



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Maestría y Doctorado en Ciencias Bioquímicas

Estudio de la función de un RNA largo no-codificante con niveles de expresión alterados
en cáncer de mama

TESIS

QUE PARA OPTAR POR EL GRADO DE:

Maestra en Ciencias

PRESENTA:

LCG Mariana del Rosario Ruiz Velasco Leyva

Dr. Félix Recillas Targa

[Instituto de Fisiología Celular \(UNAM\)](#)

MIEMBROS DEL COMITÉ TUTOR

Dra. Rosa Estela Navarro González

[Instituto de Fisiología Celular \(UNAM\)](#)

Dr. Federico Sánchez Rodríguez

[Instituto de Biotecnología \(UNAM\)](#)

MÉXICO, D. F. Agosto de 2014



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

Para mi familia, Daniela, Rosario y Edmundo, a quienes amo profundamente y guiaron mis pasos hasta este día.

Agradecimientos académicos

El presente trabajo fue elaborado en el Instituto de Fisiología Celular en el laboratorio del Dr. Félix Recillas Targa, con los apoyos DGAPA-PAPIIT IN203811, IN201114 de la Dirección General de Asuntos del Personal Académico y 128464, 220503 del Consejo Nacional de Ciencia y Tecnología.

Los estudios de maestría de la LCG Mariana del Rosario Ruiz Velasco Leyva, con número de CVU 483679, se realizaron con el apoyo de la Beca CONACYT Nacional para posgrados.

Se extiende un sincero agradecimiento al comité tutor que asesoró este proyecto, el cual estuvo constituido por:

Dr. Félix Recillas Targa	Instituto de Fisiología Celular
Dra. Rosa Estela Navarro González	Instituto de Fisiología Celular
Dr. Federico Sánchez Rodríguez	Instituto de Biotecnología

Se agradece la asesoría de la Dra. Lourdes Massieu Trigo, así como el excelente apoyo técnico de la Biol. Georgina Guerrero Avendaño, el Dr. Fernando Suaste Olmos, la Biol. Teresa Montiel Montes y el M. en C. José Luis Ventura Gallegos. Asimismo, se reconoce el trabajo de la laboratorista Gianelli Cortés González, el apoyo de la Unidad de Biología Molecular y de Leticia García Gutiérrez.

Adicionalmente, se reconoce la asesoría académica del Dr. Martín Escamilla-del-Arenal, del QFB Edgar Fernán González Buendía, del Biol. Christian Valdés Quezada, de la M. en C. Selene García de la Cadena, del LCG Rodrigo Arzate-Mejía y del M. en C. Francisco Javier Carmona Aldana.

Finalmente, se agradece a los miembros del jurado, quienes enriquecieron este trabajo:

Dr. Alejandro Manuel García Carrancá	Instituto de Investigaciones Biomédicas
Dr. José Luis Reyes Taboada	Instituto de Biotecnología
Dr. Martín Gustavo Pedraza Alva	Instituto de Biotecnología
Dr. Roberto Coria Ortega	Instituto de Fisiología Celular
Dra. Marcela Lizano Soberón	Instituto de Investigaciones Biomédicas

Agradecimientos Personales

Dedico esta tesis a mi amorosa familia, quienes siempre me han apoyado académicamente, me han inculcado que el trabajo reditúa y me impulsaron desde pequeña a saber que puedo lograr todo lo que me proponga. No hay palabras suficientes para agradecer todo lo que han hecho por mí y agradezco especialmente el que iluminaran mis días durante este periodo. Los amo.

También agradezco al resto de mi familia: a mis abuelitos, primos y tíos. Aunque los viera en menor medida siempre me hicieron sentir su amor, apoyo y todos sus buenos deseos.

Agradezco al laboratorio 122 N por abrirme sus puertas en estos dos años. Muchas gracias Dr. Recillas por su guía, tiempo y sus enseñanzas. A Geito y Fer por ser mis pilares en este periodo, gracias por extenderme su ayuda, su amistad y por darme ánimos cuando más lo necesitaba. Mil gracias a todos mis compañeros que de diversas maneras me apoyaron en este tiempo: Yael, Estela, Carla, Paco, Rodri, Martín, Christiancito, Josué, Erandi, Charis, Edgar, Karin, Adrián y Jesús. Un agradecimiento especial a Giane por todo su trabajo y por brindarme su apoyo. ¡¡Gracias por permitirme conocer a las maravillosas personas que son, por su amistad y por todos sus conocimientos!!

Muchas gracias Andrea Isabel, Rob Ary, Ariann, Sol y Germán por la amistad que surgió de la empatía de compartir tiempos de maestría. También agradezco a mis amigos entrañables Omar Esquinca y a Laura Gómez: aunque nos veamos muy poco es maravilloso reunirnos y ver que sin importar tiempos y distancia la amistad perdura.

Por último, pero no menos importante, agradezco a mi casa de estudios y alma mater, nuestra Universidad Nacional Autónoma de México. Gracias a todos los profesores que me inspiraron desde la licenciatura y en la maestría. Agradezco especialmente a Fede, a David Romero, al Dr. Caballero y al Dr. Coria. También le agradezco a Lety por toda su ayuda en este tiempo. ¡¡¡POR MI RAZA HABLARÁ EL ESPÍRITU!!!

ÍNDICE

ABREVIATURAS	10
RESUMEN	12
INTRODUCCIÓN	13
La Epigenética y sus procesos	13
<i>Metilación del DNA</i>	14
<i>Modificaciones post-traduccionales de las histonas</i>	15
<i>Complejos remodeladores de la cromatina dependientes de ATP</i>	15
<i>Organización y dinámica nuclear</i>	16
<i>Complejo represor Polycomb y activador Trithorax</i>	17
<i>RNAs no-codificantes</i>	18
Generalidades de los RNAs no-codificantes	18
<i>Definición</i>	19
<i>Funcionalidad versus ruido transcripcional</i>	19
<i>Clasificación y tipos de ncRNAs</i>	22
Transcritos largos no-codificantes y sus características generales	25
<i>Clasificación genómica de los lncRNAs</i>	26
<i>Características bioquímicas y estructurales de los lncRNAs</i>	27
<i>Patrones de expresión</i>	29
<i>Conservación de secuencia y evolución de lncRNAs</i>	30
Mecanismos de acción de los lncRNAs	31
<i>Modelos de acción de los lncRNAs</i>	32
Mecanismo de anzuelo	33
<i>Ejemplo: Gas5</i>	33
<i>Ejemplo: H19</i>	34
Mecanismo de andamiaje	35
<i>Ejemplo: Kcnq1ot1</i>	35
<i>Ejemplo: HOTAIR</i>	35
Mecanismo de guía.....	36
<i>Ejemplos: Kcnq1ot1 y HOTAIR</i>	36
Mecanismo de potenciador	37
<i>Acción en cis</i>	38
<i>Acción en trans</i>	40
Transcritos no-codificantes largos y su relación con el cáncer	42
<i>Auto-suficiencia en señales de proliferación</i>	43

