BrdU y la tinción del ADN con 7-AAD. Al igual que en el ensayo de apoptosis, no encontramos diferencias significativas en los porcentajes de células en fase G1, fase S y fase G2/M entre las células transfectadas con el LNA 520 y el LNA control (figura 49).

En conjunto, estos resultados podrían explicarse debido a una inhibición parcial del C19MC, ya que al transfectar las células con el LNA 520 sólo se bloqueó la acción de algunos de los miembros del *cluster*, es decir, la acción del resto de los microARNs podría haber enmascarado el efecto de bloquear la subfamilia miR-520 en la apoptosis y la proliferación celular. Debido a este motivo, nos propusimos eliminar el *cluster* C19MC completo mediante la técnica de edición génica CRISPR/Cas9 y evitar así efectos compensatorios generados por la gran redundancia que existe entre los microARNs del C19MC.

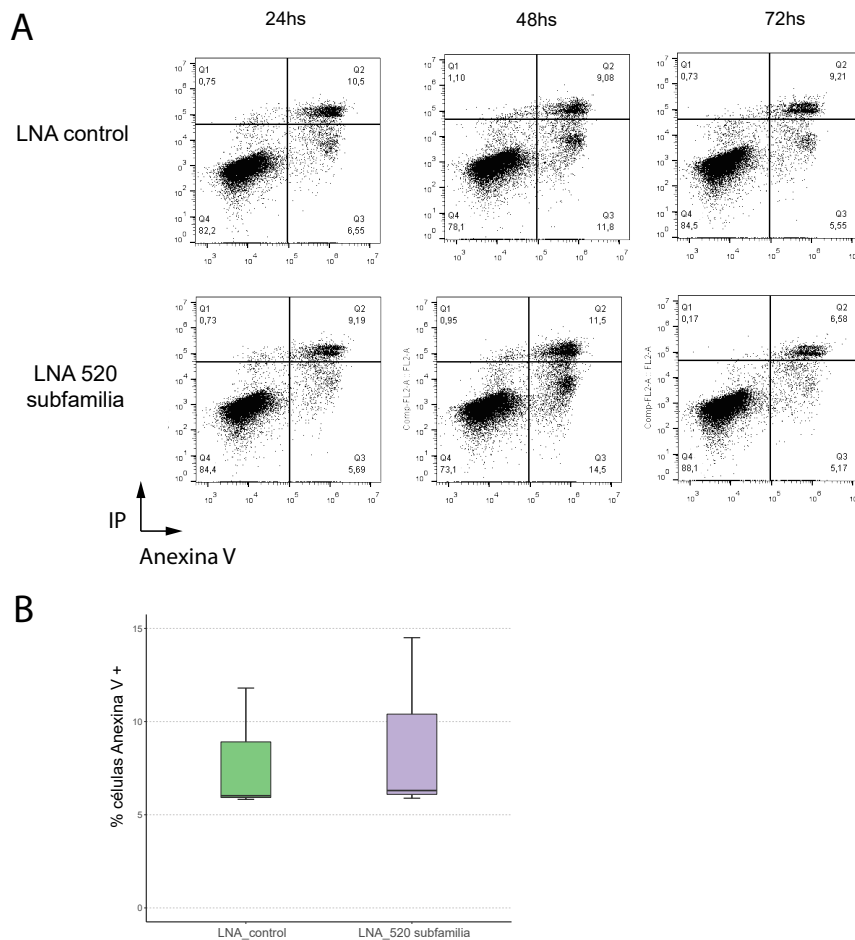


Figura 48: Ensayo de apoptosis al inhibir la subfamilia del miR-520: (A) Los dot plot son representativos de una réplica de la transfección de las células pluripotentes con el LNA control o el LNA 520 a las 24, 48 y 72hs post-transfección. Las células necróticas son positivas para el yoduro de propidio y las apoptóticas positivas para la Anexina V. En (B) se muestra el porcentaje de células positivas para Anexina V a las 48hs post transfección. En el box plot se representa la mediana y los cuartiles 25 y 75 para tres réplicas biológicas.

La técnica de edición génica CRISPR/Cas9 se basa en introducir en las células una endonucleasa de origen bacteriano, denominada Cas9, la cual produce cortes doble cadena en el ADN de manera dirigida. Para ésto, es necesario diseñar un ARN guía (sgRNA) que dirija a la endonucleasa hacia el sitio específico que se quiere modificar en el genoma, el cual debe estar próximo a una secuencia corta reconocida por la Cas9, denominada secuencia PAM (del ingles *protospacer adjacent sequence*). Una vez reconocido el sitio por complementariedad de secuencia con el sgRNA, la Cas9 genera un corte doble cadena en el ADN y el mismo es reparado posteriormente por la maquinaria propia de las células introduciendo inserciones/deleciones

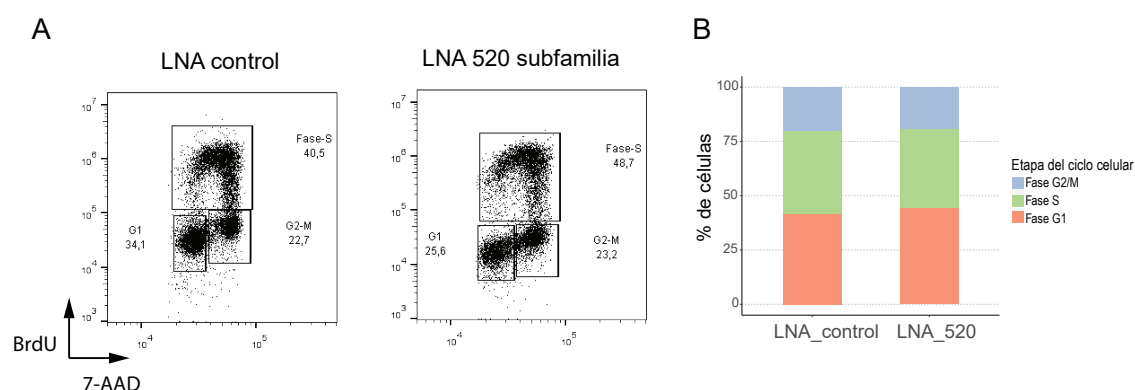


Figura 49: Análisis del perfil del ciclo celular al bloquear la subfamilia miR-520: Analizamos la incorporación de BrdU y el contenido de ADN luego de transfectar las células con el LNA 520 o el LNA control y determinamos mediante citometría de flujo las células positivas para BrdU con un anticuerpo anti-BrdU-FITC y el contenido de ADN mediante la tinción con 7-AAD a las 48hs luego de la transfección. En (A) se muestra un *dot plot* representativo de una de las réplicas y en (B) se observa el promedio del porcentaje de células en las distintas fases del ciclo celular para 5 réplicas biológicas.

en la zona reparada. El sistema de edición génica CRISPR/Cas9 ha sido utilizado extensamente en los últimos años y distintas variantes han sido diseñadas para ampliar su uso en diferentes estrategias (figura 50).

Una de las aplicaciones exitosas del sistema CRISPR/Cas9 ha sido la generación de deleciones grandes y/o rearrreglos cromosómicos. En el trabajo de Essletzbichler y col. (Essletzbichler et al., 2014) se demostró que el diseño de dos gRNAs que dirigen la endonucleasa Cas9 a sitios distanciados en el genoma permite deleccionar una región enorme de 30 megabases, generando así una línea celular haploide para el cromosoma 15. Además, es posible generar una deleción grande e insertar simultáneamente un gen reportero, tal como se demostró en el trabajo de Zhang y col., en el cual lograron eliminar una región de 65kpb e insertar el gen reportero *LacZ* en células madre embrionarias de ratón (Zhang et al., 2015a). La estrategia de deleción se basa en la probabilidad de que las células reparen mediante el mecanismo de NHEJ (del inglés, *Non Homologous End Joining*) los extremos alejados generados por el corte doble cadena. Debido a que la probabilidad de que los extremos alejados se unan entre sí es baja, es necesario analizar un gran número de clones. En el trabajo de Canver y col. se determinó que la frecuencia de deleción de fragmentos del genoma mediante CRISPR/Cas9, es inversamente proporcional al tamaño del fragmento que se busca eliminar (Canver et al., 2014).

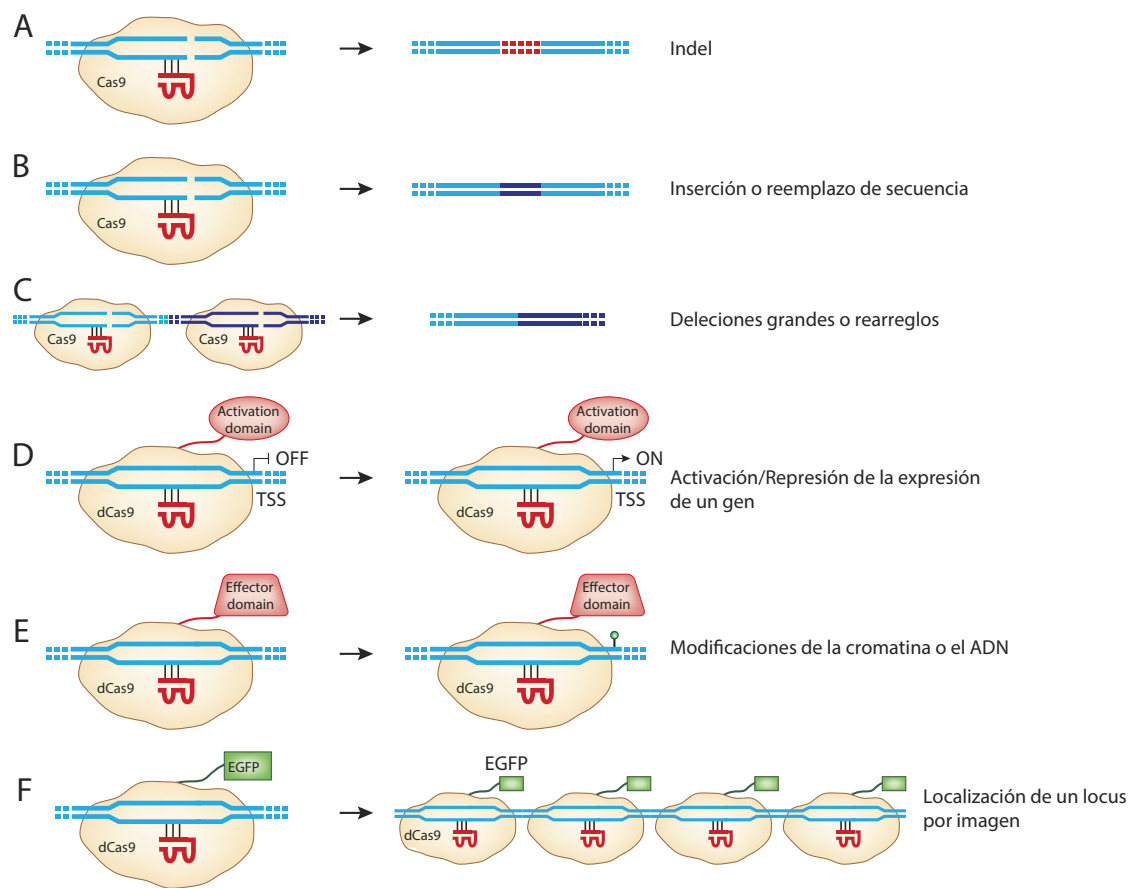


Figura 50: Aplicaciones del sistema de edición génica CRISPR/Cas9: En el es quema se muestra un resumen de distintas aplicaciones del sistema CRISPR/Cas9. (A) Mediante el uso de un sgRNA específico se puede dirigir la Cas9 hacia un sitio determinado y generar *indels* (insecciones/deleciones) o bien la inserción de una secuencia determinada por recombinación homóloga (B). (C) La introducción de dos sgRNAs diferentes puede provocar grandes deleciones o incluso rearrreglos cromosómicos. (D) Se puede regular la expresión de determinados genes mediante el sistema ON/OFF en donde se fusiona una versión mutada de la enzima Cas9 con un activador o represor de la transcripción o incluso modificadores de la cromatina, para acetilar o metilar histonas (E). (F) La fusión de la enzima Cas9 mutada con una proteína fluorescente como por ejemplo GFP permite visualizar por microscopía la posición de determinado locus en el genoma. La figura fue adaptada del trabajo de (Sander and Joung, 2014)

En este trabajo, nos propusimos deletar una zona grande de 100kpb mediante el diseño de dos gRNAs que dirigen la endonucleasa Cas9 a un sitio río arriba y río abajo del *cluster* C19MC (gRNA START y gRNA END, respectivamente) (figura 51). Para esto, las células fueron co-transfectadas con dos plásmidos, cada uno portador de un gRNA distinto y de la endonucleasa Cas9. A las 24 hs post-transfección seleccionamos las células transfectadas con

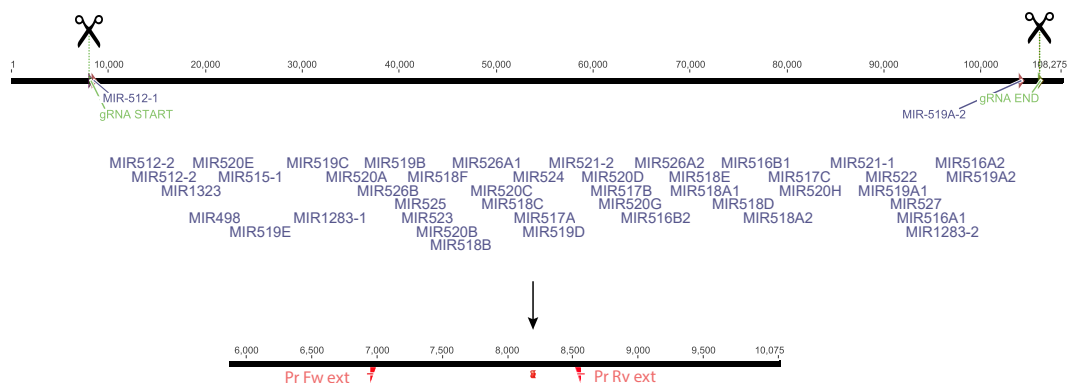


Figura 51: Esquema de la estrategia CRISPR/Cas9 para eliminar el C19MC: En el esquema se muestran los sitios de reconocimiento por los dos gRNAs (*START* y *END*) y su ubicación con respecto al *cluster* C19MC, así como también la posición de los microARNs del *cluster* (se respeta el orden pero no está a escala). Al eliminarse la región comprendida entre los dos gRNAs, queda una cicatriz en el genoma (marca roja) y los *primers* Fw y Rv externos permiten amplificar por PCR un producto de aproximadamente 1600pb.

puromicina y controlamos la eficiencia de transfección con un tercer plásmido portador del gen de la proteína fluorescente GFP. Luego, para poder identificar las células en donde ocurrieron los dos cortes y la delección del *cluster* entero, diseñamos un par de *primers* que reconocen a cada lado del *cluster*, es decir a una distancia aproximadamente de 100 kpb entre ellos. De esta manera, sólo obtuvimos un producto de PCR en aquellas células que eliminaron el *cluster* por completo. Además, diseñamos otros pares de *primers* adicionales que reconocen regiones dentro del *cluster* para controlar que éste se haya eliminado efectivamente.

Las células transfectadas post-selección con antibiótico fueron analizadas por PCR para evaluar si efectivamente el *cluster* había sido eliminado. Obtuvimos de esta manera un producto de PCR con los *primers* complementarios a las regiones flanqueantes del *cluster* ("Pr ext"), pero también con los *primers* control, lo cual era esperado ya que era una mezcla de células. Luego, aislamos las células y las crecimos de manera clonal para seleccionar el clon en el cual sólo obtengamos producto de PCR con los *primers* flanqueantes y no los *primers* control. Este *screening* fue realizado para 50 clones, de los cuales sólo dos dieron positivo en la PCR (4% de eficiencia). Sin embargo, la PCR con los *primers* control también dio positivo, por lo cual fue necesario re-aislar. Cabe aclarar que la separación clonal de éstas células es bastante compleja ya que es difícil poder crecer las células aisladas de manera unicelular. En el segundo

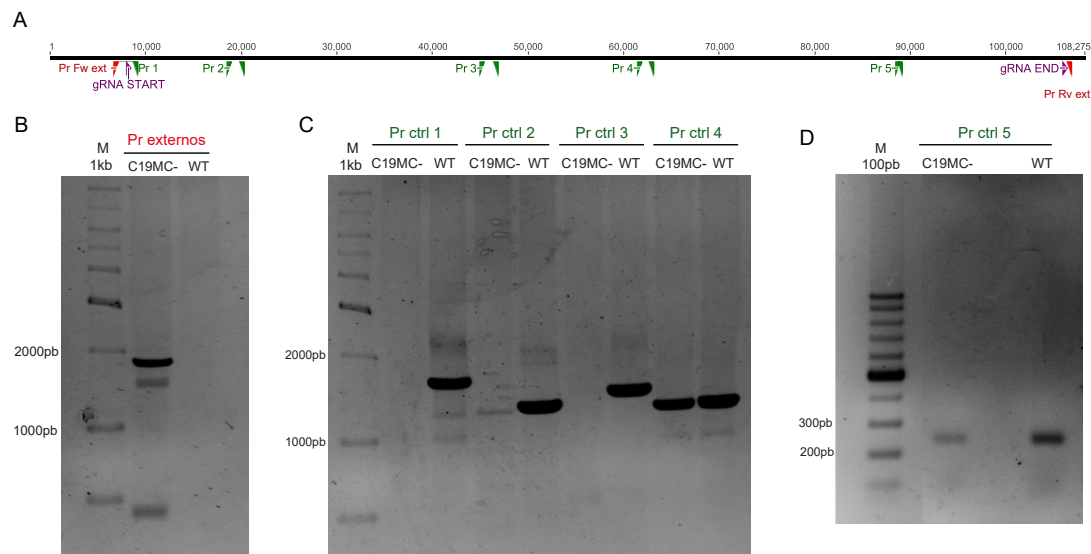


Figura 52: Análisis de la delección del *cluster* C19MC: (A) Esquema de los sitios reconocidos por los *primers* en las regiones por fuera y por dentro del *cluster*. (B) Resultado de la PCR con el par de *primers* externos (reconocen sitios río abajo y río arriba del *cluster*) en las células CMPi wt y CMPi C19MC-. (C) Resultados de la PCR con cuatro de los cinco pares de *primers* control (gel de agarosa 1 %). (D) Resultado de la PCR con el par de *primers* número 5.

screening obtuvimos un clon positivo de 14 clones analizados (figura 52B). El producto de PCR de 1600pb fue aislado y secuenciado para confirmar la identidad de la secuencia. El resultado de la secuenciación fue alineado con el genoma y coincidió con las regiones río arriba y río abajo del *cluster* C19MC. Además, diseñamos otros pares de *primers* que reconocen regiones dentro del C19MC para confirmar la delección del mismo (figura 52C y D).

Para nuestra sorpresa, dos de los cinco *primers* que reconocen regiones río abajo de la mitad del *cluster* (pares de *primers* 4 y 5) dieron positivo tanto en las células wt como en la línea mutada (C19MC(-)). Sin embargo, mediante PCR cuantitativa determinamos la ausencia de expresión de los microARNs del C19MC (figura 53). Por lo tanto, consideramos que la línea generada mediante CRISPR/Cas9 heterocigota para la delección completa de *cluster* o existe una mezcla de genotipos en la línea celular generada, pero ninguno con el C19MC completo.

A continuación, decidimos analizar el perfil del ciclo celular de las células con la delección del *cluster* C19MC (C19MC(-)) y compararlo con las CMPi wt. En la figura 54 se muestra el resultado del ensayo de yoduro de propidio para las CMPi wt y las CMPi C19MC(-). Este ensayo nos permitió determinar el contenido de ADN de las células, a partir del cual pudimos

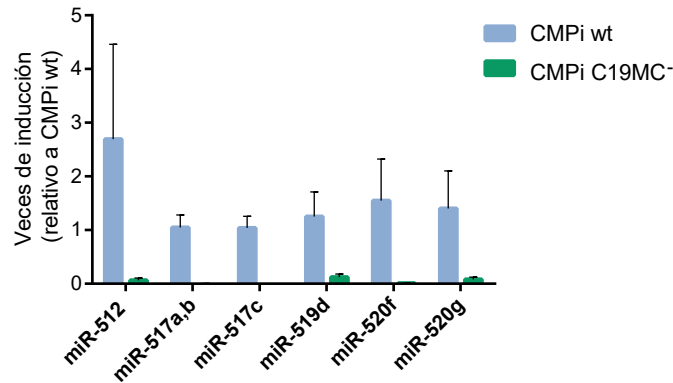


Figura 53: Análisis de la expresión de los microARNs del C19MC en las CMPi *C19MC*⁻: Se analizó por PCR en tiempo real la expresión de los microARNs miR-512, miR-517a,b, miR-517c, miR-519d, miR-520f y miR-520g. En el gráfico se muestra la media de tres réplicas biológicas $\pm$ el SEM. Los valores de expresión fueron relativizados a las CMPi wt.

determinar el porcentaje de células en fase G1, fase S y fase G2/M mediante el ajuste al modelo de Watson, realizado con el programa FlowJo. A partir de estos resultados identificamos un aumento significativo de 6,4% de las células en fase G1 de las CMPi C19MC(-) con respecto a las wt (p-valor= 0.03). Este resultado es consistente con el análisis de genes blanco predichos para *cluster* y con los resultados del ensayo de Ago-IP, en los cuales identificamos que el *cluster* C19MC estaría regulando negativamente genes relacionados con la transición G1/S del ciclo celular.