<i>Insensibilidad para señales supresoras del crecimiento</i>	45
<i>Evasión de la apoptosis</i>	46
<i>Potencial ilimitado de replicación</i>	47
<i>Angiogénesis sostenida</i>	48
<i>Invasión de tejidos y metástasis</i>	49
ANTECEDENTES PARTICULARES	52
Estudio de perfiles de expresión de lncRNAs desregulados en diversos tipos de cáncer	52
Estudio de asociación de lncRNAs a complejos de regulación epigenética	53
PLANTEAMIENTO DEL PROBLEMA	55
HIPÓTESIS	55
OBJETIVOS	56
Objetivo General	56
Objetivos Particulares	56
MODELO DE ESTUDIO	57
RESULTADOS	59
1. Identificación <i>in silico</i> de un lncRNA con asociación a complejos epigenéticos que muestra niveles de expresión alterados en tejido mamario canceroso	59
1.1 <i>Características de lincRNA_mr</i>	60
1.2 <i>Caracterización in silico de los elementos regulatorios del lincRNA_mr</i>	61
1.3 <i>Predicción computacional de interacción de lincRNA_mr y complejos epigenéticos</i>	64
2. Cuantificación de los niveles de expresión del <i>lincRNA_mr</i> en líneas celulares de tejido mamario	67
2.1 <i>Caracterización de la expresión de las 5 isoformas del lincRNA_mr</i>	68
3. Cuantificación de los genes codificantes aledaños al <i>lincRNA_mr</i>	70
4. Evaluación del gen blanco por medio de ensayos de ganancia y pérdida de función del <i>lincRNA_mr</i>	73
4.1 <i>Ensayos de pérdida de función para abatir al lincRNA_mr</i>	73
4.2 <i>Cuantificación de la expresión del gen <i>cbl-b</i> para determinar mecanismo en cis</i>	74
4.3 <i>Ensayo de ganancia de función al sobre-expresar al lincRNA_mr2 en MCF7</i>	75
4.4 <i>Cuantificación de la expresión del gen <i>cbl-b</i> en un contexto de sobre-expresión</i>	76
5. Evaluación de la fisiología celular por ensayos de pérdida de función	77
5.1 <i>Análisis de división celular y fenotipos celulares</i>	77
5.2 <i>Análisis de TUNEL en un contexto celular con lincRNA_mr abatido</i>	80

DISCUSIÓN..... 82
CONCLUSIONES 86
PERSPECTIVAS..... 87
MATERIALES Y MÉTODOS 89
REFERENCIAS..... 99

ABREVIATURAS

DNA	Ácido desoxirribonucleico
RNA	Ácido ribonucleico
lncRNA	RNA largo no-codificante
lincRNA	RNA intergénico largo no-codificante
sRNA	RNA pequeño no-codificante
mRNA	RNA mensajero
rRNA	RNA ribosomal
cDNA	DNA complementario
gDNA	DNA genómico
ssDNA	DNA de cadena sencilla (single-stranded DNA)
shRNA	short hairpin RNA (RNAs interferentes que forman tallos-asa)
pb	pares de bases
kb	kilobases
nt	Nucleótidos
5mC	Metilación del carbono 5 de la citosina
5hmC	Hidroximetilación del carbono 5 de la citosina
PcG	Grupo Polycomb
PRC2	Complejo represor 2 del grupo Polycomb
TrxG	Grupo Triterax
ENCODE	Enciclopedia de los elementos del DNA
ChIP	Inmunoprecipitación de la cromatina
ChIRP	Aislamiento de la cromatina por purificación de RNA
RIP	Inmunoprecipitación de RNA
seq	Sufijo que indica el uso de tecnologías de secuenciación masiva
CHART	Análisis de captura de blancos de RNA por hibridación
H3K4me1	Monometilación de la lisina 4 en la histona H3
H3K4me3	Trimetilación de la lisina 4 en la histona H3
H3K27ac	Acetilación de la lisina 27 en la histona H3
H3K27me3	Trimetilación de la lisina 27 en la histona H3
RNA Pol II	RNA Polimerasa II
Poli-(A)	Poliadenilación
RNP	Ribonucleoproteína
ER	Receptor de estrógeno

PR	Receptor de progesterona
SAGE	Análisis serial de expresión génica
VV	Vector vacío
KD	Knockdown
rpm	Revoluciones por minuto
ON	Over-night, incubación de toda la noche
Tm	Temperatura media o de alineamiento
TF	Factor de transcripción
TRAIL	Ligando inductor de apoptosis relacionado con TNF

RESUMEN

Los RNAs largos no-codificantes (lncRNAs) se han visto involucrados en distintos procesos celulares y en algunas enfermedades como el cáncer. Los cambios en los niveles de expresión de los lncRNAs pueden llevar a defectos en los niveles de transcripción de sus genes blanco. El objetivo de este estudio es identificar lncRNAs humanos involucrados en la regulación epigenética mediante cambios en los niveles de expresión de genes relacionados con el desarrollo de un proceso tumoral. En particular, centramos nuestros esfuerzos en cáncer de mama. A partir de un grupo de transcritos se escogió un lncRNA candidato al cual denominamos *lincRNA_mr*. Nuestros datos muestran una reducción en los niveles de expresión de este transcrito en la línea celular de cáncer mamario MCF7. Con base en este resultado, nos interesamos en estudiar la función de este lncRNA por estudios de ganancia y pérdida de función. Interesantemente, se observó una correlación positiva entre los niveles de expresión de la isoforma *lincRNA_mr2* y el gen codificante aledaño *cbl-b*. Más aun, cuando se abate la expresión del *lincRNA_mr2* en un contexto normal, se puede ver una disminución del 90% en los niveles del gen *cbl-b* y la sobre-expresión de este RNA largo no-codificante no incrementa la expresión del gen *cbl-b*. Estos resultados indican que *lincRNA_mr* actúa en *cis* sobre el gen *cbl-b*, aunque desconocemos el mecanismo de acción. Basados en literatura que asocia la baja expresión del gen *cbl-b* con la promoción de apoptosis vía extrínseca y con base en experimentos de tiempo de división celular, que mostraron una disminución en el número de células cuando se abate *lincRNA_mr*, se decidió estudiar si existían cambios en la proporción de muerte celular consecuentes a la pérdida de función. Se realizaron ensayos de TUNEL que preliminarmente indican un incremento de núcleos positivos en la línea celular donde se abatió a *lincRNA_mr*. De esta manera, nuestros datos sugieren que *lincRNA_mr* puede estar involucrado en el proceso de muerte celular al regular en *cis* al gen *cbl-b*.

INTRODUCCIÓN

La epigenética gira en torno a la estructuración del genoma en la cromatina, su efecto regulador sobre la expresión de los genes y su participación en múltiples procesos celulares fundamentales (Felsenfeld & Groudine, 2003; Recillas-Targa & Escamilla-del-Arenal, 2004). Se consideran procesos de regulación epigenéticos todos aquellos que influyen en la expresión heredable de los genes sin que ocurran cambios en la secuencia del DNA. Esto es, la información genética que viene codificada en la molécula de DNA no se ve afectada y se transmite su información sin problema alguno (Recillas-Targa & Escamilla-del-Arenal, 2004). En enfermedades humanas, en particular en el cáncer, múltiples estudios han demostrado que defectos epigenéticos y genéticos son detonadores comunes de esta enfermedad (Esteller, 2011). Para poder adentrarse en la asociación entre la epigenética y su influencia en el desarrollo de un proceso tumoral, en el siguiente trabajo se mencionarán los mecanismos epigenéticos, ahondando en los RNAs no-codificantes, además de que se recordarán algunos de los principios que gobiernan la biología del cáncer.

La Epigenética y sus procesos

Cada célula eucarionte contiene información genética en la forma de DNA que se compacta en grados variables dentro de un espacio nuclear confinado (Hake & Allis, 2006). En la especie humana, esta información genética se organiza en 23 pares de cromosomas, que juntos suman aproximadamente tres mil millones de pares de bases nucleotídicas. Si se extendiera la molécula de DNA en los humanos, su longitud total alcanzaría alrededor de dos metros, por lo que existe la necesidad de condensarlo unas 10,000 veces para que pueda residir en el núcleo (Allis, *et al.*, 2007). Aun con su alto grado de compactación, el DNA se empaqueta de tal manera que permite su lectura, replicación y reparación en respuesta a las necesidades celulares y a los estímulos externos (Hake & Allis, 2006).

La compactación del genoma se logra gracias a la estrecha interacción entre el DNA y unas proteínas denominadas histonas, conformando así una estructura fisiológica conocida como cromatina (Bernstein & Allis, 2005; Hake & Allis, 2006). La unidad fundamental de la cromatina es el nucleosoma, el cual está constituido por el octámero de histonas (H2A, H2B, H3 y H4, dos de cada una de las proteínas) alrededor del cual interaccionan 147 pares de bases (bp) (Recillas-Targa & Escamilla-del-Arenal, 2004). Una de las funciones principales de la cromatina consiste en regular la expresión génica por medio de diversos procesos que influyen en los componentes de esta estructura fisiológica, pero no alteran a la secuencia de DNA y que son heredables, es decir, que son epigenéticos (Bernstein & Allis, 2005; Felsenfeld & Groudine,

2003; Recillas-Targa & Escamilla-del-Arenal, 2004). De esta manera, la cromatina provee un medio para dividir al genoma en regiones transcripcionalmente activas o inactivas (Narlikar, *et al.*, 2013).

Dentro de los procesos epigenéticos que más correlacionan con la actividad génica y que se han estudiado mejor resaltan: la metilación/desmetilación del DNA, las modificaciones post-traduccionales de las histonas, los complejos remodeladores de la cromatina dependientes de ATP, la organización/dinámica nuclear y los complejos represor y activador (Polycomb y Trithorax) (Felsenfeld & Groudine, 2003). Sin embargo, en años recientes también se ha observado la acción coordinada de RNAs no-codificantes modulando la estructura de la cromatina (Bernstein & Allis, 2005). A continuación se describe brevemente cada uno de estos procesos epigenéticos con el fin de introducirnos al tema central de esta tesis que tiene que ver con los RNAs no-codificantes.

Metilación del DNA

La metilación del DNA es una de las modificaciones más estudiadas en los últimos años y desempeña un papel central en la transmisión de la información epigenética. Esta modificación consiste en la adición de un grupo metilo a la posición 5' de la citosina del DNA cuando se encuentra en el contexto del dinucleótido CpG (Allis, Jenuwein, & Reinberg, 2007; Felsenfeld & Groudine, 2003; Recillas-Targa & Escamilla-del-Arenal, 2004). La distribución del DNA metilado a lo largo del genoma muestra enriquecimiento en las regiones no codificantes y en elementos repetidos, pero no en las islas CpG de los genes activos (Allis, *et al.*, 2007). Siguiendo con lo anterior, la metilación del DNA ayuda a reclutar proteínas específicas que reprimen la expresión génica y compactan a la cromatina (Felsenfeld & Groudine, 2003). Además, esta modificación participa en el silenciamiento de algunos genes de manera permanente en distintas etapas del desarrollo. Por ejemplo, al momento del nacimiento se lleva a cabo una fuerte oleada de metilación que contribuye al “apagado” de genes embrionarios que no serán expresados en la etapa adulta de un organismo (Recillas-Targa & Escamilla-del-Arenal, 2004).

En años recientes, se ha acumulado evidencia que sugiere que la metilación del DNA puede revertirse en las células de mamíferos. Uno de los indicios más fuertes es el descubrimiento de la 5-hidroximetilación de la citosina (5hmC). La ruta de desmetilación del DNA comienza cuando la 5mC es hidroxilada por las enzimas TET para su remoción, generando a la 5hmC como intermediario. La 5hmC es deaminada, resultando en la producción de 5 hidroximetiluracilo el cual es procesado por glicosilasas de DNA que forman parte de la maquinaria de reparación por escisión de bases. El producto de esta serie de reacciones es el

reemplazo de la citosina y la desmetilación del DNA. Cabe resaltar que la 5hmC es especialmente abundante en tejidos como el cerebro y las células embrionarias pluripotentes (Bhutani, *et al.*, 2011).

En resumen, la metilación del DNA contribuye a mantener la estabilidad del genoma, además de participar en procesos celulares importantes como la impronta genética y la inactivación del cromosoma X (Bhutani, *et al.*, 2011; Recillas-Targa & Escamilla-del-Arenal, 2004). A su vez, es importante mencionar que en estos ejemplos mencionados existe una interdependencia entre los diversos procesos epigenéticos, como entre la metilación del DNA y las modificaciones covalentes de las histonas, de las cuales se hablará a continuación (Recillas-Targa & Escamilla-del-Arenal, 2004).

Modificaciones post-traduccionales de las histonas

Los nucleosomas pueden ser modificados químicamente tanto en la región amino-terminal de las histonas (que sobresalen de la superficie globular) como dentro del cuerpo del octámero (Felsenfeld & Groudine, 2003). Algunas de las moléculas que se adicionan a diversos residuos de lisina, arginina o serina son pequeños grupos químicos (acetilación, metilación, fosforilación) o péptidos de mayor tamaño (ubiquitinación y sumoilación) (Allis, *et al.*, 2007). Estas modificaciones covalentes sirven principalmente para controlar la estructura de la cromatina, permitir o interrumpir la unión específica de proteínas y para contribuir a los diversos patrones de expresión génica, reparación de DNA o replicación (Allis, *et al.*, 2007; Felsenfeld & Groudine, 2003). Las marcas de las histonas pueden afectarse entre sí y pueden estar positiva o negativamente correlacionadas, favoreciendo la inter-comunicación entre nucleosomas para regular a la cromatina (Felsenfeld & Groudine, 2003; Recillas-Targa & Escamilla-del-Arenal, 2004). Además, la estrecha relación entre las modificaciones de las histonas y los complejos remodeladores de la cromatina dependientes de ATP permite esculpir el paisaje de la cromatina en una escala genómica (Narlikar, *et al.*, 2013).

Complejos remodeladores de la cromatina dependientes de ATP

Consiste en una serie de complejos macro-moleculares cuya función depende de la energía que produce la hidrólisis de ATP para regular la estructura de cromatina. Una de sus acciones consiste en ayudar a exponer secuencias de DNA que permiten la unión de factores transcripcionales, ya sea por medio del desplazamiento o por el re-posicionamiento de los nucleosomas (Recillas-Targa & Escamilla-del-Arenal, 2004). Existen tres modelos que proponen explicar el funcionamiento de estos complejos. Uno indica que estos complejos tienen la

capacidad de movilizar a los nucleosomas por deslizamiento del octámero a través de la molécula de DNA. Los desplazamientos ocurren a distancias cortas y sin que los nucleosomas se disocien por completo de sus interacciones con el DNA (Recillas-Targa & Escamilla-del-Arenal, 2004). El segundo modelo tiene que ver con un cambio conformacional del propio nucleosoma, mientras que el tercer modelo propone cambios en la composición de las histonas canónicas por variantes específicas de las mismas (Allis, *et al.*, 2007; Narlikar, *et al.*, 2013; Recillas-Targa & Escamilla-del-Arenal, 2004).

Estos complejos dependientes de ATP, en coordinación con modificaciones post-traduccionales de las histonas, proveen una manera de modificar rápidamente la estructura cromatínica que se requiere para eventos transcripcionales, de replicación del DNA, de reparación y para el establecimiento, mantenimiento y la organización de estados de la cromatina específicos en el compartimento nuclear (Narlikar, *et al.*, 2013).

Organización y dinámica nuclear

Existe evidencia creciente que sugiere que el arreglo de los cromosomas, genes, loci y cuerpos nucleares no es aleatorio y que presenta características de auto-organización en espacio y tiempo. La organización nuclear depende de una variedad de señales celulares influenciadas por el ambiente, el estado del desarrollo y el ciclo celular (Rajapakse & Groudine, 2011). En metazoarios, la cromatina se organiza en regiones compactas (heterocromatina) y zonas más abiertas (eucromatina) (Felsenfeld & Groudine, 2003).

Uno de los principales elementos estructurales en la organización de la cromatina dentro del núcleo es la lámina nuclear, que es una red de filamentos intermedios situada en la superficie nucleo-plásmica de la membrana interna nuclear (Peric-Hupkes, *et al.*, 2010). Existe evidencia de que la lámina nuclear se involucra directamente en la represión génica además de anclar a cromatina transcripcionalmente inactiva (Peric-Hupkes, *et al.*, 2010; Rajapakse & Groudine, 2011). De esta manera, las interacciones entre la lámina y el genoma se involucran en el control de los programas de expresión génica durante la diferenciación de linajes celulares (Peric-Hupkes, *et al.*, 2010).