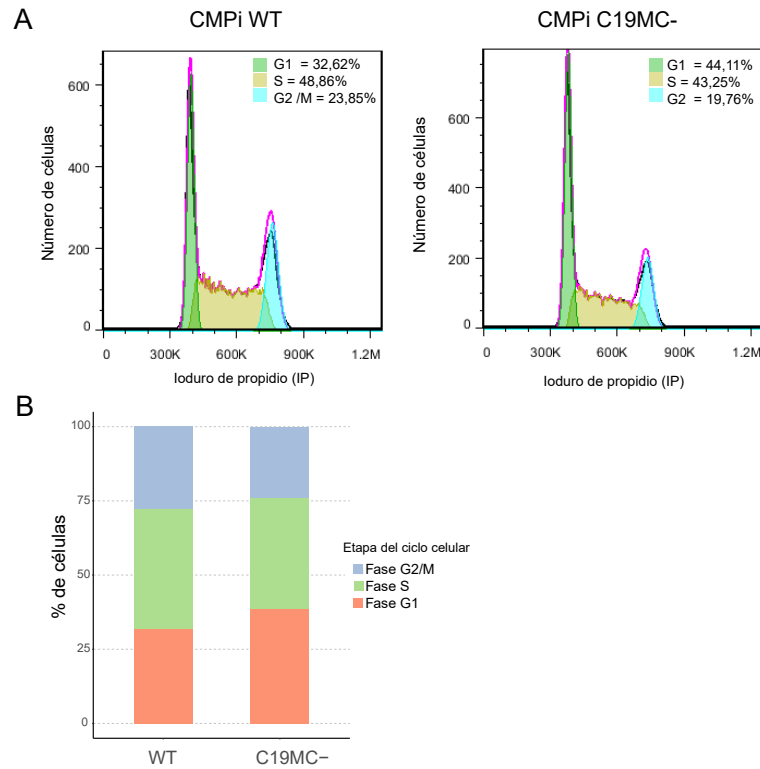


Figura 54: Análisis del perfil del ciclo celular del clon CMPi C19MC-: (A) Los histogramas representan el perfil del ciclo celular de las células CMPi wt (derecha) y las CMPi C19MC- (izquierda) para una de las 5 replicas biológicas realizadas. Arriba a la derecha se detallan los porcentajes de las células en cada fase, según el ajuste al modelo de Watson. (B) En el gráfico de barras se muestra el promedio del porcentaje de células en cada una de las fases para 5 réplicas biológicas de las células CMPi wt y CMPi C19MC-.

Conclusiones parciales del capítulo:

En este capítulo estudiamos en profundidad el *cluster* C19MC, formado por casi 60 microARNs y exclusivo de primates. Identificamos mediante secuenciación masiva 41 microARNs del *cluster* expresados exclusivamente en las células pluripotentes, y que una vez formado el progenitor de mesodermo temprano sus niveles de expresión disminuyen y se mantienen bajos incluso en los cardiomiocitos.

Dado que los microARNs del *cluster* están localizados en una misma región cromosómica exploramos de manera *in silico* los sitios de unión a la ARN polimerasa II y las marcas de transcripción activa. Identificamos que en las células pluripotentes humanas habrían sitios de unión de la ARN polimerasa II río arriba del *cluster*, así como también marcas de unión del factor de transcripción Nanog, unión de CTCF y p300. Además, encontramos marcas de histonas

acetiladas y di/tri metiladas, tanto al inicio como a lo largo del *cluster*. Estas observaciones sugieren que la región que comprende al C19MC estaría transcripcionalmente activa en las células pluripotentes humanas, lo cual va en línea con nuestros resultados de secuenciación. Por otro lado, mediante el alineamiento múltiple de sus secuencias identificamos que existen regiones muy conservadas entre la mayoría de los microARNs, lo cual a su vez se ve reflejado en el alto nivel de solapamiento que existe entre los genes blanco predichos que regularían. El análisis de los términos de *Gene Ontology* de dichos genes blanco, sugiere que los microARNs del *cluster* están involucrados en la regulación del desarrollo de diversos tejidos tales como tejidos del sistema nervioso, el musculo, el sistema renal y el esqueleto entre otros. Así como también en la regulación de importantes vías de señalización como es la vía de Wnt. Además, algunos microARNs del *cluster* poseen una gran similitud en el extremo 5' con la familia miR-302 y los miR-371-373, ambos grupos muy expresados también en las células pluripotentes. Al analizar los genes blanco predichos y los procesos celulares asociados a éstos, determinamos que los tres grupos de microARNs regularían 487 genes en común relacionados con procesos muy importantes en las células indiferenciadas como son el ciclo celular, la apoptosis y la diferenciación celular.

Luego estudiamos el rol del C19MC de manera experimental, para lo cual analizamos mediante un ensayo de co-inmunoprecipitación del ARN los genes blanco del miR-520a. Este ensayo nos permitió identificar que el miR-520a regula genes relacionados con la diferenciación celular, el desarrollo embrionario del corazón, y genes que participan de la vía de Wnt, la vía de Notch, de las MAPK, así como también genes involucrados en la regulación de la transición G1/S en el ciclo celular. Por otro lado, al estudiar los genes *target* del miR-520a en los cardiomiocitos, identificamos que este microARN podría regular genes relacionados con la diferenciación celular hacia otros linajes provenientes del mesodermo, tal como el linaje mioide, y genes relacionados con el desarrollo de otros tipos celulares de origen embrionario distinto, como el sistema nervioso. A su vez, podría regular genes involucrados en la mitosis y en el metabolismo de las células.

En base a éstos resultados y al análisis *in silico* del C19MC, estudiamos el efecto de inhibir la subfamilia del miR-520 en el porcentaje de células apoptóticas y en el perfil del ciclo celular. Sin embargo, a pesar de inhibir 9 miembros del *cluster* C19MC, no encontramos diferencias en la apoptosis ni en la proliferación de las células pluripotentes. Decidimos entonces, generar una línea celular que sea *Knock-Out* para el *cluster* C19MC (C19MC-) para evitar los posi-

bles efectos compensatorios entre los microARNs. Para ello, desarrollamos mediante la técnica CRISPR/Cas9 una línea que posee una delección grande en la región del C19MC que afecta la expresión de los microARNs del *cluster*. A partir de esta línea estudiamos de manera preliminar el perfil del ciclo celular de las células *C19MC*- y determinamos que las células que poseen la mutación en dicho *cluster* presentan un aumento en la fase G1 del 6% con respecto a las células wt. Este último resultado, contrastado con el resultado del LNA que bloquea la subfamilia miR-520, nos indicaría el gran efecto compensatorio que pueden tener los microARNs de una misma familia/cluster, y por lo tanto plantea el debate de la aproximación experimental que debe realizarse al estudiar los efectos de un microARN.

Capítulo 4:

Estudio del miR-205 en la diferenciación a mesodermo temprano

La familia del miR-200 y el miR-205 han sido estudiados extensamente en la regulación de la transición epitelio mesenquimal (TEM) que atraviesan las células durante el desarrollo tumoral. Tal como se desarrolló en la introducción, estos microARNs regularían negativamente la expresión de factores importantes en la TEM, tales como los inhibidores de E-cadherina, Snail1, Zeb1 y Zeb2 (Gill et al., 2011). Además, la familia del miR-200 ha sido identificada como una de las familias de microARNs característica de las células embrionarias humanas en estado de indiferenciación (Laurent et al., 2008).

En base a esto, previo a estudiar los niveles de expresión de los microARNs mediante secuenciación masiva, analizamos el perfil de expresión de la familia del miR-200 (miR-200a, miR-200b y miR-200c) y el miR-205 en nuestro modelo de diferenciación al progenitor temprano de mesodermo, como parte de la puesta a punto del análisis de los microARNs. Para ello, tomamos muestra de ARN en los días 0 y 3.5 del protocolo de diferenciación a mesodermo y analizamos por PCR en tiempo real la regulación de estos microARNs (figura 55). Tal como se puede observar en la figura, el miR-200a y el miR-200c no presentaron cambios significativos en sus niveles de expresión en el MPC con respecto a las PSC. Por el contrario, el miR-205 disminuye su expresión en el MPC, mientras que el miR-200b aumenta sus niveles de expresión en progenitor.

A partir del análisis de los niveles de expresión por qPCR y de lo reportado anteriormente en la bibliografía, decidimos estudiar el efecto de modular la expresión del miR-205 en el proceso de diferenciación hacia el progenitor de mesodermo temprano a partir de las células pluripotentes. Para esto, transfectamos las células con una molécula sintética que mimetiza el microARN (mimic miR-205) o con una molécula que se une de manera complementaria al microARN e inhibe su función (inhibitor miR-205). En una primera etapa, fue necesario poner a punto la concentración de los oligos sintéticos en la transfección de las células pluripotentes. Para esto, analizamos el aumento en los niveles de expresión del miR-205 al transfectar las células con 5, 10, 20 y 30 nM del mimic miR-205 o el mimic control negativo (figura 56A). Es

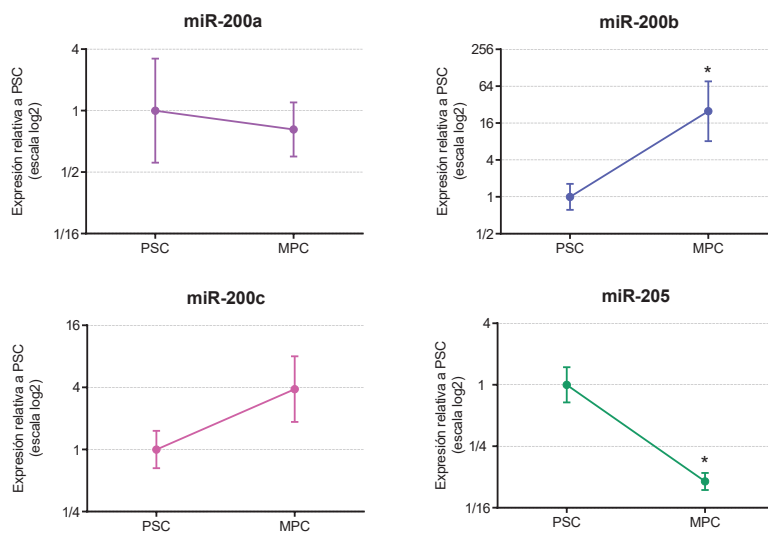


Figura 55: Análisis del perfil de expresión de la familia del miR-200 y el miR-205 en la diferenciación al progenitor temprano de mesodermo: En los gráficos se pueden observar los niveles de expresión relativa de los microARNs miR-200a, miR-200b, miR-200c y miR-205 en las células pluripotentes (PSC) y en el progenitor temprano de mesodermo (MPC), relativizados a las PSC. En cada gráfico se muestra la media $\pm$ el SEM de al menos tres réplicas biológicas. * p-valor <0.05.

importante aclarar que para el *inhibitor* del miR-205 no fue posible analizar por qPCR la disminución de la expresión del miR-205, ya que en realidad el oligo sintético bloquea la acción del microARN pero no promueve su degradación, es decir, no hay una disminución en su nivel de expresión. A su vez, analizamos los niveles de expresión del miR-205 a lo largo de los días para determinar su sobreexpresión durante la diferenciación al progenitor de mesodermo temprano (figura 56B). Tal como se puede observar, la concentración del miR-205 bajó conforme pasaron los días, pero aún se mantuvo alta durante los primeros 4 días, lo cual fue suficiente ya que el progenitor de mesodermo se analizó al día 3.5 post-inicio de la diferenciación.

En base a estos resultados, decidimos transfectar las células pluripotentes con el mimic miR-205 en una concentración 5nM y el miR-205 inhibitor con una concentración 100nM (máxima concentración recomendada por el fabricante) en el día -1 del protocolo de diferenciación a mesodermo. En el día 3.5 de iniciada la diferenciación, analizamos la formación del progenitor temprano de mesodermo (MPC) al modular los niveles del miR-205 mediante citometría de flujo. En la figura 57 se observa que al sobreexpresar el miR-205 aumentó casi dos veces el porcentaje de la población del progenitor de mesodermo (CD56+/CD326-). Por

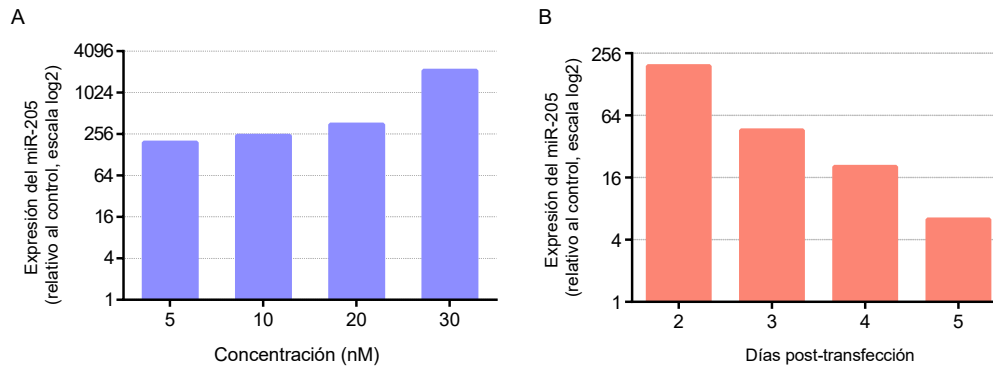


Figura 56: Sobreexpresión del miR-205 mediante la transfección del mimic miR-205: (A) Se analizó por qPCR la sobreexpresión del miR-205 a las 48hs post-transfección de las células pluripotentes, con un mimic miR-205 en las distintas concentraciones. Los valores de expresión son relativos a el mimic control negativo. (B) Se analizó la sobreexpresión del miR-205 durante 5 días post-transfección con la concentración 5nM del mimic miR-205. Los valores de expresión son relativos al mimic control negativo.

el contrario, al inhibir el miR-205 no obtuvimos cambios significativos en el porcentaje del progenitor.

A continuación, decidimos analizar el efecto de sobreexpresar el miR-205 en las células indiferenciadas (día 0) con el fin de identificar cambios en la expresión de posibles genes blanco, como también genes que indirectamente podrían ser afectados por la sobreexpresión del miR-205. Por lo tanto, estudiamos la expresión de los genes de pluripotencia (Oct-4 y Nanog) y de los genes relacionados con la TEM (Zeb1, Zeb2, Snail1, Snail2 y E-cadherina) a las 48hs post-transfección del mimic miR-205 (figura 58). Tal como se puede observar en el *box plot*, no se obtuvieron cambios significativos en la expresión de los genes estudiados, con respecto al control negativo.

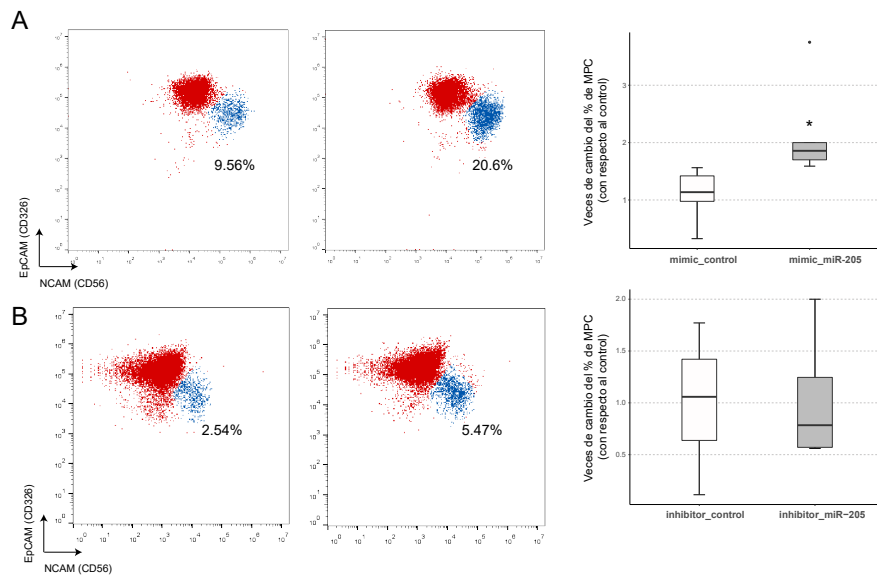


Figura 57: Modulación de la expresión del miR-205 en la diferenciación a mesodermo temprano: Se analizó por citometría de flujo el porcentaje del progenitor de mesodermo (MPC) luego de transfectar las células con el mimic miR-205 (A) o el inhibidor miR-205 (B), cada panel de citometría es una replica representativa de cinco replicas biológicas para el mimic miR-205 y cuatro replicas biológicas para el inhibidor miR-205. En los gráficos de *box plot* se muestra el promedio del cambio en el porcentaje del MPC con respecto al control negativo. * p-valor=0,04.

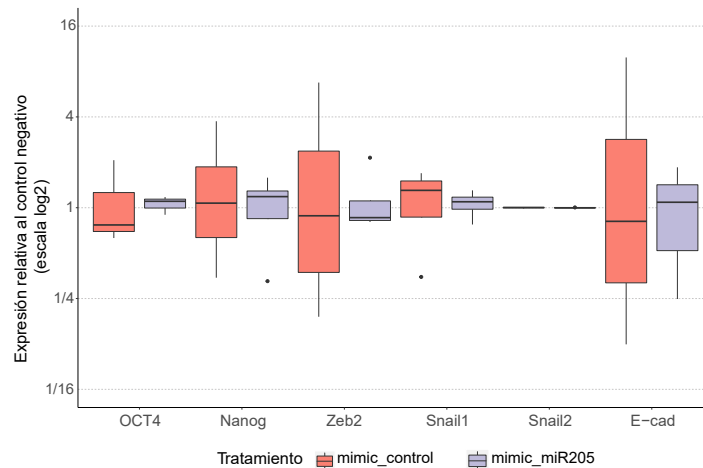


Figura 58: Análisis de la expresión de genes de indiferenciación y marcadores de la TEM al sobreexpresar el miR-205: Se analizó mediante qPCR la expresión de los genes de indiferenciación (Oct-4 y Nanog) y los genes marcadores de la TEM (Zeb2, Sanil1, Snail2 y E-cadherina). En el *box plot* se muestra la mediana de al menos tres replicas biológicas y los niveles de expresión fueron relativizados al mimic control negativo. *Nota: El nivel de expresión de cada replica fue relativizado a su propio control negativo para remarcar el efecto del mimic miR-205, ya que obtuvimos una alta variabilidad entre las replicas biológicas.*

Conclusiones parciales del capítulo:

En este capítulo estudiamos el perfil de expresión de la familia del miR-200 y el miR-205 en la diferenciación de las células pluripotentes (PSC) hacia el progenitor de mesodermo temprano (MPC), con el fin de estudiar su dinámica de expresión en nuestro modelo de estudio.

Identificamos que dentro de la familia del miR-200, el miR-200b presentó un aumento significativo en su nivel de expresión en el progenitor de mesodermo, mientras que el miR-200a y miR-200c no estarían siendo regulados en la formación de dicho progenitor. Por otro lado, determinamos que el miR-205 disminuye su nivel de expresión al diferenciarse las células al progenitor de mesodermo, lo cual va en línea con los trabajos reportados previamente, en los que se relacionó la expresión del miR-205 con un fenotipo del tipo epitelial, en donde la disminución de su expresión estaría relacionada con la TEM en un contexto tumoral.

No obstante, al sobreexpresar dicho microARN mediante la transfección de las células pluripotentes con un mimic miR-205, determinamos un aumento significativo del doble en el porcentaje de las células MPC con respecto al mimic control negativo. Este resultado fue inesperado para nosotros dado que en base a la dinámica de expresión observada en las PSC y MPC esperábamos obtener una disminución en el porcentaje de MPC. Estas observaciones serán discutidas en la próxima sección.

Por último, con el fin de comprender el rol del miR-205 en las células pluripotentes, estudiamos el efecto de sobreexpresar dicho microARN en la expresión de genes de pluripotencia (Oct-4 y Nanog) y marcadores de la TEM (Zeb2, Snail1, Snail2 y E-Cadherina). En este análisis no encontramos diferencias significativas en la expresión de dichos genes al sobreexpresar el miR-205, con respecto al control negativo.

DISCUSIÓN

4. Discusión

En las últimas décadas el estudio de las células pluripotentes humanas y sus potenciales usos en la medicina regenerativa ha estado en el centro de la atención de la investigación traslacional debido a su gran potencial en la aplicación en tratamientos de diferentes enfermedades y en su utilización en modelos de enfermedad. Estas células además, presentan muchas ventajas como modelo de desarrollo embrionario ya que permiten recapitular y estudiar los cambios que sufren las células durante la diferenciación celular en la etapa embrionaria. En particular, en el presente trabajo de tesis nos centramos en la formación de un progenitor de mesodermo temprano y la diferenciación hacia el linaje cardíaco a partir de las células pluripotentes humanas. Durante el proceso de diferenciación, diversos factores están involucrados en el establecimiento de la identidad celular, entre los cuales los microARNs tienen un rol fundamental. Estos pequeños ARNs participan en el silenciamiento post-transcripcional de genes que están relacionados con diversas vías de señalización durante el desarrollo embrionario, formando complejas redes de regulación.

El objetivo principal de este trabajo fue estudiar e identificar el conjunto de microARNs que estarían involucrados en el proceso de diferenciación hacia el mesodermo temprano y el linaje cardíaco, así como también estudiar el rol de determinados microARNs en las células pluripotentes y en la diferenciación celular.