Por el contrario, los loci que muestran transcripción activa (eucromatina) se localizan cercanos a los poros nucleares así como en compartimentos nucleares internos específicos. De esta manera, la organización espacial de los cromosomas representa un factor importante en la regulación génica y la estabilidad del genoma (Rajapakse & Groudine, 2011). Diversos estudios han reportado que los cromosomas ocupan distintos territorios bien definidos al interior del núcleo eucarionte en interfase. Estos territorios son evidentes en múltiples tipos celulares y

parecen estar conservados a lo largo de la evolución, por lo que constituyen una característica fundamental de la arquitectura nuclear (Rajapakse & Groudine, 2011; Recillas-Targa & Escamilla-del-Arenal, 2004). Asimismo, se pueden formar interacciones inter-cromosomales de largo alcance entre múltiples loci (Rajapakse & Groudine, 2011). Un complejo proteico que ayuda en este tipo de estructuración es el grupo Polycomb, que juega adicionalmente otros papeles cruciales en la regulación epigenética junto con otro complejo con funciones antagónicas, Trithorax.

Complejo represor Polycomb y activador Trithorax

Los grupos de genes Polycomb (PcG) y Trithorax (TrxG) actúan en conjunto para regular la dinámica de la cromatina y las cascadas de expresión génica, reteniendo así la memoria celular a lo largo de la vida de un organismo (Mills, 2010). La función de PcG consiste en mediar el silenciamiento génico por medio de la regulación de la estructura de la cromatina, mientras que TrxG permite la activación transcripcional (Mills, 2010; Recillas-Targa & Escamilla-del-Arenal, 2004). Ambos complejos se encuentran evolutivamente conservados (Mills, 2010).

Los componentes del grupo TrxG se puede clasificar en: los modificadores de histonas, encargados de depositar la trimetilación de la lisina 4 en la histona 3 (H3K4me3); los remodeladores de los nucleosomas dependientes de ATP (incluyen proteínas que pueden leer las marcas de metilación en las histonas) y las proteínas que se unen específicamente a secuencias de DNA. La mayoría de estas proteínas ejercen sus funciones como parte de complejos multi-proteicos en los cuales la suma de sus actividades es cooperativa para la regulación de la expresión génica (Schuettengruber, *et al.*, 2011).

La represión mediada por PcG es dinámica y el estado de silenciamiento se mantiene de un ciclo celular al siguiente (Schwartz & Pirrotta, 2013). En mamíferos, existen dos complejos principales de PcG: el complejo represor Polycomb 2 y 1 (PRC2 y PRC1) (Margueron & Reinberg, 2011). PRC2 inicia la compactación de la cromatina al catalizar la deposición de la H3K27me3 (Schwartz & Pirrotta, 2013). Las subunidades canónicas que funcionan como holoenzima de PRC2 están conservadas desde *Drosophila melanogaster* hasta mamíferos y son EZH1/2, SUZ12, EED y RbAp46/48. Sin embargo, se ha observado que existen componentes adicionales con contribuciones aditivas para maximizar la actividad de la holoenzima y que participan en el reclutamiento de PRC2, incluyendo RNAs no-codificantes (Margueron & Reinberg, 2011). PRC1 cataliza la mono-ubiquitinación de la histona H2A, contribuyendo a esta compactación (Margueron & Reinberg, 2011).

Una característica importante para TrxG y PcG es la capacidad que poseen sus diversos componentes para asociarse con moléculas como DNA, modificaciones post-traduccionales de las histonas y RNAs no-codificantes (ncRNA). Esta característica les confiere una especificidad para mantener patrones de expresión específicos en el genoma, modulando así la estructura de la cromatina (Peric-Hupkes, *et al.*, 2010; Ponjavic, *et al.*, 2009).

RNAs no-codificantes

Existen cada vez más estudios que reportan de manera consistente que la mayor parte del genoma (~70%) se transcribe en mamíferos (Lee, 2012). Sin embargo, en un estudio reciente llevado a cabo por el consorcio ENCODE, se concluyó que solamente el ~1.2% del genoma humano es codificante (Cheetham, *et al.*, 2013). El resultado es la producción dinámica y diferencial de un amplio y complejo rango de RNAs no-codificantes (Mercer & Mattick, 2013). Esto demuestra que la transcripción no sólo se limita a regiones que se traducen en proteínas, sino que es constitutiva a lo largo del genoma, como se observa con proyectos de gran escala y librerías genómicas (Wilusz, *et al.*, 2009).

Los estudios apuntan a que los ncRNAs constituyen un nivel de regulación génica crítico en los organismos complejos (Mattick, 2005). Más aún, existe evidencia de que los ncRNAs juegan un papel significativo en el control de la regulación epigenética, la dinámica nuclear y las interacciones de largo alcance. Se ha reportado que estos transcritos, comenzados a estudiar recientemente, tienen un impacto en procesos celulares normales de un organismo como la diferenciación y el desarrollo (Bernstein & Allis, 2005). Debido a la relevancia que se les ha atribuido a estos transcritos en distintos mecanismos epigenéticos, nuestro grupo de trabajo comenzó a interesarse en estudiar ncRNAs humanos involucrados en la regulación epigenética. A continuación se discutirá más acerca de los ncRNAs, su clasificación y su relevancia biológica.

Generalidades de los RNAs no-codificantes

Las regiones no-codificantes del genoma sirven no sólo como sustrato para las proteínas con unión a DNA -que a su vez gobiernan la expresión génica y la arquitectura tridimensional de la cromatina- sino que funcionan también de templado para la transcripción de un vasto número de RNAs no-codificantes que exhiben patrones de expresión altamente específicos en tiempo y espacio (Cheetham, *et al.*, 2013).

Definición

El término RNA no-codificante se emplea comúnmente para referirse a aquellas moléculas de RNA que pueden ser transcritas pero que no sirven de templado para la síntesis proteica (Guttman & Rinn, 2012; Mattick & Makunin, 2006; Rinn & Chang, 2012). Cabe resaltar que esto no significa que los ncRNAs no contengan información o no posean una función. A estos transcritos se les ha establecido una nomenclatura y clasificación gracias a los esfuerzos de organizaciones encargadas de descifrar a los elementos regulatorios del genoma como ENCODE y FANTOM.

El potencial codificante comúnmente es calculado por medio de algoritmos bioinformáticos que comparan la secuencia de genes proteicos conocidos y de los posibles ncRNAs, además de secuencias de transcritos no-codificantes que han sido validados experimentalmente (Derrien, *et al.*, 2012). Otra estrategia adoptada en algunos estudios es evaluar los posibles marcos de lectura (ORF) de un ncRNA y hacer búsquedas de homología (con BLASTX) a lo largo de bases de datos que contengan familias y dominios proteicos, como PFAM o Swissprot. Asimismo, para descartar que los ncRNAs contengan secuencias codificantes para proteínas nuevas (que recientemente hayan evolucionado) o péptidos pequeños, se pueden hacer análisis de sustitución de frecuencia de codones (Rinn & Chang, 2012). Lo que se ha observado es que los ncRNAs tienen un potencial codificante similar entre sí, pero muy distinto al de los transcritos proteicos. Aun cuando algunos transcritos no-codificantes poseen pequeños ORFs la calidad esperada del péptido no es mayor a la de una secuencia escogida al azar (Derrien, *et al.*, 2012; Guttman & Rinn, 2012).

Un método experimental para validar el potencial codificante de un ncRNA es por el perfil de expresión ribosomal, que consiste en medir la traducción *in vivo* e identificar posibles RNAs que se unen y son traducidos por el ribosoma. Otra ventaja de este método es que, si se reportan ncRNAs unidos, se sabría qué región del transcrito es ocupada, permitiendo continuar la búsqueda de potencial codificante con análisis computacionales sobre esa secuencia acotada. En el caso de que una porción de los ncRNAs contenga un péptido pequeño, se considera que esto no excluye la potencial doble naturaleza del RNA, indicando que podría funcionar tanto como un transcrito como por medio de sus productos proteicos (Rinn & Chang, 2012).

Funcionalidad versus ruido transcripcional

El papel central del RNA en el flujo de información genética vino a la luz en 1961 con el artículo de Jacob y Monod titulado “*Mecanismos regulatorios genéticos en la síntesis de proteínas*”,

donde se estableció el concepto de RNAs mensajeros (mRNAs) (Saldaña-Meyer, *et al.*, 2014). Hoy en día es muy claro que el RNA se emplea ampliamente en la regulación génica, como lo postularon estos científicos hace más de 50 años, pero ahora también es evidente que el RNA posee un gran rango de funciones biológicas en los eucariontes (Mattick & Makunin, 2006).

Cuando se observó que existen miles de ncRNAs en mamíferos y que la mayoría del genoma se transcribe, la visión céntrica de las proteínas se vio confrontada, ya que en ese entonces se consideraba a los genes como sinónimos de proteínas (Mattick & Makunin, 2006; Taft & Mattick, 2003). Esto abrió dos grandes posibilidades: que el conjunto de transcritos que no eran mRNAs fueran sólo residuos acumulados de la evolución temprana de los genes, es decir “basura o ruido transcripcional”, o que estos transcritos contuvieran información en forma de RNA que es importante para el desarrollo y los procesos celulares de los organismos superiores (Mattick & Makunin, 2006).

Se sabía desde tiempo atrás que los RNAs infraestructurales –aquellos RNAs constitutivos que confieren especificidad de secuencia y actividad catalítica (rRNAs, tRNAs y snRNAs)- llevan a cabo una serie de funciones, incluyendo el procesamiento, la edición del RNA, la traducción o la destrucción de mRNAs (Mattick & Makunin, 2006). También parecía que el RNA jugaba un papel virtualmente en todos los niveles de regulación génica (Mattick, 2005). El descubrimiento de que la relación entre las regiones no-codificantes del genoma y el DNA genómico total correlaciona con la complejidad biológica de los organismos (aun cuando se toma en cuenta la diversificación de las proteínas tanto por procesamiento alternativo como por regulación post-traducciona) refleja la importancia de estas secuencias en la evolución (Fig. 1) (Taft & Mattick, 2003). Esto sugiere que los mecanismos regulatorios basados en RNA tienen un papel relevante en la evolución de la complejidad en los eucariontes (Fatica & Bozzoni, 2014). Esta visión se ha fortalecido gracias a los múltiples estudios realizados en años recientes, que han comenzado a demostrar la importancia y abundancia de los ncRNAs (Mattick, 2005; Mattick & Makunin, 2006; Taft & Mattick, 2003).

Actualmente, existe evidencia obtenida de una variedad de organismos modelo que demuestra que múltiples ncRNAs juegan un papel significativo en el control de la regulación epigenética, la dinámica cromosomal, las interacciones de largo alcance y en fenómenos que tienen un impacto en la diferenciación celular y desarrollo de un organismo (Bernstein & Allis, 2005). Es decir, que muchos ncRNAs cuentan con funciones cruciales en distintos procesos celulares (Cabili, *et al.*, 2011; Esteller, 2011; Gutschner & Diederichs, 2012; Guttman & Rinn, 2012; Moran, *et al.*, 2012; Ulitsky, *et al.*, 2011).

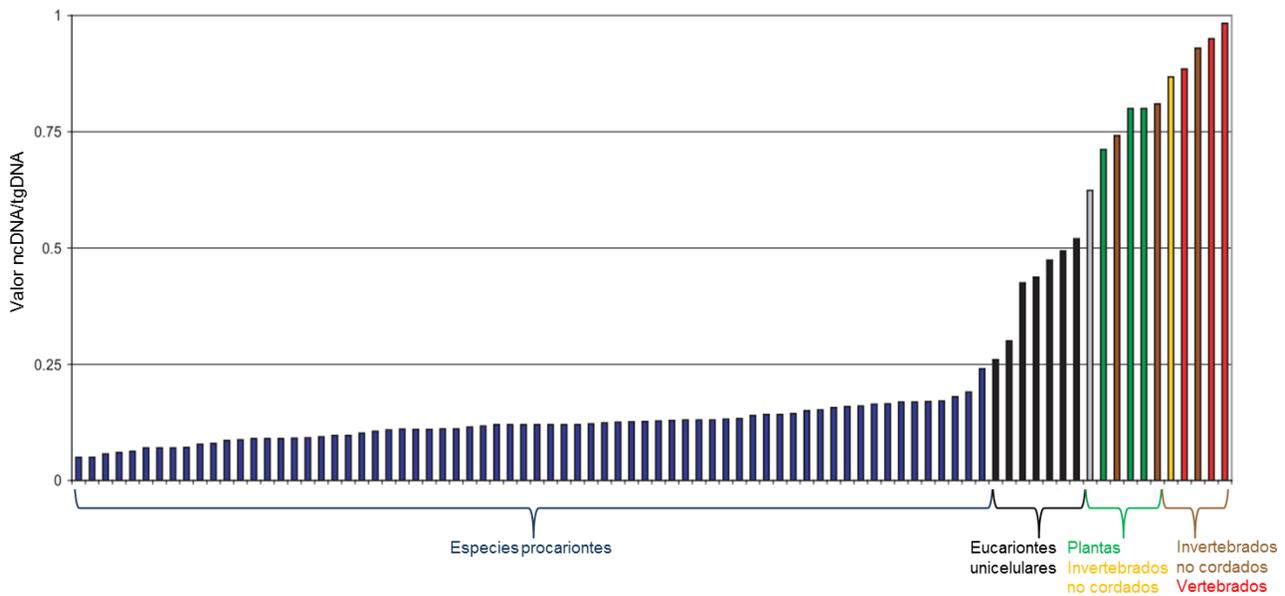


Figura 1: Tasa de ncDNA/tgDNA reportada para diversos organismos. Existe una correlación positiva entre la complejidad biológica de un organismo y la tasa de DNA no-codificante (ncDNA) dado su DNA genómico total (tgDNA) o ncDNA/tgDNA. Dicha tasa alcanza su valor más alto para la especie humana. Modificado de Taft & Mattick, 2003.

Existen también estudios novedosos que demuestran por evidencia genética *in vivo* que algunos de estos transcritos no sólo son funcionales sino que, si se abaten por completo en el contexto de un organismo en desarrollo, pueden ser responsables de letalidad temprana (Sauvageau, *et al.*, 2014). De esta manera, se puede ver el papel de algunos ncRNAs en una versión modificada del “Dogma central de la biología molecular” postulado por Francis Crick, donde la información genética codificada en el DNA y transcrita a RNA puede dirigirse tanto a proteína como permanecer en forma de RNA regulatorio, para influir en la transcripción o traducción de otros genes codificantes o no-codificantes (Fig. 2) (Wahlestedt, 2013).

Debido a la posible funcionalidad que confieren estos ncRNAs, existe un creciente interés en su identificación, clasificación y estudio. Más aún, existen reportes de ncRNAs que muestran un cambio en sus niveles de expresión entre estados normales y alterados, abriendo la posibilidad de que sean un elemento clave en la patología de diversas enfermedades. Un ejemplo es en el cáncer, donde estos transcritos se han visto fuertemente implicados (Gutschner & Diederichs, 2012; Mitra, *et al.* 2012; Qiu, *et al.*, 2013; Rinn & Chang, 2012). En el presente trabajo de tesis se podrá ahondar más en cómo los ncRNAs y sus cambios en niveles de expresión en un contexto de cáncer pueden desencadenar en consecuencias en la biología de la célula.