4.1. Diferenciación hacia un progenitor temprano de mesodermo y hacia el linaje cardíaco

En la primera parte del trabajo describimos el establecimiento del modelo de estudio. Este modelo consistió en la diferenciación de las células pluripotentes humanas en un progenitor temprano de mesodermo y en cardiomiocitos inmaduros. Tal como se describió en el trabajo del Dr. Evseenko (Evseenko et al., 2010), el progenitor de mesodermo temprano identificado por primera vez en dicho trabajo tiene la característica de conservar la potencialidad de generar todos los tipos celulares que derivan del mesodermo. Esta población celular fue identificada por presentar en paralelo un aumento de la expresión de NCAM (CD56) y una caída en la expresión de la molécula EpCAM (CD326) (figura 2). El progenitor de mesodermo temprano que aparece en el día 3,5 de iniciada la diferenciación se caracteriza por haber sufrido uno de los procesos más tempranos del desarrollo del mesodermo, conocido como la transición

epitelio-mesenquimal (TEM). Durante éste proceso de cambio, las células pierden las uniones de contacto célula-célula y adquieren un fenotipo mesenquimal, incluyendo las características propias de una célula que tiene la capacidad de migrar. Tal como se puede observar en la figura 3, el progenitor temprano de mesodermo presentó niveles de expresión altos de los genes de mesodermo temprano *Mixl1*, *Brachyury* y *Mesp1*, así como también de los marcadores característicos de la TEM *Zeb1*, *Zeb2*, *Snail1*, *Snail 2* y *Vimentina* (figura 3). Este progenitor de mesodermo fue definido como el más temprano que aparece luego de que las células sufren la transición epitelio mesenquimal y dado su multipotencia para formar los distintos tejidos que derivan del mesodermo, representa un modelo de estudio adecuado para identificar los mecanismos que regulan tanto la formación del mesodermo como del resto de los tejidos que derivan del mismo, como por ejemplo los condrocitos articulares (Craft et al., 2015).

A su vez, diferenciamos las células pluripotentes hacia el linaje cardíaco basándonos en un protocolo de diferenciación *in vitro* en el que las células se agregan por centrifugación para formar los cuerpos embrioides (EB) y se cultivan en un medio definido suplementado con distintos factores de crecimiento (figura 5). Se cree que la estructura tridimensional de los EB permite recapitular lo que sucede en el desarrollo embrionario y facilita la interacción y la comunicación entre las células. Además, la formación de los EB por centrifugación a partir de un número determinado de células favorece la homogeneidad de los EB y una mayor eficiencia en el proceso de diferenciación (Ng et al., 2005). El protocolo de diferenciación cardíaca se basó en inducir durante los primeros días la formación del mesodermo con un medio suplementado con el factor BMP4 (*Bone Morphogenetic Protein 4*), Activina A y bFGF (*Fibroblast growth factor b*), y luego a partir del día 4 inhibir la vía de Wnt con el inhibidor IWR-1, para inducir de esta manera la diferenciación cardíaca. Este protocolo se basó en los trabajos de Yang y col. (Yang et al., 2008) y Willems y col. (Willems et al., 2011). En la figura 5 se observa como aumentan los niveles de expresión de *Brachyury* y *Mesp1* en el día 4 y en el día 14 los genes cardíacos cTNT, MYH6 y NKX2.5 alcanzan sus mayores niveles de expresión. Cabe destacar que a partir del día 10 se empiezan a observar regiones con actividad contráctil espontánea, lo cual define claramente la presencia de cardiomiocitos en el cultivo. El trabajo de Yang y col. fue uno de los pioneros en describir la diferenciación dirigida específicamente al linaje cardíaco. Sin embargo la eficiencia del protocolo era baja, ya que el porcentaje de cardiomiocitos obtenido mediante este protocolo era de un 20-40 %, dependiendo de la línea celular. Por este motivo para poder trabajar con un cultivo enriquecido en células cardíacas es nece-

sario purificar los cardiomiocitos, por ejemplo mediante selección con antibiótico o mediante FACS (*Fluorescent Activated Cell Sorting*). Posteriormente, se desarrollaron protocolos más sencillos y baratos que se basan en cultivar las células en monocapa y modular la vía de Wnt con pequeñas moléculas en determinadas etapas de la diferenciación (Lian et al., 2012). En la actualidad los protocolos más utilizados se basan en ésta última estrategia.

Por otro lado, los microARNs tienen un rol esencial en la regulación de la expresión génica durante el desarrollo embrionario y la cardiomiogénesis. En la figura 7 se observa que la familia del miR-302 mantuvo un nivel de expresión alto en el día 4 de diferenciación, para luego bajar en el día 14. Esta familia de microARNs ha sido identificada como una de las familias clave en el mantenimiento de la pluripotencia, regulando procesos muy importantes como la división celular y la apoptosis (Zhang et al., 2015b). El miR-145 presentó niveles significativamente más altos en el día 14, lo cual era esperado ya que este microARN regula factores clave de la pluripotencia como Oct-4, Sox2 y Klf4 (Xu et al., 2009). A su vez, determinamos que el miR-1, miR-133 y miR-208 presentan niveles altos de expresión en el día 14 de la diferenciación. Estos microARNs han sido muy estudiados en el desarrollo del linaje cardíaco, en particular el miR-1 y el miR-133a,b han sido identificados como reguladores clave en el desarrollo del corazón (Chen et al., 2004; Liu et al., 2007) y el miR-208 interviene en la respuesta a estrés regulando procesos como la hipertrofia y fibrosis en el corazón (van Rooij et al., 2007). Asimismo, se ha determinado que éstos tres microARNs están des-regulados en una situación patológica como es el infarto del miocardio, lo cual podría estar relacionado, según lo autores del trabajo, con una reprogramación de la expresión de los genes cardíacos y una activación de genes que se expresan en la etapa fetal (Boštjančič et al., 2010).

4.2. Identificación del miRNoma relacionado con la diferenciación al progenitor de mesodermo temprano y al linaje cardíaco

Los microARNs cumplen un rol fundamental en el desarrollo embrionario y en particular en el desarrollo del corazón (Cordes and Srivastava, 2009), ya que regulan los mecanismos que establecen la identidad celular (Ivey et al., 2008). En el presente trabajo estudiamos el perfil de expresión de los microARNs en la diferenciación cardíaca mediante secuenciación masiva de ARNs pequeños, con el fin de determinar el miRNoma de las células pluripotentes humanas (PSC), del progenitor temprano de mesodermo (MPC) y de los cardiomiocitos (CM)

(figura 12). Tal como se puede observar en *heatmap* cada población presentó un patrón bien definido en la expresión de los microARNs, lo cual sugiere que la expresión de los mismos está muy regulada en el proceso de diferenciación celular, relacionado con el rol que tienen en la determinación del destino celular. En el 2010 ya se conocían algunos cientos de microARNs en el genoma humano y el rol que éstos tienen principalmente en la regulación de factores que dirigen el desarrollo hacia los distintos tipos celulares (Ivey and Srivastava, 2010). Al comparar las tres poblaciones celulares, observamos que en base a la expresión de los microARNs, las células PSC y MPC estarían más relacionada entre sí que los CM. Esto sugiere que existe una dinámica en la expresión de los microARNs, la cual se hace más evidente en los CM, probablemente debido a que son células temporalmente mucho más alejadas de las PSC y MPC en el desarrollo embrionario (figura 14).

Los microARNs forman complejas redes de regulación génica. Parte de esta complejidad se debe a que un mismo microARN regula simultáneamente cientos de ARN mensajeros y un mismo ARN mensajero es regulado a la vez por muchos microARNs. Esta manera de regular la expresión génica genera un nivel extra de complejidad en la represión de los genes blanco, lo cual refleja el riguroso control al que están sometidos determinados genes durante el desarrollo embrionario. Además, gracias al avance en la tecnología de secuenciación y en el análisis de datos, se han identificado variantes de los microARNs denominados isomiRs, lo cual adiciona un nivel más de complejidad dado que un microARN puede tener cientos de isoformas que refuerzan la regulación y expanden los genes blanco. Estas isoformas presentan variaciones en las secuencias de los microARNs canónicos, que pueden estar en los extremos 5' (isomiR-5'), 3' (isomiR-3') y a lo largo de la secuencia (SNPs). Inicialmente fueron consideradas errores de la secuenciación, pero en la actualidad se han identificado numerosas variantes de los microARNs que presentan niveles de expresión equivalente o mayores al de los microARNs canónicos (figura 15). Tal como se puede observar en la figura 15, los isomiRs-5' son menos abundantes que los isomiRs-3', lo cual estaría relacionado con la funcionalidad que tiene el extremo 5' en el reconocimiento de los ARN mensajeros *target* y el impacto que tendrían dichas modificaciones en las células (Neilsen et al., 2012). Sin embargo, modificaciones en el extremo 3' también podrían tener un impacto en las células ya que afectarían la estabilidad de los microARNs y la afinidad con la proteína Argonauta (Neilsen et al., 2012). Además, al analizar la abundancia de los isomiRs en las tres poblaciones celulares, no encontramos diferencias significativas en las mismas. Esto sugiere que existirían mecanismos generales relacionados con la modifica-

ción de los microARNs en la diferenciación cardíaca. Observaciones similares se hicieron en el trabajo de Morin y col. en el cual analizaron y compararon la expresión de los microARNs y los isomiRs entre las células embrionarias humanas y los EBs (Morin et al., 2008).

Con el fin de explorar el posible rol que tienen los isomiRs, estudiamos los genes *target* predichos para el miR-302a y sus isomiRs-5' (figura 17). Este caso representa un ejemplo en el que los isomiRs-5' son más abundantes que el microARN canónico ("exacto") y los genes *target* predichos para ambos son parcialmente distintos. A su vez, al analizar los términos GO asociados con los genes *target* del miR canónico y los términos GO al expandir la lista con los *targets* de los isomiRs-5', aparecieron nuevos términos relacionados por ejemplo con el desarrollo del sistema nervioso. Estos resultados son consistentes con la idea de que los isomiRs ejercen una función en las células que a veces difiere con la acción del propio microARN canónico, generando un gran impacto en la interacción y reconocimiento de los genes blanco (Tan et al., 2014; Gong et al., 2012; Ahmed et al., 2015). En base a esto, consideramos importante tener en cuenta en el estudio de la expresión de los microARNs, la presencia y abundancia de los isomiRs. Futuros experimentos son necesarios para comprender mejor el impacto que los isomiRs tienen en la regulación de los procesos celulares y la diferenciación celular.

Generalmente, el análisis de los microARNs y la descripción del miRNoma de las células se ha basado en el estudio de los microARNs individualmente. Sin embargo, tal como se ha desarrollado en el presente trabajo, los microARNs pueden ser agrupados en familias al presentar una gran similitud en la secuencia semilla, y en *clusters* al estar codificados en regiones muy próximas en el genoma, lo cual se refleja en muchos casos en la regulación de genes o vías de señalización en común. A partir del análisis de familias y *clusters*, identificamos que hay 41 familias y 63 *clusters* de microARNs que participan de la regulación génica en la diferenciación cardíaca. En las PSC, la familia más expresada es el miR-302, la cual coincide con el *cluster* Chr4.1 (denominado así en nuestro trabajo por ser el primer *cluster* en el cromosoma 4). Esta familia de microARNs ha sido extensamente estudiada en células pluripotentes, tal como mencionamos anteriormente en este trabajo. La familia del miR-515 perteneciente al *cluster* C19MC, presentó también niveles altos de expresión en las PSC (posteriormente en este capítulo discutiremos en detalle acerca de ésta). Otro *cluster* interesante que se expresa en las PSC y MPC es el Chr13.2 el cual está formado por las familias del miR-17, miR-19 y miR-25, los cuales se ha reportado que cumplen un rol importante en el desarrollo embrionario así como también en el desarrollo de distintos tipos de tumores (Mogilyansky and Rigoutsos,

2013; Xu et al., 2012). A su vez, en las MPC la familia del miR-26 (miR-26a y miR-26b) presentó niveles altos de expresión, la cual también está relacionada con la carcinogénesis (Gao and Liu, 2011). La expresión de estos últimos dos grupos de microARNs podría asociarse con la transición epitelio mesenquimal que ocurre durante la formación del mesodermo en la diferenciación celular, el cual es equivalente a lo que ocurre en un contexto tumoral. Por último, en los CM los cinco *clusters* que presentaron un mayor nivel de expresión fueron el Chr5.2, Chr11.7, Chr19.8, Chr8.7, Chr1.2. Estos *clusters* están formados principalmente por la familia del miR-143, la familia de let-7 y la familia del miR-30. En consistencia con trabajos anteriores, el miR-143 regula la diferenciación del músculo liso y la familia del let-7 disminuye su expresión en la diferenciación temprana, pero aumenta en el tejido cardíaco ya diferenciado (Ohtani and Dimmeler, 2011; Thum et al., 2008).

A partir de los genes blanco predichos para cada familia/*cluster* de los microARNs seleccionados, analizamos los términos de *Gene Ontology* asociados e identificamos procesos celulares relacionados con el desarrollo del corazón, desarrollo muscular y principalmente relacionados con el desarrollo embrionario. Es interesante destacar que, conforme las células van adquiriendo un fenotipo más "maduro" existe un aumento en la especificidad de los procesos regulados en cuanto a la etapa del desarrollo embrionario, lo cual reflejaría la función que tienen los microARNs en la determinación de la identidad celular. Por ejemplo, en las PSC los términos de GO asociados con el miR-302 abarcan una gran variedad de procesos de diferenciación, desde el desarrollo del corazón, el sistema óseo, el tejido muscular, el sistema nervioso y la diferenciación mesenquimal. En las MPC, los términos GO están relacionados con procesos celulares relacionados con el desarrollo muscular y el desarrollo del sistema nervioso y por el contrario, en los CM los genes blanco predichos están relacionados con procesos más específicos como el desarrollo del sistema urogenital y el desarrollo del tejido óseo. Estas observaciones sugieren que una vez que las células alcanzaron un fenotipo más maduro, los microARNs regularían procesos relacionados con tipos celulares específicos y en particular en los CM, se estarían silenciando genes relacionados con tejidos que derivan también del mesodermo. Esto es coherente con lo postulado en el trabajo de Ebert y col., en donde se planteó que durante el desarrollo embrionario los microARNs cumplirían la función de reprimir genes que se expresan en otros tejidos de progenitores en común, que tienen o bien una expresión residual o una expresión "*leaky*" que es necesario controlar para el correcto desarrollo del tejido (Ebert and Sharp, 2012).

Por otro lado, en el análisis de la correlación de los *scores* de afinidad, identificamos que en la mayoría de los casos la afinidad de cada microARN por su gen blanco es específica y diferente. Notablemente, al analizar los términos GO para cada microARN de la familia del miR-302, los procesos celulares identificados fueron distintos a los obtenidos en el análisis global con todos los miembros de la familia, lo cual sugiere que existe una colaboración entre ellos en la regulación de los genes blanco y que no son completamente redundantes. En base a esto, creemos que es importante considerar la expresión de las familias y *clusters* de los microARNs para entender mejor el rol que éstos tienen en la diferenciación cardíaca.

Finalmente, comparamos nuestros resultados de secuenciación de ARN pequeños con datos publicados y cargados en otras bases de datos, con el fin de reducir la lista de microARNs que específicamente contribuyen al desarrollo de los CM. A partir de este análisis, identificamos que hay 848 microARNs expresados diferencialmente entre las 6 líneas celulares estudiadas, de las que cada una presentó un determinado perfil de expresión de microARNs reforzando nuevamente la importancia que éstos tendrían en la determinación de la identidad celular. De esta manera redefinimos el miRNoma de los CM de manera más precisa (tabla 3). Algunos de los microARNs identificados han sido reportados previamente en otros trabajos, como el miR-1, el miR-499, el miR-133 y el miR-208 (Chistiakov et al., 2016), pero otros no han sido relacionados previamente con el linaje cardíaco, como por ejemplo el miR-675, miR-585, miR-1248 y el miR-509. Futuros experimentos son necesarios para comprender la importancia y el rol de éstos microARNs en la diferenciación cardíaca.

En conclusión, en el presente trabajo logramos determinar los microARNs que están involucrados en la diferenciación de las células pluripotentes humanas hacia el mesodermo temprano y al linaje cardíaco. Estos resultados permitieron expandir el conocimiento acerca de los microARNs y sus isomiRs, durante el proceso de diferenciación celular y el rol que éstos tienen en la regulación de mecanismos moleculares importantes para la determinación del destino celular. Dado que el miRNoma de las células involucra cientos de microARNs funcionalmente muy relacionados entre sí, creemos que es necesario para su estudio un abordaje a nivel *genome-wide* para poder comprender mejor el rol que tienen estos pequeños ARNs no codificantes. Asimismo, es necesario futuros ensayos funcionales que consideren todos los miembros de las familias/clusters de microARNs para lograr de esta manera determinar realmente la relevancia que estos tienen en el proceso de diferenciación celular.

4.3. Análisis del *cluster* C19MC en las células pluripotentes

El *cluster* C19MC está formado por casi 60 microARNs que comparten no solo la localización en el genoma, sino que también en base a la similitud de sus secuencias la mayoría forma parte de una misma familia de microARNs (la familia miR-515). Además, comparten una dinámica de expresión en la diferenciación a mesodermo y al linaje cardíaco a partir de las células pluripotentes humanas (figura 31). La regulación de la transcripción de los microARNs del C19MC ha sido un tema de bastante controversia ya que hasta ahora no se ha identificado una región promotora próxima al *cluster*. Varios trabajos han planteado diferentes hipótesis para explicar cómo se transcriben estos microARNs. En el trabajo de Bouchert y col. se propuso que las secuencias *Alu*, dispersas a lo largo de todo el *cluster*, reclutarían a la ARN polimerasa III favoreciendo la transcripción de los microARNs río abajo y más próximos a dichas secuencias (Borchert et al., 2006). Sin embargo, en el trabajo de Bortolin-Cavaille y col. se propuso que en realidad estos microARNs se generan a partir de un intrón que está presente en un gen no codificante, el cual es regulado por la ARN polimerasa II (Dance and Weber, 2009). Este último mecanismo coincidiría con el modelo más clásico de transcripción de los microARNs llevado a cabo por la ARN pol II. A pesar de estos dos trabajos, todavía no se conoce con certeza el mecanismo de transcripción de los genes de este *cluster*. En el presente trabajo, exploramos de manera *in silico* las marcas de cromatina activa y los sitios de unión de la ARN pol II y de co-factores de transcripción e identificamos sitios que podrían ser regiones promotoras río arriba del *cluster* (figura 33), pero no sitios de unión de la Pol III. Experimentos adicionales son necesarios para dilucidar y comprender de qué manera se transcriben los microARNs del C19MC, como por ejemplo, estudiar la estructura tridimensional de la cromatina en la región del *cluster*.

A su vez, identificamos de manera *in silico* que los microARNs del C19MC regularían genes en común, lo cual se representó gráficamente en la figura 35. Esta observación va en línea con la alta homología que existe entre las secuencias de dichos microARNs (figura 34) y es consistente con el perfil de expresión similar que tienen en la diferenciación cardíaca. Al estudiar los procesos celulares relacionados con los genes *target* predichos, determinamos que el C19MC participaría en la regulación de la vía de Wnt y del desarrollo de distintos tejidos como el sistema nervioso, el tejido muscular, el sistema urogenital y el desarrollo óseo. Además, un subgrupo dentro del C19MC presenta alta homología con la familia miR-302 y el

cluster miR-371-373, ambos grupos de microARNs característicos de las células pluripotentes y relacionados con el mantenimiento de la pluripotencia (Suh et al., 2004). Recientemente en el 2017, dos trabajos publicados por el mismo grupo estudiaron el *cluster* C19MC en células pluripotentes inducidas derivadas de células mesenquimales y determinaron que la sobreexpresión del miR-524-5p aumenta la eficiencia de reprogramación de fibroblastos humanos, al disminuir la expresión de los genes de la TEM, ZEB1 y SMAD4, y el gen TP53INP1, regulador de la apoptosis mediante la activación de p53 (Nguyen et al., 2017a,b). Estos trabajos y nuestros resultados sugieren que los microARNs del C19MC podrían tener un rol fundamental en la regulación de procesos celulares importantes para mantener el estado de indiferenciación de las células pluripotentes.

Por otro lado, estudiamos los genes blanco regulados por el miR-520a en las células pluripotentes humanas y en los cardiomiocitos mediante el ensayo de Ago-IP. De esta manera determinamos que en las células pluripotentes el miR-520a estaría involucrado en la regulación de genes relacionados con el desarrollo del corazón y del sistema nervioso, así como también de vías de señalización importantes en el desarrollo embrionario como la vía canónica de Wnt, la vía de Notch y la vía de MAPK (reguladora de la proliferación celular) (figura 45). Dentro de los genes *target* identificados, 114 coinciden con los genes *target* predichos, los cuales consideramos *targets* directos del miR-520a ya que poseen en sus secuencias la región complementaria a la secuencia semilla (figura 47). Por otro lado, los 305 genes restantes podrían ser considerados *targets* indirectos que se ven afectados por la sobreexpresión del miR-520a, es decir que cambios en la expresión de determinados genes (dado la sobreexpresión del miR-520a) permiten que estos ARN mensajeros sean reconocidos y cargados en RISC. Futuros ensayos de validación son necesarios para confirmar la interacción del miR-520a con los 419 genes identificados en el ensayo de Ago-IP. En los CM, identificamos que el miR-520a regularía 152 genes relacionados con procesos de diferenciación de tejidos provenientes del mesodermo así como también procesos metabólicos, el ciclo celular y la apoptosis. En base a estos resultados, creemos que sería interesante explorar si la sobreexpresión del miR-520a en los cardiomiocitos tiene un efecto en la proliferación de las células, ya que estas células normalmente se dividen durante el desarrollo pero luego su tasa de proliferación es muy baja una vez que los cardiomiocitos están maduros (Li et al., 1997).