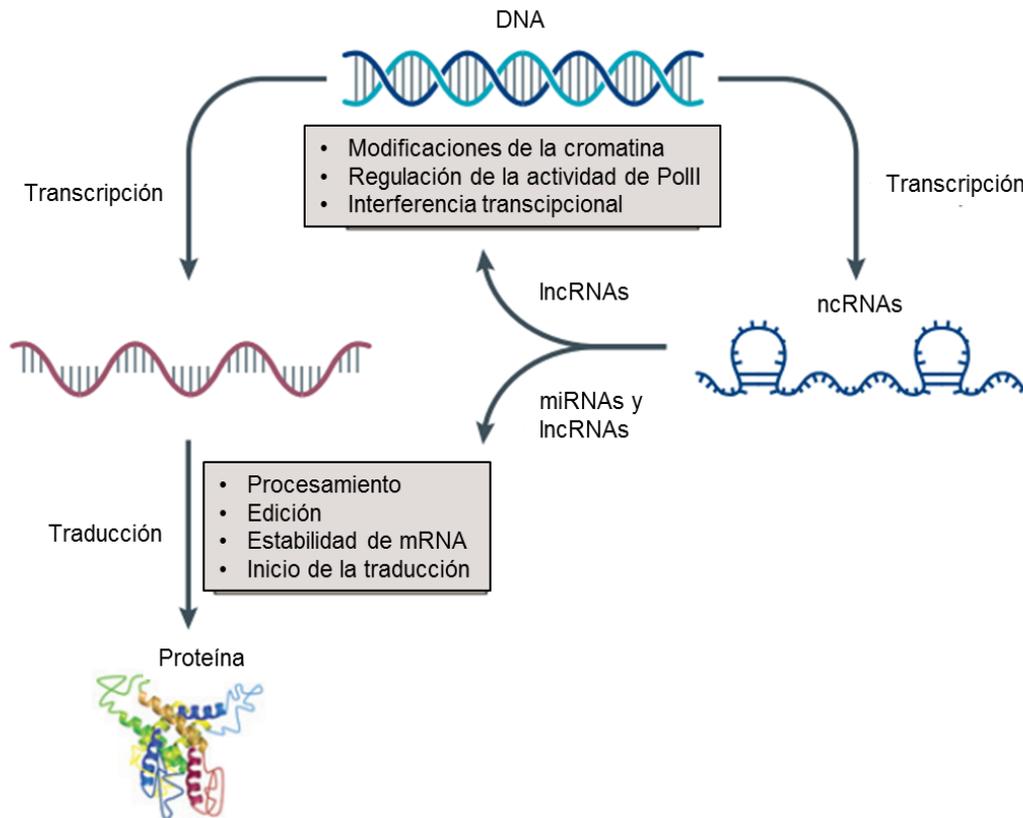


Figura 2: Dogma central de la biología molecular en el contexto de ncRNAs regulatorios. El “dogma central”, propuesto por Francis Crick, postula que “La transferencia de información de un ácido nucleico a otro ácido nucleico, o de un ácido nucleico a una proteína es posible, pero la transferencia de una proteína a una proteína o de una proteína a un ácido nucleico es imposible” (Hanahan & Weinberg, 2000). Esta visión se ha visto desafiada recientemente al descubrirse el universo de los RNAs no-codificantes; por tal razón, se propone incorporar a estas moléculas en el dogma central, donde los ncRNAs pueden actuar de diversas maneras. Modificado de Wahlestedt, 2013.

Clasificación y tipos de ncRNAs

Uno de los objetivos del consorcio ENCODE consiste en estudiar cuál es la proporción del genoma humano que es transcrita; con este fin, analizaron la expresión global en múltiples líneas celulares. Según las publicaciones más recientes de esta investigación, si se consideran los transcritos de todas las líneas celulares estudiadas, se reporta que el ~75% del genoma codifica para transcritos primarios; sin embargo, en ninguna línea celular aislada se expresa más del 56.7% (Djebali, *et al.*, 2012). Los transcritos no-codificantes se han clasificado en dos categorías dependiendo de la longitud de su secuencia. Si son menores a 200 nucleótidos (nt) se denominan RNAs pequeños no-codificantes (sRNAs) y si poseen más de 200 nt (o hasta

algunas kilobases, kb) se catalogan como RNAs largos no-codificantes (lncRNA) (Derrien, *et al.*, 2012; Djebali, *et al.*, 2012; Guttman & Rinn, 2012; Guttman, *et al.*, 2009).

La versión más reciente de la base de datos GENCODE reporta que existen en total 7,053 sRNAs anotados, de los cuales el 85% se puede dividir en cuatro clases: 1944 son RNAs pequeños nucleares (snRNAs), 1521 son pequeños nucleolares (snoRNAs), 1756 son microRNAs (miRNAs) y 624 son RNAs de transferencia (tRNAs) (Djebali, *et al.*, 2012). Los snRNAs, tRNAs y snoRNAs son RNAs infraestructurales que fueron descubiertos tiempo atrás y cuyas funciones están bien establecidas. Tanto el procesamiento como la traducción requieren de estos transcritos para reconocer secuencias específicas en los sustratos de RNA (Mattick & Makunin, 2006).

La relevancia de la función que ejercen los sRNAs es particularmente notoria en los miRNAs, que son secuencias de 19-24 nt que participan en el silenciamiento post-transcripcional del 60% de los genes codificantes, al controlar la traducción de un mRNA hacia proteínas por un apareamiento imperfecto (Esteller, 2011; Mattick & Makunin, 2006). El sitio de reconocimiento del miRNA en el mRNA, conocido como semilla, consiste en tan solo 6-8 nt y comúnmente se localiza en el 3'UTR de los mRNAs. La biogénesis de los miRNAs es un proceso que involucra a las enzimas Drosha y Dicer, que son RNAsas III, resultando en la producción de miRNAs maduros de ~22 nt. El complejo Dicer carga a estas moléculas a un miembro de la subfamilia de proteínas Argonauta, para formar un complejo silenciador inducido por RNA (RISC). RISC dirige la regulación del mRNA al reconocer a la secuencia semilla y la acción del miRNA impide la traducción del mRNA ya sea por degradación del mensajero o por inhibición del inicio de la traducción (Esteller, 2011).

En tanto algunos miRNAs regulan blancos individuales de forma específica, otros pueden funcionar como reguladores maestros de un proceso, regulando cientos de genes simultáneamente (Esteller, 2011). Un ejemplo del segundo caso es la familia de miRNAs *let-7*, que fue encontrada en *C. elegans* y que se encuentra ampliamente conservado incluso en humanos, donde fue la primera familia de miRNAs en identificarse y se conforma por 13 miembros. En este nemátodo, *let-7* resulta un gen esencial, ya que si se muta no permite la transición correcta del estadio larvario 4 hacia el gusano adulto, provocando su muerte (de ahí su nombre: "lethal-7"). En mamíferos, se ha reportado que los niveles de *let-7* aumentan durante la embriogénesis y el desarrollo cerebral; algunos de los procesos en los que se ha asociado a estos miRNAs incluyen el desarrollo, la formación muscular, la adhesión celular y la regulación génica. Se ha observado que existen niveles reducidos de *let-7* en múltiples tipos de

cáncer, debiéndose quizá en parte a que este miRNA guarda un circuito de retroalimentación negativo con Dicer (Boyerinas, *et al.*, 2010).

Otra clase de sRNA que se ha visto desregulada en cáncer es la de los RNAs que interactúan con las proteínas PIWI (piRNAs). Estos RNAs pequeños miden 24-30nt, son independientes de Dicer y se unen a la PIWI, que es una subfamilia de las proteínas Argonata involucrada en el mantenimiento de la estabilidad del genoma en las células germinales. Los piRNAs se transcriben a partir de regiones del genoma que contienen elementos transponibles expresados y de otras secuencias repetidas. El complejo que se forma entre los piRNAs y las proteínas PIWI suprime a los elementos transponibles expresados y su movilización. Se ha encontrado que existen piRNAs involucrados en la tumorigénesis de tejido testicular y de otros tipos de tumores, aunque se desconoce participación específica (Esteller, 2011).