A pesar de los resultados obtenidos en el análisis bioinformático y el ensayo de Ago-IP, al bloquear la acción de la subfamilia miR-520 no encontramos un efecto significativo en la

proliferación y la apoptosis de las células pluripotentes (figuras 48 y 49). Este resultado podría explicarse dada la redundancia y el solapamiento que existe entre los microARNs del C19MC, lo cual en este caso podría estar enmascarando el efecto de sólo bloquear la acción de un grupo de microARNs. A su vez, dado que los microARNs del C19MC tienen individualmente un nivel de expresión moderado, creemos que para obtener un efecto más evidente que nos permita comprender el rol de estos microARNs en las células pluripotentes es necesario inhibir la acción de todo el *cluster* por completo. En base a esto, generamos una línea celular mediante CRISPR/Cas9 en la que se eliminó todo o una gran parte del C19MC. Esta línea celular posee una delección que comprende toda la región del C19MC confirmado por PCR y secuenciación. Sin embargo, dado que detectamos por PCR la presencia de una región del *cluster*, creemos que la delección completa de los 100kpb ocurrió en heterocigosis o que la línea mutante generada no es completamente clonal y existe una mezcla de genotipos en donde en algunos la delección fue completa y en otros parcial. Es importante aclarar que para generar la línea clonal con células pluripotentes humanas es difícil crecerlas de manera unicelular, ya que a pesar de cultivarlas diluidas para que cada colonia provenga de una única célula, éstas tienden a migrar y a unirse entre células vecinas. A pesar de esto, al analizar la expresión de los microARNs del *cluster*, identificamos una ausencia de expresión en relación con las células wt. Esto podría deberse a tres motivos distintos en nuestro modelo. En primer lugar, la mayoría de las células C19MC(-) no tiene presente el *cluster* en su totalidad, y solo una parte menor conserva unas 40,000 bases al final del *cluster*. En segundo lugar, puede deberse además a una mutación en una región regulatoria necesaria para la transcripción del *cluster*. Finalmente, dado que la mutación es grande puede existir un mecanismo que afectó la estructura tridimensional de la cromatina, necesaria para la correcta transcripción de los microARNs. Cabe aclarar que en cualquiera de los dos casos, se logró generar una línea de células pluripotentes humanas con una delección grande que afecta la funcionalidad del *cluster* C19MC.

Finalmente, analizamos el perfil del ciclo celular de la línea *knock-out* del C19MC y determinamos que existe un aumento de las células en la fase G1 del ciclo celular con respecto a las células wt. Esto es consistente con nuestros resultados anteriores y lo reportado previamente en una línea celular de coriocarcinoma, en la cual determinados microARNs del C19MC promueven la proliferación celular inhibiendo el arresto de la transición G1/S mediante la represión de p21Cip1/Waf1 (Wu et al., 2010).

En conclusión, el *cluster* C19MC está involucrado en la regulación de la expresión génica

que controla en las células pluripotentes humanas los niveles de expresión de genes relacionados con el mantenimiento de la pluripotencia y la diferenciación celular. A partir de la línea *knock-out* generada con CRISPR/Cas9 logramos establecer un modelo de estudio que permite evaluar el rol que tiene el *cluster* C19MC, lo cual dada la complejidad de este grupo numeroso de microARNs es difícil de estudiar mediante ensayos más clásicos de pérdida de función, como los ensayos con los oligos LNA que sólo permiten bloquear la acción de algunos pocos miembros del *cluster*. Experimentos adicionales con esta línea nos permitirán comprender mejor la función del C19MC en las células pluripotentes humanas y en el proceso de diferenciación celular.

4.4. Estudio del miR-205 en la diferenciación al progenitor de mesodermo temprano

En el último capítulo del presente trabajo estudiamos el perfil de expresión de microARNs relacionados con la transición epitelio mesenquimal, en particular la familia del miR-200 y el miR-205. La familia del miR-200 ha sido identificada como una de las familias de microARNs muy relacionadas con la malignidad de los tumores, ya que su expresión se ha encontrado desregulada en distintos contextos tumorales (Park et al., 2008; Peter, 2009; Gill et al., 2011). Al igual que el miR-205, se ha determinado que estos microARNs reprimen la expresión de genes importantes de la transición epitelio mesenquimal (Gregory et al., 2008a). En las células pluripotentes humanas aún no se ha estudiado el rol que cumpliría el miR-205.

En nuestro modelo de diferenciación hacia el mesodermo temprano encontramos diferencias significativas en la expresión del miR-200b y del miR-205. El perfil de expresión del miR-205 es coherente con lo reportado previamente en la bibliografía, mientras que el perfil de expresión del miR-200b no. Dado que el miR-200b está localizado en el mismo *cluster* que el miR-200a y bajo los mismos elementos regulatorios, creemos que podrían existir mecanismos post-transcripcionales que afecten diferencialmente los niveles de expresión de estos microARNs. Por otro lado, determinamos que la sobreexpresión del miR-205 aumenta al doble el porcentaje de la población MPC (figura 57). Si bien estos resultados no son consistentes con el perfil de expresión en la diferenciación al progenitor de mesodermo, se ha reportado previamente que en determinados casos los microARNs actúan como activadores de la expresión génica, mediante el reclutamiento de la maquinaria transcripcional a zonas promotoras o fa-

voreciendo la traducción de los ARN mensajeros blanco cargados en RISC (Vasudevan et al., 2007), lo cual podría explicar que un aumento de la expresión el miR-205 afecte positivamente a la diferenciación a mesodermo. Por otro lado, sería interesante evaluar la dinámica de expresión del miR-205 a períodos más cortos de la diferenciación, ya que podría existir un aumento transitorio en las primeras horas de iniciada la diferenciación que no fue posible detectar en nuestro modelo experimental. Por último, es importante aclarar que los experimentos de sobreexpresión con los oligos sintéticos (mimics) generan una condición no fisiológica en donde se refuerzan las interacciones del microARN con los ARN mensajeros y los efectos observados pueden ser consecuencia de esta interacción "anormal" (Vidigal and Ventura, 2015). Al estudiar los niveles de expresión de genes de pluripotencia y genes relacionados con la TEM, luego de sobreexpresar el miR-205, no encontramos diferencias significativas. Este resultado podría ser interpretado de dos maneras diferentes, por un lado podría ser que los genes analizados no sean efectivamente genes blanco directos o indirectos del miR-205, o bien podría ser que la regulación de dichos genes sea a nivel de las proteínas, ya que en determinados casos los microARNs regulan los ARN mensajeros mediante el bloqueo de la traducción. En conclusión, para poder comprender mejor el rol del miR-205 en las células pluripotentes y la diferenciación a mesodermo es necesario realizar ensayos adicionales como por ejemplo un ensayo de Ago-IP que permita identificar los ARN mensajeros que se cargan en el complejo RISC al sobreexpresar el microARN.

5. Conclusiones generales

En resumen, las conclusiones generales que se desprenden de este trabajo de tesis son las siguientes:

- Al rededor de 700 microARNs presentan un patrón de expresión bien definido en la diferenciación de las células pluripotentes humanas hacia el progenitor temprano de mesodermo y hacia los cardiomiocitos, regulando mecanismos involucrados en la determinación de la identidad celular.
- En base a los niveles de expresión de los microARNs, se identificó que las células pluripotentes presentan una mayor similitud con el progenitor de mesodermo temprano que con respecto a los cardiomiocitos.
- Analizamos la expresión los isomiRs en la diferenciación hacia el mesodermo y el linaje cardíaco, e identificamos por un lado que las isoformas más abundantes son aquellas con modificaciones en el extremos 3', que no existen diferencias en la abundancia de los isomiRs en el proceso de diferenciación y que la modificación más abundante es la eliminación de tres nucleótidos y la eliminación de un nucleótido en los isomiRs-5' e isomiRs-3', respectivamente. A su vez, determinamos que las modificaciones dentro de la secuencia se dan mayoritariamente en el extremo 5'.
- Determinamos que el nivel de expresión de los isomiRs puede ser mayor que el de los microARNs canónicos, siendo este el caso del miR-302a, en donde los isomiR-5' son más abundantes y expanden tanto los genes *target* regulados así como también los procesos celulares relacionados.
- Determinamos el miRNoma de las PSC, MPC y CM en función de las familias y *clusters* de microARNs, los cuales presentan un patrón definido a lo largo del proceso de diferenciación que es característico de cada población celular.
- Exploramos mediante un análisis *in silico* el nivel de solapamiento de los genes blanco predichos para los microARNs de una misma familia/cluster, y determinamos que regularían procesos en común.

- Definimos el miRNoma de los cardiomiocitos en base a los microARNs más abundantes en estas células, pero que a su vez presentaron una expresión diferencial con respecto a las células pluripotentes, al progenitor de mesodermo y a seis tipos celulares distintos reportados en GEO.
- Determinamos que la mayoría de los microARNs que forman el *cluster* C19MC comparten su perfil de expresión en la diferenciación celular, presentando niveles de expresión altos en las células pluripotentes.
- Identificamos *in silico* que el C19MC regularía genes relacionados con procesos celulares importantes en la pluripotencia, como la proliferación y diferenciación celular.
- Analizamos experimentalmente los genes *target* del miR-520a en las células pluripotentes humanas y en los cardiomiocitos e identificamos que regularía genes relacionados con la vía de Wnt, la vía de Notch, el ciclo celular y procesos de diferenciación en las PSC, mientras que en los CM regularía genes relacionados con diferenciación celular, el metabolismo, la proliferación y la apoptosis.
- Generamos una línea *knock-out* para el *cluster* C19MC mediante CRISPR/Cas9 e identificamos que la ausencia de expresión de los microARNs de este *cluster* afecta la transición G1/S aumentando el porcentaje de células en G1.
- Estudiamos el perfil de expresión de la familia del miR-200 y el miR-205 en la diferenciación a mesodermo temprano e identificamos que el miR-200b aumenta significativamente su expresión, mientras que el miR-205 disminuye con respecto a las células pluripotentes.
- Determinamos que al sobreexpresar el miR-205 aumenta al doble el porcentaje de células del progenitor temprano de mesodermo derivado de células pluripotentes humanas.

MATERIALES Y MÉTODOS

6. Materiales y Métodos

6.1. Células utilizadas

Para la realización del presente trabajo utilizamos la línea de CMEh, WA09 (H9) adquirida del Insitituto WiCell de la Universidad de Wisconsin y las células pluripotentes inducidas otorgadas por la Universidad de Stanford, DCM wt y líneas de fibroblastos embrionarios de ratón (MEF, del inglés, murine embryonic fibroblast).

6.2. Co-Cultivo de CMEh sobre MEFi

Las CMEh fueron mantenidas en estado indiferenciado cultivándolas sobre una capa nutricia de MEFi, en un medio de cultivo denominado HES (DMEM/F12 (Gibco) suplementado con 20% *Knock Out Serum Replacement* (KSR) (Gibco), aminoácidos no esenciales (2mM concentración final), L-glutamina (2mM concentración final), 50 U/ml P/E, β -Mercaptoetanol 55mM (Gibco) y 4ng/ml bFGF (Invitrogen). El medio de cultivo fue renovado aproximadamente cada 24hs. Los pasajes de las CMEh se realizaron cuando las colonias, observadas al microscopio con un aumento de 100X, alcanzaban aproximadamente el tamaño del campo. Para ello, las células se lavaron con PBS y luego se incubaron durante aproximadamente 40 minutos con 1ml de una solución 1 mg/ml de Colagenasa Tipo IV (Gibco); al confirmar bajo el microscopio la separación parcial de las colonias de la capa nutricia se inactivó la Colagenasa por dilución, agregando 5 ml de medio HES. Luego de una centrifugación durante 5 minutos a 200 x g, se eliminó el sobrenadante, se agregó medio HES y se resuspendió suavemente de manera que las colonias se rompieran parcialmente. Este punto es de gran importancia ya que si las colonias resultantes son muy pequeñas, la cantidad de muerte celular resulta muy elevada.

6.3. Derivación de MEF

Con el objetivo de obtener los embriones de ratón, ratones hembras en día 12 o 13 de preñez de la cepa CF1 fueron obtenidas del Bioterio de la Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires. Estos ratones fueron sacrificadas por dislocación cervical. En cabina de flujo vertical se procedió a limpiar el abdomen con etanol 70% v/v, se abrió la cavidad peritoneal y se removieron los cuernos uterinos que contienen a los embriones. Los

cuernos uterinos fueron colocados en una placa de Petri y lavados 3 veces con PBS (del inglés phosphate buffered saline) (NaCl 8 mg, KCl 200 mg, KH₂PO₄ 200 mg y Na₂HPO₄ 1,14 mg, por litro de solución, pH 7,4) con el objetivo de remover la sangre. Posteriormente, se separaron los embriones de las placentas y de sus membranas fetales y una vez libres de deciduas se colocaron en otra placa de Petri, lavando nuevamente 3 veces con PBS. Bajo lupa estereoscópica los embriones fueron decapitados y eviscerados, descartando luego este excedente. El remanente se colocó en una nueva placa de Petri y se lavó 2 veces con PBS. Los cuerpos de los embriones fueron homogeneizados mecánicamente utilizando hojas de bisturí; luego se agregó Tripsina/EDTA (0,25 % v/v Tripsina 0,5mg/ml (Invitrogen) en PBS + 0,53 mM EDTA) y se continuó el tratamiento mecánico de los mismos hasta disgregar completamente el tejido. Se agregaron 5 ml de Tripsina/EDTA a la placa y se incubaron los preparados durante 30 minutos a 37°C. Luego se colocaron los embriones procesados en un tubo cónico de 50 ml y se agregaron aproximadamente 30 ml de medio DMEM (del inglés, Dulbecco's Modified Eagle's Medium) (Gibco) suplementado con aminoácidos no esenciales (2mM concentración final) (Gibco), 50 U/ml Penicilina/Estreptomicina (Gibco) 200µM. A este medio suplementado se le agregó además 10 % v/v SFB (suero fetal bovino) (Gibco). Se dejó decantar la preparación recuperando luego el sobrenadante, el cual contiene a los MEF. El volumen obtenido de sobrenadante se dividió en X botellas de cultivo T75 (X = Número de Embriones Obtenidos/3). Esta siembra de células fue considerada pasaje cero. Se mantuvieron las botellas T75 durante 2 ó 3 días en incubadora a 37°C y atmósfera al 5 % CO₂. Luego de que la monocapa de fibroblastos alcanzó un 100 % de confluencia se procedió a su amplificación o a su criopreservación.

6.4. Cultivo y criopreservación de MEF

Los MEF fueron cultivados en medio DMEM con suplementos y con 10 % v/v SFB. Para realizar los pasajes y amplificación de las células, estas disgregadas enzimáticamente con Tripsina/EDTA. Los pasajes sucesivos se realizaron cuando las células alcanzaron un 100 % de confluencia. Para evitar los procesos biológicos asociados a la senescencia replicativa las células no fueron amplificadas por más de 3 pasajes.

Para la criopreservación de los MEF, las células fueron resuspendidas en 1ml de DMEM con suplementos y con 20 % v/v SFB y 10 % v/v dimetilsulfóxido (DMSO) (Sigma Aldrich). El medio de criopreservación fue agregado a las células de forma lenta y mezclando suave-

mente. Los criotubos fueron guardados rápidamente en un contenedor para congelamiento con isopropanol por 24 horas a -80°C , permitiendo un descenso de la temperatura a razón de $1^{\circ}\text{C}/\text{minuto}$. Finalmente se los almacenó en un tanque de nitrógeno líquido.

6.5. Cultivo de CMEh libre de MEFi

Para los protocolos de diferenciación celular en adhesión y para los experimentos de transfección (tanto de plásmidos como pequeñas moléculas de ARN) fue necesario crecer las CMEh sobre placas *multiwell* pre-coateadas con una matriz comercial rica en componentes de matriz extracelular, Geltrex (ThermoFisher), evitando así el co-cultivo con MEFi. Las células fueron mantenidas de ésta manera en estado de indiferenciación cultivándolas con medio mTesR (Stem Cell Technologies) o E8 (ThermoFisher) y mantenidas en incubadora a 37°C y 5 % CO_2 . Los sucesivos pasajes fueron realizados una vez alcanzada una confluencia de 70-80%. Para ello, se disgregaron las células con la enzima *TrypLE Select 1X* (ThermoFisher) durante 3-5 min y luego diluida en PBS. Luego, las células unicelulares fueron cultivadas en mTesR o E8 suplementado con inhibidor de Rock $10\mu\text{M}$ (necesario para mantener la sobrevivencia de las células en estado unicelular).

6.6. Inducción del progenitor temprano de mesodermo

Las colonias de CMEh fueron pasadas manualmente de MEFi a *wells* pre-tratados con Geltrex (diluido 1:1000 de una concentración $15\text{mg}/\text{ml}$) y cultivadas en medio mTeSR (StemCell Technology) hasta adquirir una confluencia del 70-90 % (2 a 4 días). Una vez alcanzada la confluencia adecuada se les cambió el medio por StemPro34 (ThermoFisher) suplementado con Activina A ($10\text{ng}/\text{ml}$), bFGF ($10\text{ng}/\text{ml}$), VEGF ($10\text{ng}/\text{ml}$) y BMP4 ($10\text{ng}/\text{ml}$). Las células fueron cultivadas en este medio por 24hs (día 1), luego el medio fue cambiado a StemPro34 suplementado con bFGF ($10\text{ng}/\text{ml}$), VEGF ($10\text{ng}/\text{ml}$) y BMP4 ($10\text{ng}/\text{ml}$) y se las cultivo en este medio hasta el fin del protocolo (todos los factores de crecimiento utilizados son de ThermoFisher). En el día 3,5 se analizó mediante citometría de flujo la aparición del progenitor temprano de mesodermo.

6.7. Diferenciación cardíaca mediante la formación de EB

Las colonias de CMEh crecidas sobre MEFi fueron disgregadas utilizando la enzima TrypLE Select 1X (ThermoFisher) durante 5min hasta lograr una suspensión unicelular. Luego fueron diluidas en PBS y centrifugadas por 5 minutos a 1000 rpm. Luego se las resuspendió en medio mTeSR suplementado con inhibidor de Rock 10 μ M (Cayman Chemicals) y BMP-4 10ng/ml (ThermoFisher). Se sembraron las células en placas no adherentes de 96 *wells* con fondo en V (25.000 células/well, contadas en cámara de Neubauer), se centrifugaron por 7min a 1300 rpm y se incubaron O.N a 37°C y 5 % de CO₂. Al día siguiente, se retiraron manualmente los EBs formados y se cultivaron en suspensión por 4 días en placas de petri no adherentes en medio StemPro34 suplementado con 10ng/ml BMP4, 5ng/ml bFGF y 3ng/ml Activina A. El día 4 de la diferenciación, los EBs fueron adheridos en placas de 12, 24 o 6 *wells*, dependiendo del experimento, pre-tratadas con gelatina 0,1 % bovina (0,1 % v/v en PBS) (Sigma) en medio StemPro34 suplementado con 10ng/ml BMP4, 5ng/ml bFGF , 3ng/ml Activina A y 5 μ M IWR-1 (Sigma). A partir del día 5 y hasta el día 14 las células fueron mantenidas en medio StremPro34 suplementado con 5ng/ml VEGF, 10ng/ml bFGF y 5 μ M IWR-1.

6.8. Diferenciación cardíaca dirigida en monocapa

Las células CMPi fueron crecidas sobre *wells* pre-tratados con Geltrex (diluido 1:1000 de una concentración 15mg/ml), y cultivadas con mTesR. Una vez alcanzado el 100% de confluencia se cultivaron las células con el medio RPMI (Thermo Fisher) suplementado con B27 sin insulina (B27 (-)) (Thermo Fisher) y con el inhibidor de GSK3 α , CHIR99021, en una concentración 10 μ M durante 24hs (al inicio del protocolo es importante que no haya ningún espacio del *well* sin estar cubierto de células para alcanzar una buena eficiencia en la diferenciación). Luego de 24hs se retiró del medio el inhibidor y se cultivaron las células con el medio RPMI/B27 (-). En el tercer día se agregó al medio el inhibidor de Wnt, IWR-1, en una concentración 5 μ M. En el día 5 de iniciada la diferenciación se retiró del medio el inhibidor IWR-1 y se cultivaron las células en RPMI/B27 (-) durante 4 días (cambiando el medio día por medio). En el día 9 aproximadamente comenzaron a aparecer regiones con actividad contráctil espontánea y el medio fue cambiado a RPMI suplementado con B27 con insulina (B27(+)). Las células fueron mantenidas de esta manera hasta que casi toda la monocapa de células presentó regiones con actividad contráctil (día 11-12). Una vez esto, se cultivaron las células

en un medio específico para enriquecer el cultivo en células cardíacas, durante 4-5 días. Este medio está compuesto por RPMI sin glucosa, BSA 500 μ g/ml, *L-ascorbic acid 2-phosphate sesquimagnesium* 213 μ g/ml y DL-lactato de sodio 7mM. Una vez enriquecido el cultivo en células con actividad contráctil, se volvieron a cultivar en RPMI/B27(+) durante 2-3 días a modo de recuperación. Finalmente, para tener un cultivo homogéneamente distribuido en los *wells* las células fueron disgregadas y vuelto a plaquear utilizando TrypLE Select 10X durante 4-10min y agitando la placa suavemente. La enzima fue inactivada con un medio de "replaqueo" (RPMI + B27 (+) + SFB 20%) y cultivadas sobre *wells* pre-tratados con Geltrex con el medio de "replaqueo" durante 24hs y luego RPMI/B27+. Aproximadamente en el día 22 de iniciada la diferenciación se realizó la transfección con el *mimic* del miR-520a o el *mimic* control.

6.9. Inmunomarcaciones y microscopía de fluorescencia

Para las inmunomarcaciones las células fueron cultivadas en placas de 24 *wells* (Corning). Las células fueron lavadas con PBS y fijadas con una solución 4% m/v de paraformaldehído (PFA) (Sigma) en PBS durante 20 minutos a temperatura ambiente. A continuación, se realizaron 3 lavados con albúmina sérica bovina (ASB) 0,1 % v/v en PBS (Gibco). La permeabilización y el bloqueo se realizó con una solución compuesta por 0,1 % v/v Tritón X-100 (Sigma); 0,1 % v/v ASB; 10% v/v Suero de Cabra Normal (SCN) (Gibco) en PBS, durante 30 minutos a temperatura ambiente. Luego se realizaron 3 lavados durante 5 minutos cada uno, con 0,1 % v/v ASB en PBS. Las inmunomarcaciones se realizaron incubando las células con los anticuerpos primarios específicos en una solución de dilución de anticuerpos compuesta por: 0,1 % v/v ASB, 10% v/v SCN en PBS durante 2 horas a temperatura ambiente en agitador mecánico o durante 16 hs a 4°C. Luego se realizaron 3 lavados de 5 minutos cada uno con PBS, 0,1 % v/v ASB. Posteriormente, se realizó una incubación de 2 horas en agitación de las células con el o los anticuerpos secundarios conjugados a los fluoróforos, en una dilución 1:400 (en la misma solución de dilución empleada para los anticuerpos primarios). Se realizaron nuevamente 3 lavados durante 5 minutos cada uno, con 0,1 % v/v ASB en PBS y finalmente se agregó un volumen apropiado de PBS. Los anticuerpos primarios y secundarios utilizados en este trabajo se detallan en la tabla 7. El análisis de las imágenes se realizó fotografiando las células con una cámara digital Nikon DXN1200F acoplada al microscopio invertido de fluorescencia con lám-

para de mercurio (Nikon Eclipse TE2000-S) y controlada por el software EclipseNet versión 1.20.0 Build 61.

Anticuerpo primario	Especie-tipo	Marca	Número de cat.
α -Nkx2.5	Conejo-policlonal	<i>Santa Cruz</i>	<i>sc-14033</i>
α -MHCA (MF20)	Ratón-monoclonal	<i>DSHB</i>	<i>AB_2147781</i>
α -cTnT	Ratón-monoclonal	<i>Abcam</i>	<i>ab8295</i>
Anticuerpo secundario	Especie-Fluoróforo	Marca	Número de cat.
α -IgG de ratón	Pollo-policlonal- Alexa Fluor 488	<i>Invitrogen</i>	<i>A-21200</i>
α -IgG de conejo	Cabra-policlonal- Alexa Fluor 555	<i>Invitrogen</i>	<i>A-21429</i>

Tabla 7: Anticuerpos primarios y secundarios utilizados en inmunofluorescencia

6.10. Citometría de Flujo

Para el análisis de la expresión de Oct-4, las células fueron disgregadas utilizando una solución de tripsina-EDTA 0,25 % v/v, transferidas a tubos de hemólisis y centrifugadas a 300 x g durante 5 minutos. Luego fijadas en 300 μ l de PFA 4% e incubadas durante 15 a 20 min. Luego de 2 lavados con PBS se permeabilizaron utilizando la solución PermBufferIII (BD) durante 20 min a 4 °C. La inmunomarcación fue realizada agregando el anticuerpo anti-Oct-4 (Santa Cruz, SC-5279) utilizado en una dilución 1/400 e incubando las muestras 30-40min a 4°C. Las células luego fueron lavadas e incubadas con el anticuerpo secundario Alexa fluor 488 anti-ratón IgG (H+L) (Life Technologies). Como control de especificidad se incubaron células con el anticuerpo secundario pero no marcadas con el anticuerpo primario. Todas las muestras fueron incubadas a 4°C durante 20 min en oscuridad. Luego se realizaron dos lavados agregando 1ml de PBS en cada tubo, centrifugando a 200 x g, aspirando completamente el sobrenadante y resuspendiendo las células en 500 μ l de PBS.

Para analizar por citometría la diferenciación hacia el MPC, las células fueron despegadas de la superficie en las que se las cultivó utilizando TrypLe Select 1X durante 3-4min. Una vez desprendidas completamente, se diluyó la enzima con PBS y se centrifugó 5min a 200xg. Luego, se resuspendió el *pellet* de células en una dilución de 1:50 del anticuerpo específico anti-EpCAM(CD326) (clon VU-1D9, Life Technologies) y anti-NCAM(CD56) (clon B159, BD), conjugados con PE y BB515 respectivamente en 100 μ l de PBS con 0,1 % v/v ASB y se incu-

bó en hielo durante 20 min. Luego se realizaron dos lavados con PBS/ABS 0,01 % v/v y se resuspendió las células en PBS para ser analizadas.

Las mediciones se realizaron en el citómetro Accuri c6 de BD (Becton Dickinson) y los resultados fueron analizados con el software BDaccuri.

6.11. Extracción de ADN genómico

Para la extracción de ADN genómico, a partir de 2×10^5 células disociadas se realizó un lavado con PBS y se las resuspendió en 100 μ l de *buffer* de lisis (10 mM Tris-HCl pH 8.3, 50 mM KCl, 2 mM MgCl₂, 0.001 % gelatina, 0.5 % NP-40, 0.5 % Tween-20) y se agregó 1 μ l de proteinasa K (10 μ g/ml, Thermo Fisher). Luego de incubar la muestra a 65°C durante 1 hora en agitación se precipitó en frío con 600 μ l de isopropanol 100 % durante 30 min a -20°C. Se centrifugó la muestra a 12.000xg durante 15 min a 4°C, se lavó el *pellet* de ADN con etanol 70 % y se lo resuspendió en 30 μ l de agua milliQ. Se favoreció la solubilización del DNA genómico mediante la incubación de las muestras a 65°C durante 1 h. Conservamos las muestras de DNA genómico a -20°C.

Alternativamente, para el *screening* inicial de los clones de CRISPR/Cas9, se realizó la extracción de ADN genómico sin los pasos de precipitación y lavado. Básicamente, luego del incubación de 1 hr con la Proteinasa K, se incubó la muestra 15 min a 95°C y para la reacción de PCR se utilizó 1 μ l de esa solución.

6.12. Extracción de ARN total

La extracción del ARN se llevó a cabo utilizando el reactivo Trizol (Thermo Fisher) de acuerdo a las indicaciones del fabricante. Brevemente, se retiró el medio de cultivo a las células, se lavó con PBS y se colectaron en un volumen óptimo de Trizol según el número de células de las cuales se partió. Luego se realizó la extracción orgánica con cloroformo (Merck), se precipitó el ARN con isopropanol (Merck), se lavó con etanol 75 % y se resuspendió en agua deionizada libre de ribonucleasas. En el caso de tener una muestra con poco número de células, se agregó, previo a la precipitación con isopropanol, en cada tubo Glyco Blue 0,5 μ l (15mg/ml, Thermo Fisher), el cual funciona como *carrier* favoreciendo la precipitación del ARN. Al final se puso la solución de ARN en un termobloque a 55°C durante 10 min para favorecer la resuspensión. La concentración del ARN aislado (ng de ARN/ μ l de solución) se determinó utilizando un espec-

trofotómetro automático NanoDrop 1000. Se verificó también que la relación OD260/OD280 fuese de entre 1,7 y 2 para asegurar una elevada pureza del ARN total obtenido. Este ARN fue conservado a a -80°C hasta el momento de uso.

6.13. Extracción de pequeños ARN para secuenciación

Para el análisis y estudio de los microARNs por secuenciación masiva (*Small RNA-Seq*) se realizó la extracción de los pequeños ARN utilizando el kit *miRvana* de Ambion (*Life Technologies*), siguiendo las instrucciones del fabricante. Brevemente, las células en suspensión fueron lisadas utilizando el *buffer* de lisis/unión y mediante acción mecánica con *vortex* o pipeta hasta lograr un aspecto homogéneo en el lisado. Luego, se realizó la extracción orgánica del ARN mediante ácido fenol-cloroformo y centrifugación. Seguidamente, se agregó etanol 100 % (1/3 del volumen recuperado) a la fase acuosa y se colocó en una columna que retiene las moléculas de ARN largas, pero deja pasar los pequeños ARNs. Se centrifugo, y se volvió a agregar etanol 100 % (2/3 del volumen) al eluido y se volvió a pasar por una nueva columna. Esta columna se lavó por centrifugación agregando las soluciones 1 y 2/3. Finalmente se eluyó el ARN con 50-100 μ l de agua MiliQ (pre-calentada a 95°C) y libre de ARNasas. La concentración del ARN aislado (ng de ARN/ μ l de solución) se determinó utilizando un espectrofotómetro automático NanoDrop 1000 y se conservó a -80°C.

6.14. Retrotranscripción de ARNm y microARNs maduros

La síntesis del ADNc a partir del ARN se llevó a cabo según el siguiente protocolo para retrotranscribir tanto los ARNm como los microARNs. En una primera etapa se preparó la siguiente mezcla por cada tubo: 0,25 μ g de oligo dT, entre 500ng - 1 μ g de ARN (tratado o no con ADNasa, según cada caso), dNTPs a una concentración final de 0,4 mM cada uno y el SLO *Stem Loop Primer* específico para cada microARN o una mezcla de éstos en una concentración final de 1 μ M. El volumen de la mezcla se completó con agua Milli-Q hasta 13,5 μ l. La mezcla de reacción se incubó durante 5min a 65 °C y a continuación en hielo al menos 1 min. Luego se adicionó a cada tubo de reacción 4 μ l de 5X *First Strand Buffer*, 1 μ l de 0,1 M DTT, 1 μ l *RNase OUT Recombinant Ribonuclease Inhibitor* (Invitrogen) y 100 U de *SuperScript III Reverse Transcriptase*. Se mezcló con micropipeta y se incubó primero 30 min a 16 °C. Luego 30 min a 42 °C y una hora a 50°C. Por último se inactivó la reacción mediante la incubación a 70 °C

durante 15 min y finalmente el ADNc a usar como molde en la reacciones de amplificación se diluyó al menos 10 veces en agua Milli-Q.

6.15. PCR a tiempo final

La reacción de PCR para el *screening* de los clones de CRISPR/Cas9 se realizó en un volumen final de 25 μ l y se colocaron los siguientes reactivos por tubo de PCR: 2,5 μ l de Buffer de PCR (Tris-HCl 200mM (pH 8.4), KCl 500mM), 1 μ l de MgCl₂ 50mM, 0,5 μ l de una mezcla de los 4 deoxi-nucleótidos trifosfato (dNTPs) 10mM (dATP, dGTP, dCTP y dTTP) (Thermo Fisher), 0,6 μ l de oligonucleótido sentido (primer forward) 10 μ M, 0,6 μ l de oligonucleótido anti-sentido (primer reverse) 10 μ M, 0,2 μ l de la ADN polimerasa TransTaq HiFi 5U/ μ l (Trans Bionovo) y 2 μ l de ADN genómico, y completando el volumen a 25 μ l con agua libre de DNasas. El programa de PCR utilizado fue el siguiente:

1. Desnaturalización inicial y activación de la polimerasa: 5 minutos a 94°C.
2. Desnaturalización del ADN molde: 30 segundos a 94 °C
3. Apareamiento: 30 segundos a 65 °C
4. Extensión: 40 segundos a 72 °C
5. Extensión Final: 5 minutos a 72 °C

Los pasos 2 a 4 se realizaron 35 ciclos

Para secuenciar el producto de PCR, se corrió el producto de la reacción en un gel de agarosa (con la concentración adecuada según el tamaño del producto) y luego se cortó la banda del tamaño correspondiente y se purificó utilizando el kit "*PureLink*"*Quick Gel Extraction Kit* (Invitrogen) siguiendo las especificación del fabricante. Brevemente, se pesó el fragmento de agarosa con el ADN para luego agregar en una relación 3:1 el volumen del buffer L3 que permite solubilizar la agarosa (mg de agarosa : ml de buffer). La mezcla se calentó a 55°C durante 10 min o hasta que el gel este completamente disuelto. A continuación se incubó 5 min más a la misma temperatura y se agregó un volumen de isopropanol equivalente al peso inicial de agarosa. Luego se purificó el ADN utilizando una columna con una membrana de silica. Se cargó la muestra en la columna y se centrifugó 1 min a 12.000xg. Se descartó el eluido, se lavó con 500 μ l de Wash buffer y se volvió a centrifugar 1 min a 12.000xg. Luego se descartó una

vez más el eluido y se centrifugó 2-3 min a la máxima velocidad para eliminar todo el volumen remanente. Por último, se eluyó el ADN con 30-40 μ l de agua milliQ previamente calentada a 50°C centrifugando 1 min a 12.000xg. Para maximizar la elución, se repitió este último paso una vez más.

6.16. PCR cuantitativa

Los pares de oligonucleótidos específicos (*primers*) utilizados en este trabajo para analizar la expresión de los genes estudiados fueron diseñados en el portal BLAST de NCBI con la aplicación *primer-BLAST*. Para evitar la amplificación espuria de posible ADN geonómico contaminante en las muestras, los *primers* fueron diseñados de manera tal que cada miembro del par sentido y anti-sentido hibride en un exón diferente del gen en estudio. En los casos en los que esto no fue posible, luego de la extracción de ARN se incluyó un paso de digestión de ADN, incubando las muestras con ADNasa 1 durante 30 minutos a 37°C. Los *primers* fueron sintetizados por el servicio de la empresa Invitrogen o Genbiotech. Las secuencias de los oligonucleótidos utilizados se indican en la tabla X. Las reacciones se realizaron en un volumen final de 20 μ l. Se colocaron los siguientes reactivos por tubo de PCR: 10 μ l de SYBR-Green master mix (ThermoFisher); 0,4 μ l de ROX (ThermoFisher); 3,6 μ l de primer sentido (5 μ M); 3,6 μ l de primer anti-sentido (5 μ l); 0,4 μ l de H₂O y 2 μ l de ADNc. Se utilizó el termociclador Step One Plus (Applied Biosystems). Cada muestra a analizar fue determinada por duplicado. En todos los casos el programa utilizado fue el siguiente:

Desnaturalización inicial y activación de la polimerasa:

1. 2 min a 50°C.
2. 10 minutos a 94°C.
3. Apareamiento y extensión: 30 segundos a 60 °C.
4. Desnaturalización: 40 segundos a 94 °C.
5. 40 segundos a 58 °C.
6. 40 segundos a 58 °C
7. 40 segundos a 72 °C

Se realizaron 40 ciclos de los pasos 2-5. La lectura de la fluorescencia fue realizada al finalizar el paso 5 de cada ciclo. Para cada uno de los ADNc analizados se realizaron curvas de temperatura de apareamiento para verificar la presencia de un único producto de amplificación. Esto fue confirmado al sembrar los productos de la reacción en geles de agarosa (2 %) teñidos con bromuro de etidio, evidenciando una única banda.

El análisis de los resultados se realizó utilizando en la mayoría de los casos el programa de uso libre LinReg (Ruijter et al., 2009), a partir del cual se obtuvieron las eficiencias de la reacción de PCR para cada par de *primers* y luego se calcularon los niveles teóricos iniciales de cDNA (N0) para cada condición. El análisis de la qPCR de los genes en la diferenciación cardíaca (figura 5) se realizó mediante el método delta delta Ct. Se evaluó un gen *housekeeping* el RPL7, el cual fue utilizado para normalizar los niveles de expresión de los genes estudiados y el RNU6B, el cual fue utilizado como normalizador para los niveles de los microARNs. Todos los experimentos fueron realizados con al menos tres réplicas biológicas.

Nombre del gen	Primer Forward	Primer Reverse
RPL7	AATGGCGAGGATGGCAAG	TGACGAAGGCGAAGAAGC
NANOG	AAAGAATCTTCACCTATGCC	GAAGGAAGAGGAGAGACAGT
OCT4	CTGGGTTGATCCTCGGACCT	CACAGAACTCATACGGCGGG
T (Brachyury)	TCCCAGGTGGCTTACAGATGA	GGTGTGCCAAAGTTGCCAAT
NKX2.5	CCCACGCCCTTCTCAGTCAA	GTAGGCCTCTGGCTTGAAGG
CTNT	ATGATGCATTTTGGGGGTTA	CAGCACCTTCCTCCTCTCAG
MYH6	CAACAATCCCTACGACTAC	TCTCCTCTGAAGTGAAGC
KDR	TATCCTCTCCTACCGGCACCCGC	GAGGCGCGGAGGTGGAACG
MESP1	GCCACTTCACACCTCGGGCTC	CCAGGCCGCAGAGAGCATCCA
MIXL1	GGTACCCCGACATCCACTT	TGGAAGGATTTCCTACTCTG
SNAI2	TCAAGGACACATAGAACTCAC	CTACACAGCAGCCAGATTC
SNAI1	ATGCACATCCGAAGCCACA	GAGGGTCAGCGGGGACATC
ZEB1	TTACACCTTTGCATACAGAACCC	TTACGATTACACCCAGACTGC
ZEB2	GGAGACGAGTCCAGCTAGTGT	CCACTCCACCTCCCTTATTTC
ECAD	AAAGGCCCATTTCTTAAAAACCT	TGCGTTCTCTATCCAGAGGCT
VIM	TGCCGTTGAAGCTGCTAACTA	CCAGAGGGAGTGAATCCAGATTA

Tabla 8: Secuencias de los primers utilizados en la PCR cuantitativa en sentido 5'-3'

6.17. Diseño de *primers* para analizar los microARNs maduros

Para la retrotranscripción de los microARNs fue necesario diseñar primers específicos denominados *Stem Loop Primers* (SLO), los cuales consisten en una región de constante de 44 nucleótidos (*GTTGGCTCTGGTGCAGGGTCCGAGGTATTTCGCACCAGAGCCAAC*) y 6 nucleótidos en el extremo 3' que son específicos para cada microARN a estudiar. Luego, para la PCR cuantitativa se diseñaron *primers forward* específicos para cada microARN el cual consiste en 14 nucleótidos complementarios al extremo 5' del microARN y una región de 7 nucleótidos que funciona como adaptador (*GCGGCCGG*). La región adaptadora permite balancear el contenido GC con el *primer reverse* y así sus temperaturas de *annealing*. Este *primer reverse* es "universal" y se utiliza el mismo para todas las reacciones de qPCR de los microARN (reconoce una región del SLO).

miRNA	SLO primer	Primer Fw (5'-3')
RNU6B	GTCGTATCCAGTGCAGGGTCCGAGGTATTTCGCACTGGATACGACAAAATATGGAAC	CGCAAGGATGACACGCAAATTC
mir-302a	GTCTCCTCTGGTGCAGGGTCCGAGGTATTTCGCACCAGAGGAGACTCACCA	GCGGCCGTAAGTGCTTCCATG
mir-302b	GTCTCCTCTGGTGCAGGGTCCGAGGTATTTCGCACCAGAGGAGACCTACTA	
mir-302c	GTCTCCTCTGGTGCAGGGTCCGAGGTATTTCGCACCAGAGGAGACCCACTG	
mir-302d	GTCTCCTCTGGTGCAGGGTCCGAGGTATTTCGCACCAGAGGAGACACACTC	
mir-200a	GTCGTATCCAGTGCAGGGTCCGAGGTATTTCGCACTGGATACGACACATCG	GCGGCCGTAACACTGTCTGGTAAC
mir-200b	GTCGTATCCAGTGCAGGGTCCGAGGTATTTCGCACTGGATACGACTCATCA	GCGGCCGTAATACTGCCTGG
mir-200c	GTCGTATCCAGTGCAGGGTCCGAGGTATTTCGCACTGGATACGACTCCATC	GCGGCCGTAATACTGCCTGG
mir-205	GTCGTATCCAGTGCAGGGTCCGAGGTATTTCGCACCAGAGGAGACACT	GCGGCCGTCCTTCATCCACCG
mir-145	GTCTCCTCTGGTGCAGGGTCCGAGGTATTTCGCACCAGAGGAGACAGGGAT	GCGGCCGGTCCAGTTTCCAG
mir-296	GTCTCCTCTGGTGCAGGGTCCGAGGTATTTCGCACCAGAGGAGACACAGGA	GCGGCCGAGGGCCCCCTCAATC
mir-1	GTCTCCTCTGGTGCAGGGTCCGAGGTATTTCGCACCAGAGGAGACATACAT	GCGGCCGTGGAATGTAAAGAAG
mir-133a	GTCTCCTCTGGTGCAGGGTCCGAGGTATTTCGCACCAGAGGAGACAGCTG	GCGGCCGTTTGGTCCCCTTCAAC
mir-133b	GTCTCCTCTGGTGCAGGGTCCGAGGTATTTCGCACCAGAGGAGACTAGCTG	GCGGCCGTTTGGTCCCCTTCAAC
mir-206	GTCTCCTCTGGTGCAGGGTCCGAGGTATTTCGCACCAGAGGAGACCCACAC	GCGGCCGTGGAATGTAAAGAAG
mir-208	GTCTCCTCTGGTGCAGGGTCCGAGGTATTTCGCACCAGAGGAGACACAAGC	GCGGCCGATAAGACGAGCAAAAAG
mir-512-3p	GTCGTATCCAGTGCAGGGTCCGAGGTATTTCGCACTGGATACGACACCTC	GCGGCCGAAGTGCTGTCATAGC
mir-516b	GTCGTATCCAGTGCAGGGTCCGAGGTATTTCGCACTGGATACGACAAAGTG	GCGGCCGATCTGGAGGTAAGAAG
mir-517a,b	GTCGTATCCAGTGCAGGGTCCGAGGTATTTCGCACTGGATACGACACACTC	GCGGCCGATCGTGATCCCTTTAG
mir-517c	GTCGTATCCAGTGCAGGGTCCGAGGTATTTCGCACTGGATACGACACACTC	GCGGCCGATCGTGATCCTTTTAG
mir-519d	GTCGTATCCAGTGCAGGGTCCGAGGTATTTCGCACTGGATACGACACTCT	GCGGCCGCAAGTGCTCCCTCC
mir-520a	GTCGTATCCAGTGCAGGGTCCGAGGTATTTCGCACTGGATACGACACAGTC	GCGGCCGAAAGTGCTTCCTTTG
mir-520f	GTCGTATCCAGTGCAGGGTCCGAGGTATTTCGCACTGGATACGACAACCT	GCGGCCGAAGTGCTTCCTTTAG
mir-520g	GTCGTATCCAGTGCAGGGTCCGAGGTATTTCGCACTGGATACGACACACTC	GCGGCCGACAAAGTGCTTCCC
mir-526	GTCGTATCCAGTGCAGGGTCCGAGGTATTTCGCACTGGATACGACACAGAA	GCGGCCGCTCTTGAGGAAGCAC
mir-Reverse	ATCCAGTGCAGGGTCCGAGG	
RNU6B Rev	CCAGTGCAGGGTCCGAGGT	

Tabla 9: Secuencias de los primers utilizados para los microARNs en la RT-qPCR.