Los snoRNAs, que aunque se clasifican como sRNA suelen medir 60-300 bp, son otro ejemplo de ncRNAs asociados al cáncer. Estos RNAs residen en el nucleolo y son componentes de ribonucleoproteínas (RNPs) responsables de marcar al RNA ribosomal (rRNA) con la 2'-O-metilación y la pseudouridilación, facilitando el plegamiento y estabilidad de los rRNAs (Esteller, 2011). En vertebrados, la mayoría de los snoRNAs están codificados en intrones y existe un subgrupo de lncRNAs que contiene a estos RNAs pequeños (Askarian-Amiri, *et al.*, 2011). El primer indicio de que los snoRNAs tenían papeles potenciales en el cáncer surgió del reporte de una disminución sustancial en meningiomas, seguido de otros datos que mostraban desregulación en cáncer pulmonar, de próstata y de mama (Esteller, 2011). Asimismo, un estudio mostró que el lncRNA *Zfas1*, localizado dentro de los ductos y alveolos de la glándula mamaria, pero con baja expresión en contexto tumoral, contiene tres snoRNAs (Snord12, Snord12b y Snord12c). Interesantemente, se demostró que *Zfas1* tiene asociación con la proliferación celular y la diferenciación (Askarian-Amiri, *et al.*, 2011). Se piensa que los snoRNAs contribuyen a la tumorigénesis por un efecto en los ribosomas y la traducción proteica, que comúnmente se perturba en células cancerosas (Esteller, 2011).

Al parecer, hallar a dos clases de ncRNAs juntos, como se menciona en el lncRNA *Zfas1*, es una característica aparentemente más general de lo que se pensaba. El estudio de GENCODE reporta que el 27% de los sRNAs anotados reside dentro del 8% de los genes codificantes y el 5% de estos transcritos pequeños en el 3% de los lncRNAs (Djebali, *et al.*, 2012). Un análisis más a fondo reveló que los exones de los lncRNAs están enriquecidos para todas las clases de sRNAs, excepto para snRNAs, cuando se compara con otros dominios genómicos incluyendo a los intrones de los lncRNAs. Existe un enriquecimiento de snoRNAs en los exones de los lncRNAs, aunque estos RNAs pequeños también se pueden generar de

intrones de los RNAs largos no-codificantes (Derrien, *et al.*, 2012). Debido a que también se reporta que el 6% de los lncRNAs contienen sRNAs, parece ser que en algunas ocasiones los lncRNAs son precursores de los RNAs pequeños (Djebali, *et al.*, 2012).

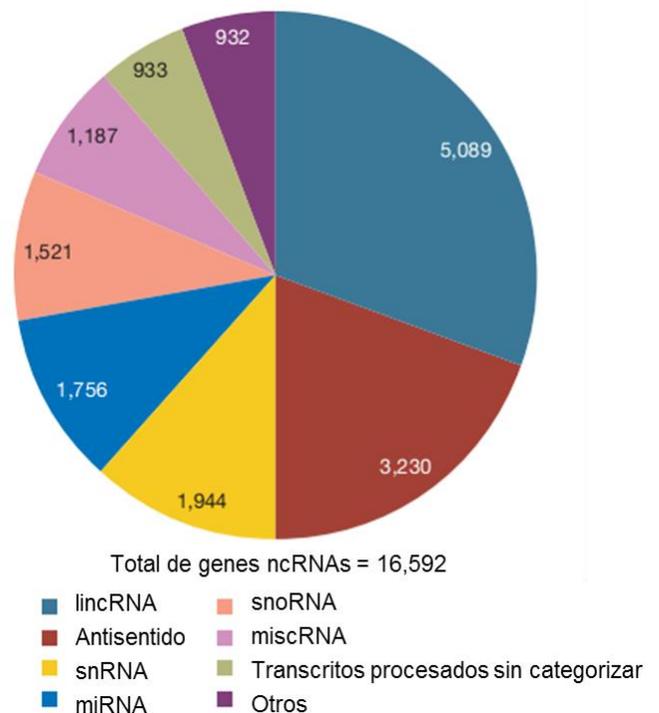
En particular, en nuestro grupo de investigación nos hemos interesado en la función reguladora a nivel epigenético de los RNAs no-codificantes largos, que día a día atraen más atención al involucrarse en casi todas las etapas de la expresión génica, así como en diversas funciones nucleares y al reportarse ejemplos que los asocian con la carcinogénesis (Gutschner & Diederichs, 2012; Mercer & Mattick, 2013).

Transcritos largos no-codificantes y sus características generales

El descubrimiento y estudio de los lncRNAs es de gran relevancia para la biología humana, ya que representan un componente funcional extenso y poco explorado del genoma. Al momento se han predicho miles de transcritos que miden más de 200 nt y tienen un potencial codificante bajo o nulo. Las cifras de lncRNAs registrados en la base de datos de GENCODE, que cuenta con transcritos anotados manualmente, reportan 14,880 lncRNAs que se originan de 9,277 loci (Derrien, *et al.*, 2012). La proporción de los lncRNAs es mayor a la de los otros tipos de ncRNAs; lo cual se puede apreciar en la gráfica realizada preliminarmente por el proyecto Havana (Fig. 3) (Baker, 2011).

Los lncRNAs se han visto implicados en la regulación de procesos biológicos importantes que mantienen las funciones normales de una célula (Mitra, *et al.*, 2012). Algunos ejemplos de estos procesos son el ciclo celular, la regulación transcripcional, la compensación por dosis cromosómica y el tráfico nuclear/citoplásmico entre otros (Huarte, *et al.*, 2010; Tsai, *et al.*, 2010; Wapinski & Chang, 2011). Sumado a esto, los transcritos largos no-codificantes juegan un papel integral en el control del desarrollo, división y diferenciación celular (Cheetham, *et al.*, 2013).

Figura 3: Proporción de cada tipo de RNA no-codificante. En esta figura se muestra la distribución de diversos ncRNAs, según datos preliminares del proyecto Havana-GENCODE 7. Modificado de Baker, 2011.



Clasificación genómica de los lncRNAs

La organización de los loci de lncRNAs se encuentra frecuentemente entrelazada con la de los genes codificantes, lo que añade un grado de complejidad en la comprensión de la estructura, función y evolución del genoma (Fatica & Bozzoni, 2014). Los lncRNAs se han sub-clasificado de acuerdo a su localización genómica como RNAs antisentido (con transcritos que intersectan cualquier exón de un gen codificante en la cadena opuesta), lncRNAs intergénicos (lincRNAs) y lncRNAs génicos (que intersectan con genes codificantes). Para estos últimos existe a su vez una sub-clasificación según la forma en que los lncRNAs coinciden con las secuencias codificantes, pudiendo ser exónicos, intrónicos e intersectados. La distribución reportada en GENCODE 7 es 3,233 antisentido, 9,518 lincRNAs y 5,362 lncRNA génicos -2,411 exónicos, 2,784 intrónicos y 167 intersectados- como se observa en la figura 4 (Derrien, *et al.*, 2012).