6.18. Diseño y clonado de los gRNA para CRISPR/Cas9

En este trabajo diseñamos dos sgRNAs para eliminar el *cluster* C19MC, utilizando la herramienta bioinformática *Benchling*. El gRNA START dirige la enzima Cas9 hacia un sitio río arriba del *cluster* y el gRNA END hacia el final del mismo. En la tabla 10 se detallan las secuencias de dichos sgRNAs.

sgRNA	Secuencia (5'-3')
sgRNA START top	CACCGCAAATCCTAGGCCTGCCCTG
sgRNA START bottom	AAACCAGGGCAGGCCTAGGATTGTC
sgRNA END top	CACCGGTAGACAAGATGAATCTCAG
sgRNA END bottom	AAACCTGAGATTCATCTTGTCTACC

Tabla 10: Secuencia de los gRNA diseñados para la edición génica del *cluster* C19MC

Los oligonucleótidos top y bottom fueron resuspendidos en agua MiliQ en una concentración final 100 μ M, hibridizados *in vitro* y clonados en el plásmido pSpCas9(BB) tal como se explica en el trabajo de (Ran et al., 2013). Se realizó un ensayo de digestión y ligación simultánea con la enzima de restricción Bbs1 y la enzima ligasa T7, respectivamente. Luego de la transformación de bacterias competentes y selección con ampicilina, seleccionamos colonias transformantes y las validamos por PCR con *primers* específicos y luego por secuenciación.

6.19. Preparación de plásmidos - Transformación de bacterias competentes

Bacterias competentes *Escherichia coli* de la cepa DH5 α , fueron transformadas con 100ng de cada plásmido clonado con los gRNA de CRISPR, mediante *shock* térmico a 42°C durante 90 segundos. A continuación se incubó en medio LB líquido (para 1L de medio: 10g de peptona, 5g de extracto de levadura y 10g de NaCl), sin antibiótico, a 37°C durante 30-45 minutos en agitación permanente. Las bacterias se plaquearon en placas de Petri, de agar-LB (LB + 15 g/L de agar) con ampicilina 100 μ g/ml, y se dejaron en estufa a 37°C durante 16 horas. Posteriormente se picaron algunas colonias y se realizaron mini cultivos en LB-Ampicilina a 37°C durante 16 horas, en agitación permanente. A partir de estos se preservó una alícuota de bacterias transformadas con el plásmido en glicerol 30%.

En cada ensayo de transformación, se realizaron los siguientes tres controles. Bacterias sin transformar, para evaluar la sensibilidad al antibiótico, bacterias con el plásmido sin clonar, para evaluar la viabilidad de las células y bacterias con el plásmido sin ligar, como control del plásmido religado sin inserto.

6.20. Preparación de plásmidos a gran escala (Maxi y midi-prep)

Los mini cultivos obtenidos fueron crecidos en 500ml de LB-Ampicilina a 37°C durante 16 horas, en agitación permanente. Se centrifugaron las bacterias 20 minutos a 3200xg y se descartó el sobrenadante. La preparación del plásmido se realizó utilizando el kit Wizard Plus Maxiprep DNA purification (Promega), siguiendo las indicaciones del fabricante. Brevemente, el *pellet* de bacterias se resuspendió en *Cell Resuspension Solution* (50mM Tris-HCl (pH 7.5), 10mM EDTA, 100µg/ml RNase A), hasta que se obtuvo un lisado homogéneo. A esto se le agregó *Cell Lysis Solution* (0.2M NaOH, 1 % SDS) y se dejó reposar durante 10min. Luego se agregó la solución de neutralización (1.32M acetato de potasio (pH 4.8)), se mezcló por inversión y se centrifugó a 14000xg durante 15min. A continuación, el sobrenadante fue transferido a un falcon filtrándolo con una gasa para retener la mayor cantidad de restos celulares. Se agregó 0,5 volúmenes de isopropanol, se mezcló por inversión y se centrifugó a 14000xg durante 15min. El *pellet* obtenido se resuspendió en 1ml de agua MiliQ o *buffer* TE (10mM Tris-HCl (pH 7.5), 1mM EDTA). A continuación, se realizó la purificación del plásmido mediante una columna con resina. Para esto, la solución con el ADN fue mezclada con 10 ml de resina y luego transferida a una maxicolumna. Esta mezcla fue empaquetada en el fondo de la columna mediante vacío. Luego, se lavó la columna con *Wash Solution* (80mM acetato de potasio, 8.3mM Tris-HCl (pH 7.5), 40µM EDTA y 55 % Etanol) y luego con etanol 80 %. Una vez eluido las soluciones de lavado por centrifugación, se dejó secar la columna y se eluyó el ADN con 1ml de agua MiliQ sin ADNasas (pre-calentada a 65-70°C). La concentración de ADN obtenida e determinó utilizando un espectrofotómetro automático NanoDrop 1000 y el ADN plasmídico fue conservado a -20°C.

6.21. Transfección con pequeñas moléculas comerciales de ARN

Para modular la expresión de los microARNs, las células CMEh y CMPih fueron transfectadas con pequeñas moléculas de ARN sintetizadas comercialmente, conocidas como *mimics*

e *inhibitors* ambas de Ambion (Life Technologies) o LNA (del inglés, *Locked Nucleic Acids*) comercializados por Exiqon (ver los oligos utilizados y sus respectivos códigos en la tabla 11). Para esto, las células fueron crecidas sobre Geltrex con mTesR hasta llegar a una confluencia del 60-70 %. Una vez alcanzada la confluencia deseada, las células fueron transfectadas con una mezcla de ARN con lipofectamina (*Lipofectamine RNAiMAX Reagent*, ThermoFisher) según las indicaciones del fabricante. Brevemente, se preparó una dilución con el ARN en Opti-MEM (ThermoFisher) y una dilución de igual volumen de lipofectamina en Opti-MEM. La concentración y cantidad de lipofectamina utilizada varió según el experimento, por ejemplo para una concentración final de 10nM del mimic/inhibitor, se utilizaron 3 μ l de lipofectamina. Ambas diluciones se mezclaron, agregando la dilución con ARN gota a gota en la que contenía la lipofectamina y se incubó entre 15-20 min. Luego, ésta mezcla fue agregada gota a gota al *well* con las células crecidas en adhesión.

Como control de la transfección se transfectaron en paralelo células con una molécula de pre-microARN acoplada al fluorocromo Cy3 (*Cy3TM Dye-Labeled Pre-miR Negative Control*, ThermoFisher) que es de fácil detección mediante microscopio de fluorescencia y permite estimar la eficiencia de la transfección.

Nombre del oligo sintético	Empresa (código)
mirVana miRNA mimic hsa-miR-205-5p	Ambion, Thermo Fisher (4464066)
mirVana miRNA mimic hsa-miR-520a-3p	Ambion, Thermo Fisher (MC10391)
mirVana miRNA mimic, negative control	Ambion, Thermo Fisher (4464058)
mirVana miRNA inhibitor hsa-mir-205-5p	Ambion, Thermo Fisher (AM11015)
mirVana miRNA inhibitor, negative control	Ambion, Thermo Fisher (4464058)
miRCURY LNA Power family inhibitor	EXIQION (460046)
miRCURY LNA Inhibitor control	EXIQION (199006-001)
Cy3-labeled Pre-miR Negative control	Ambion, Thermo Fisher (AM17120)

Tabla 11: Nombre y número de catálogo de los oligos sintéticas de los microARNs

6.22. Transfección de las CMEh con los gRNA para CRISPR/Cas9

Los plásmidos clonados con los sgRNA para CRISPR/Cas9 fueron transfectados en las células CMPi wt para generar el *knock-out* del *cluster* C19MC. Para esto, se plaquearon 2×10^5 células sobre un *well* de *multiwellX12* (mwx12) y se las cultivó en medio mTesR suplementado con inhibidor de Rock $10 \mu\text{M}$. Luego, a las 24hs fueron transfectadas con $2 \mu\text{g}$ de una mezcla con los dos plásmidos (gRNA START y gRNA END) y *Lipofectamine Stem Transfection Reagent* (Thermo Fisher) en una proporción 2:3 (μg de plásmido:lipofectamina), siguiendo las indicaciones del fabricante. A las 24hs post-transfección se seleccionaron las células transfectadas con un pulso de 48hs de puromicina (750ng/ml). En paralelo, transfectamos un plásmido con GFP para controlar la eficiencia de transfección, la cual determinamos por citometría de flujo a las 24hs. Luego de la selección con antibiótico, cultivamos las células durante 3-4 días hasta alcanzar una confluencia del 70-80%. Una vez alcanzada la confluencia, cultivamos las células de manera muy diluida con el fin de que cada colonia provenga de una única célula y cuando alcanzaron un tamaño considerable las picamos manualmente y crecimos en placas de *multiwellX48* (1 colonia por *well*). Luego, realizamos el *screening* de cada clon por PCR. Una vez identificado el clon positivo, confirmamos la delección del *cluster* C19MC mediante la secuenciación del producto de PCR.

6.23. Ensayo de incorporación de bromodeoxiuridina

Para este ensayo se utilizó el kit APC BrdU flow Kit (BD Pharmigen). Las CMPi crecidas sobre geltrex recibieron un pulso de $10 \mu\text{M}$ de bromodeoxiuridina (BrdU) durante 40min. Luego de la incubación, las células fueron disgregadas utilizando TrypLe 1X durante 5min. Luego cada muestra fue pasada a un tubo de hemólisis, centrifugada (por 5 minutos a 1000 rpm) y resuspendidas en $100 \mu\text{l}$ de BD *Cytofix/Cytoperm Buffer*. Se incubaron las muestras en hielo durante 20 minutos y se lavaron con 1 ml de 1X BD *Perm/Wash Buffer* mediante centrifugación. A continuación se resuspendieron las células en $100 \mu\text{l}$ de BD *Cytoperm Permeabilization Buffer* y se las incubó nuevamente en hielo por 10min, se las lavó y re-fijo en *buffer* de fijación. Luego de un nuevo lavado, se trató las células con la enzima DNAasa ($300 \mu\text{g/ml}$) para exponer la BrdU incorporada en el ADN y se las incubó durante 1 hora a 37°C . Luego, las células fueron lavadas por centrifugación e incubadas durante 20min a temperatura ambiente con los anticuerpos primarios específicos anti-BrdU conjugado con el fluorocromo aloficocianina

(APC, por su nombre en inglés). Una vez finalizada la tinción con el anticuerpo se realizaron dos lavados por centrifugación utilizando el buffer 1X BD *Perm/Wash* y se marcó el contenido de ADN con 20 μ l de 7-AAD durante 15min. Por último, las células fueron analizadas mediante citometría de flujo en el citómetro Accuri c6 de BD y los resultados analizados con el software *BDaccuri*.

6.24. Ensayo de apoptosis con Anexina V

Para el análisis de la presencia de Anexina V en las células apoptóticas se utilizó el kit FITC-Annexin V Apoptosis Detection Kit I de BD Bioscience. Brevemente, se recuperó el medio de cultivo en donde eran crecidas las células para recuperar las células apoptóticas que pudieran estar en suspensión y se despegaron las células crecidas en adhesión utilizando TrypLe Select 1X. A continuación, se inactivó la enzima por dilución con PBS, se juntó con el sobrenadante recuperado, y se centrifugó 5 min a 200xg. Luego las células fueron resuspendidas en 1ml de *Binding buffer* 1X y se incubaron 100 μ l de esta suspensión (1×10^5 células) con 5 μ l de IP (ioduro de propidio) y 5 μ l de FITC Anexina V, durante 15min a temperatura ambiente y oscuridad. Luego, se agregaron 400 μ l del *Binding buffer* 1X y se analizaron las muestras mediante citometría de flujo.

6.25. Ensayo de ciclo celular con ioduro de propidio (IP)

Para el análisis del perfil del ciclo celular, se analizó el contenido de ADN mediante el ensayo de ioduro de propidio. Las células de fenotipo wt y las C19MC(-) fueron contadas y plaqueadas en una determinada confluencia (200.000 cél. / well de una mw12) y se crecieron durante 3 días hasta alcanzar una confluencia del 80-90%. Luego, se disgregaron de manera enzimática utilizando la enzima Tryple 1X tal como se explicó anteriormente. El *pellet* celular fue resuspendido en 300 μ l de *Staining buffer* (PBS1X, SFB 3%) y las células fueron fijadas con 2ml de etanol 70%. Es importante que el paso de fijación se realice de manera lenta, incorporando el etanol gota a gota y con las células en constante agitación, para evitar así la formación de dobletes y tripletes de células. Luego se dejaron las muestras en la solución de fijación durante 40min a 4°C (también es posible dejar las células en ésta solución y en frío, durante varias semanas). Una vez transcurrido el tiempo de fijación, se centrifugaron las células y los pellets se resuspendieron en 1 ml de *Staining buffer*. Este paso se repitió 3 veces.

En el último lavado, se resuspendió las muestras en 500 μ l de *Staining buffer*, se agregaron 20 μ l de la enzima *RNAse* 20mg/ml y se incubaron 30 min a 37 °C, para degradar el ARN. Por último, se agregaron 10 μ l de yoduro de propidio (1mg/ml), se incubó 5 min y se analizó el contenido de ADN mediante citometría de flujo. El análisis posterior para determinar los % de células en las distintas fases del ciclo celular se realizó en el programa Flowjo, mediante el ajuste al modelos de Watson.

6.26. Ensayo de co-inmunoprecipitación de Argonauta (AgoIP)

Los ensayos de Ago-IP fueron realizados mediante el kit *miRNA Target IP kit* de la empresa Active Motif. Para realizar estos ensayos fue necesario transfectar 10 millones de células para cada reacción de inmunoprecipitación. Luego de la transfección, las células crecidas en adhesión fueron lavadas dos veces utilizando PBS1X frío y se las disgregó de manera suave utilizando un *scraper*. A continuación se centrifugaron a 400xg 5 min a 4°C y el pellet fue resuspendido en 1ml de PBS1X frío. Es importante aclarar que durante toda esta primera parte fue necesario trabajar en frío para preservar la integridad de las células y evitar la acción de nucleasas. Se contaron las células, para asegurarse de tener 10 millones para cada reacción de precipitación y luego fueron nuevamente centrifugadas 5 min a 400xg a 4°C. Los *pellets* fueron resuspendidos en 150 μ l del *Complete Lysis Buffer*. Este buffer es un buffer de lisis suave proporcionado por el kit que permite conservar las interacciones proteína-ARN. El lisado de células fue guardado *over night* (O.N.) a -80°C. Al día siguiente se realizó el bloqueo de las bolitas magnéticas, incubándolas con una solución de BSA 50mg/ml durante 10 min a temperatura ambiente en rotación *end-over-end*. Al finalizar, las bolitas magnéticas fueron precipitadas utilizando una barra magnética y lavadas dos veces con 500 μ l de *Buffer Wash AM7*. A continuación se conjugaron 50 μ l de bolitas magnéticas con 5 μ l del anticuerpo anti-Ago1/2/3 o 25 μ l del control negativo IgG, durante 30 min, en rotación y a temperatura ambiente. Una vez conjugadas, las bolitas junto con los anticuerpos fueron colectados en el fondo del tubo con la barra magnética y lavadas con 500 μ l de *Buffer Wash AM7*. Éste último paso se repitió 3 veces. A continuación se descongelaron los lisados celulares, se centrifugaron durante 15 min a 20.000xg a 4°C. Luego, del sobrenadante se guardaron 10 μ l para utilizar luego como un control del *input* y el resto se combinaron con las bolitas magnéticas conjugadas resuspendidas en el *Immunoprecipitation buffer*. Estas reacciones de inmunoprecipitación se incubaron

O.N. a 4°C en rotación *end-over-end*. Al día siguiente, se precipitaron los complejos proteína-ARN unidos a los anticuerpos/bolitas magnéticas utilizando la barra magnética, se descartó el sobrenadante y se lavó seis veces con 500µl del *Buffer Wash AM7* frío. Luego del último lavado, se resuspendieron los complejos en 10µl del *Buffer Wash AM7* y se agregó 140µl del *Proteinase K Digestion buffer* a cada tubo para degradar los anticuerpos y las proteínas presentes en las muestras. Ésta mezcla se incubó a 55°C durante 30 min en agitación. Al finalizar la incubación, se precipitaron las bolitas magnéticas con la barra magnética, se recuperó el eluido y se agregó 250µl del *Buffer Wash AM7*. A continuación, se realizó una extracción orgánica del ARN agregando 400µl de fenol:cloroformo:alcohol isoamílico (25:24:1) a cada tubo. Se centrifugó durante 15 min a 19.000xg y luego se transfirió la fase acuosa de arriba a un tubo nuevo. Se agregaron 400µl a la fase acuosa y se volvió a centrifugar durante 10 min a 19.000xg. Se recuperó la fase acuosa, se mezcló con el buffer de precipitación, (éste contiene el *precipitation buffer* con una solución de sal, glucógeno y etanol 100%), y se incubó O.N. a -80°C para precipitar el ARN. Al día siguiente, se centrifugaron las muestras durante 40 min a 20.000xg a 4°C, se descartó el sobrenadante y se lavaron los *pellets* con 500µl de etanol 80%. Finalmente se descartó el sobrenadante y se dejó secar los *pellets* para luego resuspenderlos en agua *RNase-free*. Estas muestras fueron cuantificadas utilizando un espectrofotómetro automático NanoDrop 1000 y enviadas al servicio de secuenciación masiva.

6.27. Secuenciación masiva de ARN (RNA-Seq)

- La secuenciación masiva de pequeños ARNs se realizó en el servicio de secuenciación de la Universidad de Los Angeles (TCGB Resources UCLA Path & Lab Med), gracias a la colaboración con el grupo del Dr. Evseenko. Se utilizó un secuenciador Illumina HiSeq2500, y se corrieron las 9 muestras en un único *lane*, con una profundidad de lectura de 8.5 de *reads* por muestra y en *single-end*. Para armar las librerías se utilizó 200ng de cada muestra y el kit *NEBnext small RNA library prep* de New England Biolabs, siguiendo el protocolo del fabricante. Los datos de la secuenciación fueron cargados en GEO con el *accession number*: GSE108021 (entrar al link).
- La secuenciación de ARN mensajeros del ensayo Ago-IP se realizó también en una plataforma de secuenciación de Illumina, gracias a la colaboración con el Dr. Mercola, en el servicio de secuenciación de Sanford Burnham Prebys Medical Discovery Institute,

Florida, USA. La profundidad de lectura fue de 35 millones de *reads* por muestra y en *single-end*. Para armar las librerías se utilizaron 100ng de cada muestra y el kit Ovation Ultralow Library System v2 de NuGen, siguiendo el protocolo del fabricante. Los datos de esta secuenciación serán cargados próximamente a la base de datos GEO.

6.28. Análisis bioinformático de los resultados de RNA-Seq

- Análisis de secuenciación masiva de pequeños ARNs: Los resultados crudos de la secuenciación fueron analizados siguiendo el *pipeline* CAPmiRSEQ (v1.1) (Sun et al., 2014). Brevemente, los resultados fueron analizados mediante el programa FastQC para controlar la calidad de los datos, luego se removieron las secuencias adaptadores y las bases de baja calidad, utilizando el programa Cutadapt (v1.13). Luego, se realizó el alineamiento de las secuencias con el programa Bowtie (Langmead et al., 2009) y HTSeq (Anders et al., 2015) para la cuantificación de las *reads* y las secuencias fueron mapeadas al genoma de referencia (hg19) con el programa mirDeep2 (Friedländer et al., 2008). Este programa permite identificar las secuencias de microARNs conocidos y desconocidas (*novel* microARNs) (se utilizó como base de datos miRBase v21). Por último, los datos fueron normalizados como RPKM (*counts per millon*, $1000.000 * \text{raw_counts} / \text{total_counts}$) y el análisis de expresión diferencial se realizó con el paquete de R, DESeq2 (Love et al., 2014).

Para el análisis de los isomiRs, se utilizó el paquete miraligner/isomiR (v1.6) de Bioconductor en R (Pantano et al., 2009) y el programa Chimira, el cual se utiliza desde la plataforma de la web (Vitsios and Enright, 2015). El primero, se utilizó para el análisis de el número de variantes de isoformas, su abundancia absoluta y normalizada en las tres poblaciones y la posición y frecuencia de los SNPs, para el cual fue necesario realizar nuevamente el alineamiento mediante el programa miRTOP, desarrollado por el mismo grupo [Lorena Pantano](<https://github.com/lpantano>) (Bioinformatic Core, Harvard Chan School, Boston, USA). El programa Chimira se utilizó principalmente para el análisis de las isoformas ADAR.

La predicción de genes blanco se realizó con el paquete de Bioconductor miRNAatap, el cual utiliza los algoritmos de predicción de DIANA, Miranda, PicTar, TargetScan y miRDB. El análisis de genes blanco de los isomiRs se realizó con el programa MR-microT

de DIANA Tools. El análisis de *Gene Ontology* se realizó con el paquete de Bioconductor *enrichGO* en R.

- Análisis de secuenciación masiva de ARN mensajeros (ensayo de Ago-IP):

Los resultados de la secuenciación masiva fueron inicialmente analizados mediante el programa FastQC para evaluar la calidad de la secuenciación. Estos datos fueron enviados desde el servicio de secuenciación *post-trimming* de los adaptadores, por lo cual no fue necesario remover las secuencias adaptadoras. Posteriormente, se realizó el alineamiento de las secuencias con el programa STAR (v2.5.3.a) utilizando los parámetros estandarizados y el genoma de referencia hg19. Los alineamientos que *mapearon* con más de un sitio (*multihits*) fueron filtrados con el programa SAM tools (v1.5) con un valor $q=3$. Luego mediante el paquete de Bioconductor RIP-Seeker (v1.18) se identificaron las regiones o picos en donde alinearon las secuencias y se realizó la anotación de éstas. A partir de este análisis, definimos como interacción específica con la proteína Argonau- ta aquellos ARN mensajeros que presentaron un enriquecimiento de 2 veces o más con respecto a los ARN mensajeros que interactúan con el control IgG y un número de *reads* igual o mayor a 5. A continuación, realizamos un análisis de enriquecimiento diferencial entre las muestras con el mimic miR-520a y el mimic control negativo, mediante el paquete DESeq2. En este análisis, definimos a los ARN mensajeros blanco específicos para el miR-520a a aquellos mensajeros que presentaron un enriquecimiento de dos veces o más con respecto al mimic control y un p-valor ajustado mayor a 0.1. El análisis de los genes blanco predichos para el miR-520a se realizó mediante el programa miRWalk, y se utilizaron los algoritmos de búsqueda miRWalk, TargetScan, Pictar2 y RNA22. Y el análisis de los términos de *Gene Ontology* se realizó en R mediante el paquete Enrichr (Kuleshov et al., 2016).

6.29. Análisis estadísticos

Los resultados experimentales son presentados como la media $\pm$ error estándar de la media (SEM) para al menos tres réplicas biológicas. El análisis estadístico fue realizado mediante estudios de ANOVA en bloques aleatorizados seguido de pruebas de comparación múltiples de Dunnet para los experimentos de PCR cuantitativa de la diferenciación cardíaca o mediante una prueba t de Student para los experimento de PCR cuantitativa de la diferenciación a

mesodermo, los ensayos de apoptosis, los ensayos de proliferación de incorporación de BrdU, análisis del perfil de ciclo celular con IP y los ensayos de sobreexpresión e inhibición del miR-205 en la formación del MPC. Se consideró diferencias entre grupos como estadísticamente significativas cuando el valor p fue menor o igual a 0,05. Los análisis estadísticos fueron realizados utilizando el programa Graphpad Prism (v6.01) y el InfoStat.

7. Bibliografía

- Ahmed, F., Senthil-Kumar, M., Lee, S., Dai, X., Mysore, K. S., and Zhao, P. X. Comprehensive analysis of small RNA-seq data reveals that combination of miRNA with its isomiRs increase the accuracy of target prediction in Arabidopsis. *RNA Biol*, 6286(May 2015):0, 2015.
- Akhtar, M. M., Micolucci, L., Islam, S., Olivieri, F., and Procopio, A. D. Bioinformatic tools for microRNA dissection. *Nucleic Acids Research*, 44(1), 2016.
- Amit, M., Carpenter, M. K., Inokuma, M. S., Chiu, C. P., Harris, C. P., Waknitz, M. A., Itskovitz-Eldor, J., and Thomson, J. A. Clonally derived human embryonic stem cell lines maintain pluripotency and proliferative potential for prolonged periods of culture. *Developmental Biology*, 227(2):271–278, 2000.
- Anders, S., Pyl, P. T., and Huber, W. HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics*, 31(2):166–169, 2015.
- Annis, D. S., Mosher, D. F., and Roberts, D. D. Origins and Fates of Cardiovascular Progenitor Cells. *Cell*, 27(4):339–351, 2009.
- Bartel, D. P. MicroRNAs: Target Recognition and Regulatory Functions. *Cell*, 136(2):215–233, 2009.
- Barwari, T., Joshi, A., and Mayr, M. MicroRNAs in Cardiovascular Disease. *Journal of the American College of Cardiology*, 68(23):2577–2584, 2016. ISSN 15583597.
- Bernstein, E., Kim, S. Y., Carmell, M. A., Murchison, E. P., Alcorn, H., Li, M. Z., Mills, A. A., Elledge, S. J., Anderson, K. V., and Hannon, G. J. Dicer is essential for mouse development. *Nature Genetics*, 35(3):215–217, 2003.
- Blugüermann, C., Romorini, L., Evseenko, D., Garate, X., Neiman, G., Sevlever, G., Scassa, M., and Miriuka, S. Leukemia Inhibitory Factor Increases Survival of Pluripotent Stem Cell-Derived Cardiomyocytes. *Journal of Cardiovascular Translational Research*, 2017.
- Borchert, G. M., Lanier, W., and Davidson, B. L. RNA polymerase III transcribes human microRNAs. *Nature Structural and Molecular Biology*, 13(12):1097–1101, 2006. ISSN 15459993.

- Boštjančič, E., Zidar, N., Štajer, D., and Glavač, D. MicroRNAs miR-1, miR-133a, miR-133b and miR-208 are dysregulated in human myocardial infarction. *Cardiology*, 115(3):163–169, 2010. ISSN 00086312.
- Boyer, L. A., Tong, I. L., Cole, M. F., Johnstone, S. E., Levine, S. S., Zucker, J. P., Guenther, M. G., Kumar, R. M., Murray, H. L., Jenner, R. G., Gifford, D. K., Melton, D. A., Jaenisch, R., and Young, R. A. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*, 122(6):947–956, 2005.
- Brand, T. Heart development: Molecular insights into cardiac specification and early morphogenesis. *Developmental Biology*, 258(1):1–19, 2003.
- Bruneau, B. G. The developmental genetics of congenital heart disease. *Nature*, 451(7181):943–948, 2008.
- Burridge, P., Keller, G., Gold, J., and Wu, J. Production of de novo cardiomyocytes: human pluripotent stem cell differentiation and direct reprogramming. *Cell Stem Cell*, 10(1):16–28, 2012.
- Canver, M. C., Bauer, D. E., Dass, A., Yien, Y. Y., Chung, J., Masuda, T., Maeda, T., Paw, B. H., and Orkin, S. H. Characterization of genomic deletion efficiency mediated by clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells (The Journal of Biological Chemistry (2014) 289 (21312-21324) DOI: 10.1074/jbc. *Journal of Biological Chemistry*, 292(6):2556, 2014.
- Chang, H., Yi, B., Ma, R., Zhang, X., Zhao, H., and Xi, Y. CRISPR/cas9, a novel genomic tool to knock down microRNA in vitro and in vivo. *Scientific reports*, 6(October 2015):22312, 2016.
- Chen, C., Ridzon, D. A., Broomer, A. J., Zhou, Z., Lee, D. H., Nguyen, J. T., Barbisin, M., Xu, N. L., Mahuvakar, V. R., Andersen, M. R., Lao, K. Q., Livak, K. J., and Guegler, K. J. Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Research*, 33(20):1–9, 2005.
- Chen, G., Gulbranson, D. R., Hou, Z., Bolin, J. M., Probasco, M. D., Smuga-otto, K., Howden, S. E., Nicole, R., Propson, N. E., Wagner, R., Lee, G. O., Teng, J. M. C., and Thomson, J. a.

- Chemically defined conditions for human iPS cell derivation and culture. *Nature Methods*, 8(5):424–429, 2011.
- Chen, J.-F., Mandel, E. M., Thomson, J. M., Wu, Q., Callis, T., Hammond, E. S. M., Conlon, F. L., and Wang, D.-Z. The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. *Nat Genet*, 38(2):2–7, 2004.
- Chistiakov, D. A., Orekhov, A. N., and Bobryshev, Y. V. Cardiac-specific miRNA in cardiogenesis, heart function, and cardiac pathology (with focus on myocardial infarction). *Journal of Molecular and Cellular Cardiology*, 94:107–121, 2016.
- Cloonan, N., Wani, S., Xu, Q., Gu, J., Lea, K., Heater, S., Barbacioru, C., Steptoe, A. L., Martin, H. C., Nourbakhsh, E., Krishnan, K., Gardiner, B., Wang, X., Nones, K., Steen, J. A., Matigian, N. A., Wood, D. L., Kassahn, K. S., Waddell, N., Shepherd, J., Lee, C., Ichikawa, J., McKernan, K., Bramlett, K., Kuersten, S., and Grimmond, S. M. MicroRNAs and their isomiRs function cooperatively to target common biological pathways. *Genome biology*, 12(12):R126, 2011.
- Cordes, K. R. and Srivastava, D. MicroRNA regulation of cardiovascular development. *Circulation Research*, 104(6):724–732, 2009.
- Craft, A. M., Rockel, J. S., Nartiss, Y., Kandel, R. A., Alman, B. A., and Keller, G. M. Generation of articular chondrocytes from human pluripotent stem cells. *Nature Biotechnology*, 33(6):638–645, 2015.
- Dance, M. and Weber, M. C19MC microRNAs are processed from introns of large Pol-II , non-protein-coding transcripts. *Nucleic acids research*, 37(10):3464–3473, 2009.
- Denning, C., Allegrucci, C., Priddle, H., Barbadillo-Muñoz, M. D., Anderson, D., Self, T., Smith, N. M., Parkin, C. T., and Young, L. E. Common culture conditions for maintenance and cardiomyocyte differentiation of the human embryonic stem cell lines, BG01 and HUES-7. *International Journal of Developmental Biology*, 50(1):27–37, 2006.
- Derynck, R. and Zhang, Y. E. Smad-dependent and Smad-independent pathways in TGF-beta family signalling. *Nature*, 425(6958):577–84, 2003.
- Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., Xue, C., Marinov, G. K., Khatun, J., Williams, B. A., Zaleski, C.,

- Rozowsky, J., Röder, M., Kokocinski, F., Abdelhamid, R. F., Alioto, T., Antoshechkin, I., Baer, M. T., Bar, N. S., Batut, P., Bell, K., Bell, I., Chakraborty, S., Chen, X., Chrast, J., Curado, J., Derrien, T., Drenkow, J., Dumais, E., Dumais, J., Duttagupta, R., Falconnet, E., Fastuca, M., Fejes-Toth, K., Ferreira, P., Foissac, S., Fullwood, M. J., Gao, H., Gonzalez, D., Gordon, A., Gunawardena, H., Howald, C., Jha, S., Johnson, R., Kapranov, P., King, B., Kingswood, C., Luo, O. J., Park, E., Persaud, K., Preall, J. B., Ribeca, P., Risk, B., Robyr, D., Sammeth, M., Schaffer, L., See, L. H., Shahab, A., Skancke, J., Suzuki, A. M., Takahashi, H., Tilgner, H., Trout, D., Walters, N., Wang, H., Wrobel, J., Yu, Y., Ruan, X., Hayashizaki, Y., Harrow, J., Gerstein, M., Hubbard, T., Reymond, A., Antonarakis, S. E., Hannon, G., Giddings, M. C., Ruan, Y., Wold, B., Carninci, P., Guig, R., and Gingeras, T. R. Landscape of transcription in human cells. *Nature*, 489(7414):101–108, 2012.
- Ebert, M. S. and Sharp, P. A. Roles for MicroRNAs in conferring robustness to biological processes. *Cell*, 149(3):505–524, 2012.
- Ebert, M. S., Neilson, J. R., and Sharp, P. A. MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. 4(9), 2013.
- Essletzbichler, P., Konopka, T., Santoro, F., Chen, D., Gapp, B. V., Kralovics, R., Brummelkamp, T. R., Nijman, S. M. B., and Bürckstümmer, T. Megabase-scale deletion using CRISPR/Cas9 to generate a fully haploid human cell line. *Genome Research*, 24(12):2059–2065, 2014.
- Evseenko, D., Zhu, Y., Schenke-layland, K., Kuo, J., Latour, B., Ge, S., and Scholes, J. Mapping the first stages of mesoderm commitment during differentiation of human embryonic stem cells. *Proceedings of the National Academy of Sciences*, 2010.
- Fan, Y., Siklenka, K., Arora, S. K., Ribeiro, P., Kimmins, S., and Xia, J. miRNet - dissecting miRNA-target interactions and functional associations through network-based visual analysis. *Nucleic acids research*, 44(W1):W135–W141, 2016. ISSN 13624962.
- Fazi, F. and Nervi, C. MicroRNA: basic mechanisms and transcriptional regulatory networks for cell fate determination. *Cardiovascular research*, 79(4):553–61, sep 2008.
- Filipowicz, W., Bhattacharyya, S. N., and Sonenberg, N. Mechanisms of post-transcriptional regulation by microRNAs: Are the answers in sight? *Nature Reviews Genetics*, 9(2):102–114, 2008. ISSN 14710056.

- Freund, C., Ward-van Oostwaard, D., Monshouwer-Kloots, J., van den Brink, S., van Rooijen, M., Xu, X., Zweigerdt, R., Mummery, C., and Passier, R. Insulin Redirects Differentiation from Cardiogenic Mesoderm and Endoderm to Neuroectoderm in Differentiating Human Embryonic Stem Cells. *Stem Cells*, 26(3):724–733, 2008.
- Friedländer, M. R., Chen, W., Adamidi, C., Maaskola, J., Einspanier, R., Knespel, S., and Rajewsky, N. Discovering microRNAs from deep sequencing data using miRDeep. *Nature Biotechnology*, 26(4):407–415, 2008. ISSN 10870156.
- Friedman, R. C., Farh, K. K.-h., Burge, C. B., and Bartel, D. P. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Research*, pages 92–105, 2009.
- Gao, J. and Liu, Q. G. The role of miR-26 in tumors and normal tissues (Review). *Oncology Letters*, 2(6):1019–1023, 2011. ISSN 17921074.
- Garate, X., Greca, A. L., Neiman, G., Blugüermann, C., Santin Velazque, N. L., Luzzani, C., Scassa, M. É., Sevlever, E., Romorini, L., and Miriuka, S. G. Identification of the miRNAome of early mesoderm progenitor cells and cardiomyocytes derived from human pluripotent stem cells. *Scientific Reports*, (In press), 2018.
- Gill, J., Langer, E., and Lindsley, R. Snail and the miR-200 family act in opposition to regulate EMT and germ layer fate restriction in differentiating ES cells. *Stem Cells*, 29(5):764–776, 2011.
- Gong, J., Tong, Y., Zhang, H. M., Wang, K., Hu, T., Shan, G., Sun, J., and Guo, A. Y. Genome-wide identification of SNPs in MicroRNA genes and the SNP effects on MicroRNA target binding and biogenesis. *Human Mutation*, 33(1):254–263, 2012. ISSN 10597794.
- Gregory, P. a., Bert, A. G., Paterson, E. L., Barry, S. C., Tsykin, A., Farshid, G., Vadas, M. a., Khew-Goodall, Y., and Goodall, G. J. The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. *Nature cell biology*, 10(5):593–601, may 2008a.
- Gregory, P. A., Bracken, C. P., Bert, A. G., and Goodall, G. J. MicroRNAs as regulators of epithelial-mesenchymal transition. *Cell Cycle*, 7(20), 2008b.

- Guarino, M., Rubino, B., and Ballabio, G. The role of epithelial-mesenchymal transition in cancer pathology. *Pathology*, 39(3):305–318, 2007.
- Guo, L. and Liang, T. MicroRNAs and their variants in an RNA world: implications for complex interactions and diverse roles in an RNA regulatory network. *Briefings in Bioinformatics*, (November):bbw124, 2016.
- Ha, M. and Kim, V. N. Regulation of microRNA biogenesis. *Nature Reviews Molecular Cell Biology*, 15(8):509–524, 2014.
- Hausser, J. and Zavolan, M. Identification and consequences of miRNA-target interactions-beyond repression of gene expression. *Nature Reviews Genetics*, 15(9):599–612, 2014.
- Heldin, C. H., Vanlandewijck, M., and Moustakas, A. Regulation of EMT by TGF β in cancer. *FEBS Letters*, 586(14):1959–1970, 2012.
- Huntzinger, E. and Izaurralde, E. Gene silencing by microRNAs: Contributions of translational repression and mRNA decay. *Nature Reviews Genetics*, 12(2):99–110, 2011.
- Itskovitz-Eldor, J., Schuldiner, M., Karsenti, D., Eden, A., Yanuka, O., Amit, M., Soreq, H., and Benvenisty, N. Differentiation of human embryonic stem cells into embryoid bodies compromising the three embryonic germ layers. *Molecular medicine (Cambridge, Mass.)*, 6(2): 88–95, 2000.
- Ivey, K. N. and Srivastava, D. MicroRNAs as regulators of differentiation and cell fate decisions. *Cell stem cell*, 7(1):36–41, 2010.
- Ivey, K. N., Muth, A., Arnold, J., King, F. W., Yeh, R.-F., Fish, J. E., Hsiao, E. C., Schwartz, R. J., Conklin, B. R., Bernstein, H. S., and Srivastava, D. MicroRNA regulation of cell lineages in mouse and human embryonic stem cells. *Cell stem cell*, 2(3):219–29, 2008.
- James, D. TGF β /activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells. *Development*, 132(6):1273–1282, 2005.
- Kehat, I. and Kenyagin-Karsenti, D. Human embryonic stem cells can differentiate into myocytes with structural and functional properties of cardiomyocytes. *Journal of Clinical Investigation*, 108(3):407–414, 2001.

- Keller, G. New Insights into Early Human Development : Lessons for Stem Cell Derivation and Differentiation. *Genes & development*, 19:1129–1155, 2005.
- Kim, K., Doi, A., Wen, B., Ng, K., Zhao, R., Cahan, P., Kim, J., Aryee, M. J., Ji, H., Ehrlich, L. I., Yabuuchi, A., Takeuchi, A., Cunniff, K. C., Hongguang, H., McKinney-Freeman, S., Naveiras, O., Yoon, T. J., Irizarry, R. A., Jung, N., Seita, J., Hanna, J., Murakami, P., Jaenisch, R., Weissleder, R., Orkin, S. H., Weissman, I. L., Feinberg, A. P., and Daley, G. Q. Epigenetic memory in induced pluripotent stem cells. *Nature*, 467(7313):285–290, 2010.
- Kim, P. C., Mo, R., and Hui, C.-c. Murine models of VACTERL syndrome: Role of sonic hedgehog signaling pathway. *Journal of Pediatric Surgery*, 36(2):381–384, 2001.
- Kliver, J., Gibcus, J. H., Hettinga, C., Adema, A., and Richter, M. K. S. Rapid Generation of MicroRNA Sponges for MicroRNA Inhibition. *PloS one*, 7(1):14–21, 2012.
- Krtolica, A., Genbacev, O., Escobedo, C., Zdravkovic, T., Nordstrom, A., Vabuenta, D., Nath, A., Simon, C., Mostov, K., and Fisher, S. J. Disruption of Apical-Basal Polarity of Human Embryonic Stem Cells Enhances Hematoendothelial Differentiation. *Stem Cells*, 25(9):2215–2223, 2007.
- Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S. L., Jagodnik, K. M., Lachmann, A., McDermott, M. G., Monteiro, C. D., Gunderson, G. W., and Ma'ayan, A. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic acids research*, 44(W1):W90–W97, 2016.
- Kurosawa, H. Methods for inducing embryoid body formation: in vitro differentiation system of embryonic stem cells. *Journal of Bioscience and Bioengineering*, 103(5):389–398, 2007.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, 10(3), 2009.
- Laurent, L. C., Chen, J., Ulitsky, I., Mueller, F.-J., Lu, C., Shamir, R., Fan, J.-B., and Loring, J. F. Comprehensive MicroRNA Profiling Reveals a Unique Human Embryonic Stem Cell Signature Dominated by a Single Seed Sequence. *Stem Cells*, 26(6):1506–1516, 2008. ISSN 10665099.

- Lee, R. C. The *C. elegans* Heterochronic Gene *lin-4* Encodes Small RNAs with Antisense Complementarity to *lin-14*. *Cell*, 75:843–854, 1993.
- Lewis, B. P., Shih, I.-h., Jones-Rhoades, M. W., Bartel, D. P., and Burge, C. B. Prediction of Mammalian MicroRNA Targets. *Cell*, 115(7):787–798, 2003.
- Li, C., Finkelstein, D., and Sherr, C. J. Arf tumor suppressor and miR-205 regulate cell adhesion and formation of extraembryonic endoderm from pluripotent stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, 110(12):E1112–21, 2013.
- Li, F., Wang, X., Bunger, P. C., and Gerdes, A. M. Formation of binucleated cardiac myocytes in rat heart: I. Role of actin-myosin contractile ring. *Journal of Molecular and Cellular Cardiology*, 29(6):1541–1551, 1997.
- Lian, X., Hsiao, C., Wilson, G., Zhu, K., Hazeltine, L. B., Azarin, S. M., Raval, K. K., Zhang, J., Kamp, T. J., and Palecek, S. P. Robust cardiomyocyte differentiation from human pluripotent stem cells via temporal modulation of canonical Wnt signaling. *Proceedings of the National Academy of Sciences*, 109(27):1–10, 2012.
- Liang, P., Lan, F., Lee, A. S., Gong, T., Sanchez-Freire, V., Wang, Y., Diecke, S., Sallam, K., Knowles, J. W., Wang, P. J., Nguyen, P. K., Bers, D. M., Robbins, R. C., and Wu, J. C. Drug screening using a library of human induced pluripotent stem cell-derived cardiomyocytes reveals disease-specific patterns of cardiotoxicity. *Circulation*, 127(16):1677–1691, 2013.
- Liu, N., Williams, A. H., Kim, Y., McAnally, J., Bezprozvannaya, S., Sutherland, L. B., Richardson, J. A., Bassel-Duby, R., and Olson, E. N. An intragenic MEF2-dependent enhancer directs muscle-specific expression of microRNAs 1 and 133. *Proceedings of the National Academy of Sciences*, 104(52):20844–20849, 2007. ISSN 0027-8424.
- Love, M. I., Huber, W., and Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12):1–21, 2014.
- Lytle, J. R., Yario, T. a., and Steitz, J. a. Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5' UTR as in the 3' UTR. *Proceedings of the National Academy of Sciences of the United States of America*, 104(23):9667–9672, 2007.

- Menasché, P., Vanneaux, V., Hagège, A., Bel, A., Cholley, B., Parouchev, A., Cacciapuoti, I., Al-Daccak, R., Benhamouda, N., Blons, H., Agbulut, O., Tosca, L., Trouvin, J.-H., Fabreguettes, J.-R., Bellamy, V., Charron, D., Tartour, E., Tachdjian, G., Desnos, M., and Larghero, J. Transplantation of Human Embryonic Stem Cell–Derived Cardiovascular Progenitors for Severe Ischemic Left Ventricular Dysfunction. *Journal of the American College of Cardiology*, 71(4):429–438, 2018.
- Mogilyansky, E. and Rigoutsos, I. The miR-17 / 92 cluster : a comprehensive update on its genomics , genetics , functions and increasingly important and numerous roles in health and disease. *Cell Death and Differentiation*, 20(12):1603–1614, 2013.
- Moretti, A., Bellin, M., and Laugwitz, K.-L. New England Journal. *The New England Journal of Medicine*, pages 411–421, 2010.
- Morin, R. D., Connor, M. D. O., Griffith, M., Kuchenbauer, F., Delaney, A., Prabhu, A.-l., Zhao, Y., McDonald, H., Zeng, T., Hirst, M., Eaves, C. J., and Marra, M. A. Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome research*, pages 610–621, 2008.
- Mouillet, J. F., Ouyang, Y., Coyne, C. B., and Sadovsky, Y. MicroRNAs in placental health and disease. *American Journal of Obstetrics and Gynecology*, 213(4):S163–S172, 2015.
- Mummery, C. Differentiation of Human Embryonic Stem Cells to Cardiomyocytes: Role of Coculture With Visceral Endoderm-Like Cells. *Circulation*, 107(>21):2733–2740, 2003.
- Nana-Sinkam, S. P. and Croce, C. M. MicroRNAs as therapeutic targets in cancer. *Translational Research*, 157(4):216–225, 2011.
- Narayanan, A., Hill-Teran, G., Moro, A., Ristori, E., Kasper, D. M., Roden, C. A., Lu, J., and Nicolli, S. In vivo mutagenesis of miRNA gene families using a scalable multiplexed CRISPR/Cas9 nuclease system. *Scientific Reports*, 2016.
- Neilsen, C. T., Goodall, G. J., and Bracken, C. P. IsomiRs - The overlooked repertoire in the dynamic microRNAome. *Trends in Genetics*, 28(11):544–549, 2012.
- Ng, E. S., Davis, R. P., Azzola, L., Stanley, E. G., Elefanty, A. G., and De, W. Forced aggregation of defined numbers of human embryonic stem cells into embryoid bodies fosters robust , repro-

- ducible hematopoietic differentiation Brief report Forced aggregation of defined numbers of human embryonic stem cells into embryoid bodies fost. *Blood journal*, 106(5):1601–1603, 2005.
- Nguyen, P. N. N., Choo, K. B., Huang, C.-J., Sugii, S., Cheong, S. K., and Kamarul, T. miR-524-5p of the primate-specific C19MC miRNA cluster targets TP53IPN1- and EMT-associated genes to regulate cellular reprogramming. *Stem cell research & therapy*, 8(1):214, 2017a.
- Nguyen, P. N. N., Huang, C.-j., Sugii, S., Cheong, S. K., and Choo, K. B. Selective activation of miRNAs of the primate-specific chromosome 19 miRNA cluster (C19MC) in cancer and stem cells and possible contribution to regulation of apoptosis. *J Biomed Sci*, 24:20, 2017b.
- Nishikura, K. A-to-I editing of coding and non-coding RNAs by ADARs. *Nature Reviews Molecular Cell Biology*, 17(2):83–96, 2016. ISSN 14710080.
- Noguer-dance, M., Abu-amero, S., Al-khtib, M., Lefe, A., Coullin, P., and Moore, G. E. The primate-specific microRNA gene cluster (C19MC) is imprinted in the placenta. *Human Molecular Genetics*, 19(18):3566–3582, 2010.
- Obad, S., Santos, C. O., Petri, A., Heidenblad, M., Ruse, C., Fu, C., Lindow, M., Stenvang, J., Marie, E., Hansen, H. F., Koch, T., Pappin, D., and Hannon, G. J. Silencing of microRNA families by seed-targeting tiny LNAs. *Nature Genetics*, 43(4):371–378, 2013.
- Ohtani, K. and Dimmeler, S. Control of cardiovascular differentiation by microRNAs. *Basic research in cardiology*, 106(1):5–11, 2011.
- Ong, S.-G., Lee, W. H., Kodo, K., and Wu, J. C. MicroRNA mediated Regulation of Differentiation and Transdifferentiation in stem cells. *Advanced Drug Delivery Reviews*, 22(2): 134–139, 2015.
- Pantano, L., Estivill, X., and Martí, E. SeqBuster, a bioinformatic tool for the processing and analysis of small RNAs datasets, reveals ubiquitous miRNA modifications in human embryonic cells. *Nucleic Acids Research*, 38(5), 2009.
- Park, S.-m., Gaur, A. B., Lengyel, E., and Peter, M. E. The miR-200 family determines the epithelial phenotype of cancer cells by targeting the E-cadherin repressors ZEB1 and ZEB2. *Genes & development*, pages 894–907, 2008.

- Peter, M. Let-7 and miR-200 microRNAs: guardians against pluripotency and cancer progression. *Cell Cycle*, 8(6):843–852, 2009.
- Place, R. F., Li, L.-C., Pookot, D., Noonan, E. J., and Dahiya, R. MicroRNA-373 induces expression of genes with complementary promoter sequences. *Proceedings of the National Academy of Sciences*, 105(5):1608–1613, 2008.
- Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., and Zhang, F. Genome engineering using the CRISPR-Cas9 system. *Nature protocols*, 8(11):2281–2308, 2013.
- Reinhart, B. J., Slack, F. J., Basson, M., Pasquienelll, A. E., Bettlnger, J. C., Rougvle, A. E., Horvitz, H. R., and Ruvkun, G. The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature*, 403(6772):901–906, 2000.
- Roberts, A. B. and Wakefield, L. M. The two faces of transforming growth factor beta in carcinogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, 100(15):8621–3, 2003.
- Romorini, L., Garate, X., Neiman, G., Luzzani, C., Furmento, V. A., Guberman, A. S., Sevillever, G. E., Scassa, M. E., and Miriuka, S. G. AKT/GSK3 β signaling pathway is critically involved in human pluripotent stem cell survival. *Scientific Reports*, 6(November 2015):1–15, 2016. ISSN 20452322. doi: 10.1038/srep35660.
- Ruijter, J. M., Ramakers, C., Hoogaars, W. M. H., Karlen, Y., Bakker, O., Hoff, M. J. B. V. D., and Moorman, A. F. M. Amplification efficiency : linking baseline and bias in the analysis of quantitative PCR data. *Nucleic acids research*, 37(6), 2009.
- Saga, Y., Hata, N., Kobayashi, S., Magnuson, T., Seldin, M. F., and Taketo, M. M. MesP1: a novel basic helix-loop-helix protein expressed in the nascent mesodermal cells during mouse gastrulation. *Development (Cambridge, England)*, 122:2769–2778, 1996.
- Saini, H. K., Griffiths-Jones, S., and Enright, A. J. Genomic analysis of human microRNA transcripts. *Proceedings of the National Academy of Sciences*, 104(45):17719–17724, 2007.
- Sander, J. D. and Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nature Biotechnology*, 32(4):347–350, 2014.

- Sato, N., Munoz Sanjuan, I., Heke, M., Uchida, M., Naef, F., and Brivanlou, A. H. Molecular signature of human embryonic stem cells and its comparison with the mouse. *Developmental Biology*, 260(2):404–413, 2003.
- Schwarzenbach, H., Nishida, N., Calin, G. A., and Pantel, K. Clinical relevance of circulating cell-free microRNAs in cancer. *Nature Reviews Clinical Oncology*, 11(3):145–156, 2014.
- Slack, F. J., Basson, M., Liu, Z., Ambros, V., Horvitz, H. R., and Ruvkun, G. The lin-41 RBCC gene acts in the C. elegans heterochronic pathway between the let-7 regulatory RNA and the LIN-29 transcription factor. *Molecular Cell*, 5(4):659–669, 2000.
- Smoak, I. W., Byrd, N. A., Abu-Issa, R., Goddeeris, M. M., Anderson, R., Morris, J., Yamamura, K., Klingensmith, J., and Meyers, E. N. Sonic hedgehog is required for cardiac outflow tract and neural crest cell development. *Developmental Biology*, 283(2):357–372, 2005.
- Spence, T., Perotti, C., Sin-Chan, P., Picard, D., Wu, W., Singh, A., Anderson, C., Blough, M. D., Cairncross, J. G., Lafay-Cousin, L., Strother, D., Hawkins, C., Narendran, A., Huang, A., and Chan, J. A. A novel C19MC amplified cell line links Lin28/let-7 to mTOR signaling in embryonal tumor with multilayered rosettes. *Neuro-Oncology*, 16(1):62–71, 2014. ISSN 15235866.
- Srivastava, D. Making or Breaking the Heart: From Lineage Determination to Morphogenesis. *Cell*, 126(6):1037–1048, 2006.
- Stadler, B., Ivanovska, I., Mehta, K., Song, S., Nelson, A., Tan, Y., Mathieu, J., Darby, C., Blau, C. A., Ware, C., Peters, G., Miller, D. G., Shen, L., Cleary, M. a., and Ruohola-Baker, H. Characterization of microRNAs involved in embryonic stem cell states. *Stem cells and development*, 19(7):935–50, 2010.
- Staerk, J., Dawlaty, M. M., Gao, Q., Maetzel, D., Hanna, J., and Cesar, A. Reprogramming of peripheral blood cells to induced pluripotent stem cells. *Cell Stem Cell*, 7(1):20–24, 2010.
- States, U. Postnatal isl1+ cardioblasts enter fully differentiated cardiomyocyte lineages. *Nature*, 106(3):201–207, 2017.
- Subramanyam, D., Lamouille, S., Judson, R. L., Liu, J. Y., Bucay, N., Derynck, R., and Blelloch,

- R. Multiple targets of miR-302 and miR-372 promote reprogramming of human fibroblasts to induced pluripotent stem cells. *Nature Biotechnology*, 29(5):443–448, 2011.
- Suh, M.-R., Lee, Y., Kim, J. Y., Kim, S.-K., Moon, S.-H., Lee, J. Y., Cha, K.-Y., Chung, H. M., Yoon, H. S., Moon, S. Y., Kim, V. N., and Kim, K.-S. Human embryonic stem cells express a unique set of microRNAs. *Developmental biology*, 270(2):488–98, 2004.
- Sun, Y. M., Lin, K. Y., and Chen, Y. Q. Diverse functions of miR-125 family in different cell contexts. *Journal of Hematology & Oncology*, page 6, 2013. ISSN 17568722.
- Sun, Z., Evans, J., Bhagwate, A., Middha, S., Bockol, M., Yan, H., and Kocher, J. P. CAP-miRSeq: A comprehensive analysis pipeline for microRNA sequencing data. *BMC Genomics*, 15(1): 1–10, 2014.
- Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., and Tomoda, K. Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell*, pages 861–872, 2007.
- Tam, P. P. and Behringer, R. R. Mouse gastrulation: The formation of a mammalian body plan. *Mechanisms of Development*, 68(1-2):3–25, 1997.
- Tan, G. C., Chan, E., Molnar, A., Sarkar, R., Alexieva, D., Isa, I. M., Robinson, S., Zhang, S., Ellis, P., Langford, C. F., Guillot, P. V., Chandrashekan, A., Fisk, N. M., Castellano, L., Meister, G., Winston, R. M., Cui, W., Baulcombe, D., and Dibb, N. J. 5 isomiR variation is of functional and evolutionary importance. *Nucleic Acids Research*, 42(14):9424–9435, 2014.
- Tan, P.-H., Pao, Y.-Y., Cheng, J.-K., Hung, K.-C., and Liu, C.-C. MicroRNA-based therapy in pain medicine: Current progress and future prospects. *Acta Anaesthesiologica Taiwanica*, 51 (4):171–176, 2013.
- Tay, Y., Zhang, J., Thomson, A. M., Lim, B., and Rigoutsos, I. MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature*, 455(7216): 1124–8, 2008.
- Thies, R. S. and Murry, C. E. The advancement of human pluripotent stem cell-derived therapies into the clinic. *Development*, 142(18):3077–3084, 2015.

- Thomson, J. A., Itskovitz-eldor, J., Shapiro, S. S., Waknitz, M. A., Swiergiel, J. J., Marshall, V. S., and Jones, J. M. Embryonic Stem Cell Lines Derived from Human Blastocysts. *Science*, 282 (5391):1145–1147, 1998.
- Thum, T. MicroRNA therapeutics in cardiovascular medicine. *EMBO Molecular Medicine*, 4 (1):3–14, 2012.
- Thum, T., Catalucci, D., and Bauersachs, J. MicroRNAs: novel regulators in cardiac development and disease. *Cardiovascular research*, 79(4):562–70, 2008.
- Tohyama, S., Hattori, F., Sano, M., Hishiki, T., Nagahata, Y., Matsuura, T., Hashimoto, H., Suzuki, T., Yamashita, H., Satoh, Y., Egashira, T., Seki, T., Muraoka, N., Yamakawa, H., Ohgino, Y., Tanaka, T., Yoichi, M., Yuasa, S., Murata, M., Suematsu, M., and Fukuda, K. Distinct metabolic flow enables large-scale purification of mouse and human pluripotent stem cell-derived cardiomyocytes. *Cell Stem Cell*, 12(1):127–137, 2013.
- Townley-Tilson, W. H. D., Callis, T. E., and Wang, D. MicroRNAs 1, 133, and 206: critical factors of skeletal and cardiac muscle development, function, and disease. *The international journal of biochemistry & cell biology*, 42(8):1252–5, 2010.
- Urgese, G., Paciello, G., Acquaviva, A., and Ficarra, E. isomiR-SEA: an RNA-Seq analysis tool for miRNAs/isomiRs expression level profiling and miRNA-mRNA interaction sites evaluation. *BMC Bioinformatics*, 17(1):148, 2016.
- van Rooij, E., Sutherland, L. B., Qi, X., Richardson, J. a., Hill, J., and Olson, E. N. Control of stress-dependent cardiac growth and gene expression by a microRNA. *Science (New York, N. Y.)*, 316(5824):575–9, 2007.
- Vasudevan, S., Tong, Y., and Steitz, J. A. Switching from repression to activation: MicroRNAs can up-regulate translation. *Science*, 318(5858):1931–1934, 2007.
- Vidigal, J. A. and Ventura, A. The biological functions of miRNAs: lessons from in vivo studies. *Trends cell biology*, 13(2):137–147, 2015. ISSN 09652140.
- Vijayaragavan, K. and Bhatia, M. Early cardiac development: a Wnt beat away. *Proceedings of the National Academy of Sciences of the United States of America*, 104(23):9549–50, 2007.

- Vitsios, D. M. and Enright, A. J. Chimira: Analysis of small RNA sequencing data and microRNA modifications. *Bioinformatics*, 31(20):3365–3367, 2015.
- Wang, G., Guo, X., Hong, W., Liu, Q., Wei, T., Lu, C., Gao, L., Ye, D., and Zhou, Y. Critical regulation of miR-200 / ZEB2 pathway in Oct4 / Sox2-induced mesenchymal-to-epithelial transition and induced pluripotent stem cell generation. *Proceedings of the National Academy of Sciences of the United States of America*, pages 1–6, 2013.
- Wang, Y., Medvid, R., Melton, C., and Jaenisch, R. DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. *Nature Genetics*, 39(3):380–385, 2010.
- Willems, E., Spiering, S., Davidovics, H., Lanier, M., Xia, Z., Dawson, M., Cashman, J., and Mercola, M. Small molecule inhibitors of the Wnt pathway potently promote cardiomyocytes from human embryonic stem cell derived mesoderm. *Circulation Research*, 109(4):360–364, 2011.
- Winter, J., Jung, S., Keller, S., Gregory, R. I., and Diederichs, S. Many roads to maturity: MicroRNA biogenesis pathways and their regulation. *Nature Cell Biology*, 11(3):228–234, 2009.
- Wu, S., Huang, S., Ding, J., Zhao, Y., Liang, L., Liu, T., Zhan, R., and He, X. Multiple microRNAs modulate p21Cip1/Waf1 expression by directly targeting its 3' untranslated region. *Oncogene*, 29(15):2302–2308, 2010.
- Xie, L., Mouillet, J.-f., Chu, T., Parks, W. T., Sadovsky, E., Knöfler, M., and Sadovsky, Y. C19MC MicroRNAs Regulate the Migration of Human Trophoblasts. *Reproduction-Development*, 155(February):4975–4985, 2015.
- Xu, C., Inokuma, M. S., Denham, J., Golds, K., Kundu, P., Gold, J. D., and Carpenter, M. K. Feeder-free growth of undifferentiated human embryonic stem cells. *Nature Biotechnology*, 19(10):971–974, 2001.
- Xu, N., Papagiannakopoulos, T., Pan, G., Thomson, J. a., and Kosik, K. S. MicroRNA-145 regulates OCT4, SOX2, and KLF4 and represses pluripotency in human embryonic stem cells. *Cell*, 137(4):647–58, 2009.

- Xu, X., Chen, Z., Zhao, X., Wang, J., Ding, D., Wang, Z., Tan, F., Tan, X., Zhou, F., Sun, J., Sun, N., Gao, Y., Shao, K., Li, N., Qiu, B., and He, J. MicroRNA-25 promotes cell migration and invasion in esophageal squamous cell carcinoma. *Biochemical and Biophysical Research Communications*, 421(4):640–645, 2012.
- Yang, L., Soonpaa, M. H., Adler, E. D., Roepke, T. K., Kattman, S. J., Kennedy, M., Henckaerts, E., Bonham, K., Abbott, G. W., Linden, R. M., Field, L. J., and Keller, G. M. Human cardiovascular progenitor cells develop from a KDR+ embryonic-stem-cell-derived population. *Nature*, 453(7194):524–528, 2008.
- Yin, M. and Pacifici, M. Vascular regression is required for mesenchymal condensation and chondrogenesis in the developing limb. *Developmental Dynamics*, 222(3):522–533, 2001.
- Zernicka-Goetz, M., Morris, S. A., and Bruce, A. W. Making a firm decision: Multifaceted regulation of cell fate in the early mouse embryo. *Nature Reviews Genetics*, 10(7):467–477, 2009.
- Zhang, L., Jia, R., Palange, N. J., Satheka, A. C., Togo, J., An, Y., Humphrey, M., Ban, L., Ji, Y., Jin, H., Feng, X., and Zheng, Y. Large genomic fragment deletions and insertions in mouse using CRISPR/Cas9. *PLoS ONE*, 10(3):1–14, 2015a. ISSN 19326203.
- Zhang, Z., Hong, Y., Xiang, D., Zhu, P., Wu, E., Li, W., Mosenson, J., and Wu, W. S. MicroRNA-302/367 cluster governs hESC self-renewal by dually regulating cell cycle and apoptosis pathways. *Stem Cell Reports*, 4(4):645–657, 2015b.
- Zhu, Z. and Huangfu, D. Human pluripotent stem cells: an emerging model in developmental biology. *Development*, 140(4):705–717, 2013.