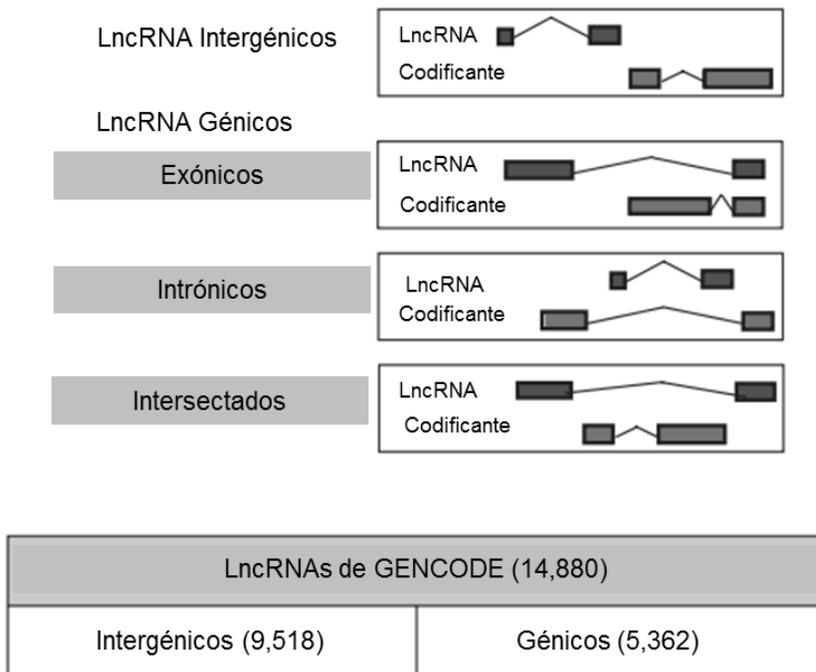


Figura 4: Anotación manual de los diversos lncRNAs en el genoma humano. En la parte de arriba se muestra la clasificación de los lncRNAs en el genoma humano basada en la intersección con los genes codificantes para proteínas. La tabla muestra el número de transcritos de cada clase encontrados en esta anotación manual. Nótese que los transcritos que intersectaron con exones codificantes en la misma cadena fueron omitidos. Modificado de Derrien *et al.*, 2012.

Para muchos lncRNAs, la localización genómica ha mostrado ser clave para su mecanismo de acción. Por ejemplo, cuando los promotores de un lncRNA y un gen codificante se encuentran cercanos en el cromosoma, el inicio de un evento transcripcional para estos genes podría estar acoplado (Ponjavic, *et al.*, 2209). De igual forma, en el caso de los RNAs antisentido, su arreglo genómico inmediatamente indica que podrían estar actuando uno sobre otro (Pelechano & Steinmetz, 2013).

Características bioquímicas y estructurales de los lncRNAs

La mayoría de los lncRNAs caracterizados se generan por la misma maquinaria transcripcional que los mRNAs, evidenciado por su transcripción a partir de la RNA Pol II (Mercer & Mattick, 2013). También se conoce que las modificaciones post-traduccionales de las histonas participan en la regulación de la expresión génica y recientemente, las técnicas como la inmunoprecipitación de la cromatina seguido de secuenciación masiva (ChIP-seq), han permitido generar mapas genómicos de la arquitectura de la cromatina (Rinn & Chang, 2012).

Los patrones de marcas de histonas en el promotor y sitio de inicio de la transcripción de los lncRNAs guardan perfiles similares a los observados en los genes codificantes cuando existe transcripción activa, con las modificaciones H3K4me2, H3K4me3, H3K9ac y H3K27ac. También poseen H3K36me3 en el cuerpo del gen, propia de elongación transcripcional (Derrien, *et al.*, 2012; Wapinski & Chang, 2011). Además, si los lncRNAs están silenciados presentan H3K27me3, modificación asociada a represión por PcG (Mercer & Mattick, 2013). En particular, las marcas H3K4me3 y H3K36me3 –también conocidas como firma K4-K36- se han considerado como firmas indicativas de actividad transcripcional y se ha tomado ventaja de estas modificaciones durante la identificación de lncRNAs en diversos organismos (Guttman, *et al.*, 2009; Khalil, *et al.*, 2009; Nam & Bartel, 2012; Rinn & Chang, 2012; Ulitsky, *et al.*, 2011). En la actualidad existen catálogos de lncRNAs para organismos tan diversos como humano (Derrien, *et al.*, 2012), ratón (Guttman, *et al.*, 2009), el nemátodo *C. elegans* (Nam & Bartel, 2012) y el pez cebra (Ulitsky, *et al.*, 2011).

Otras características estructurales de los lncRNAs es que, al igual que los mRNAs, poseen 7-metilguanosina tri-fosfato en el extremo 5', también conocida como 5' CAP. Aunque no es una generalidad, al menos el 39% de los lncRNAs se poliadenilan (poli-(A)). Esto se observó cuando se analizaron las 100 bp que rodean al extremo 3' de estos transcritos para buscar la presencia de alguno de los 6 motivos más comunes que indicaran poli-(A) (Derrien, *et al.*, 2012; Mercer & Mattick, 2013; Moran, *et al.*, 2012).

El 98% de los lncRNAs se procesan post-transcripcionalmente y la mayoría poseen dos exones; esta cifra (42%) resulta muy elevada si se considera que tan solo el 6% de los genes codificantes tienen este número. Además, los exones de los lncRNAs son ligeramente más largos que los de genes codificantes (mediana de 149 y 132 bp respectivamente) al igual que sus intrones (mediana de 2,280 bp para lncRNAs y 1,602 bp para genes codificantes). Al poseer menos exones, normalmente los lncRNAs son más cortos que los mRNAs (medianas de 592 bp y 2,453 bp, respectivamente) (Derrien, *et al.*, 2012).

Adicionalmente, >25% de los lncRNAs muestran evidencia de procesamiento alternativo con al menos dos diferentes isoformas por locus. Cabe destacar que la gran mayoría de los intrones de los lncRNAs conservan los sitios de reconocimiento canónicos GT/AG (Derrien, *et al.*, 2012). Algo que se observó es que las isoformas resultantes del procesamiento alternativo de un gen no se expresan en niveles similares, es decir, una isoforma domina por sobre las otras en una condición específica. El número de isoformas expresadas por gen crece con el número de sus isoformas anotadas (Fig. 5) (Djebali, *et al.*, 2012).

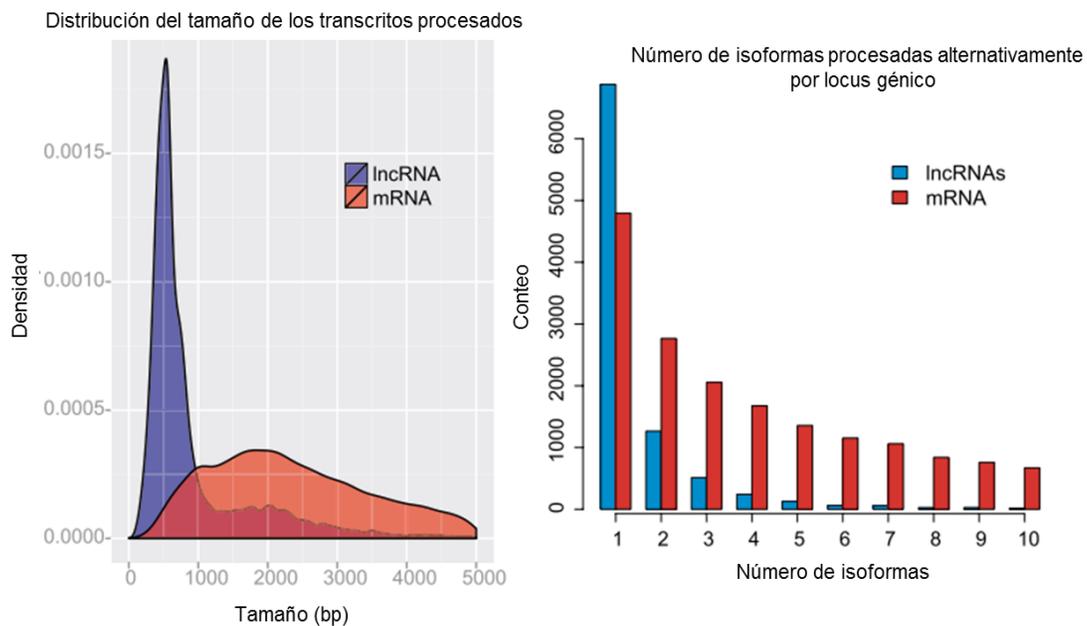


Figura 5: Distribución del tamaño y número de isoformas de lncRNAs comparado con el de los mRNAs. Entre las características estructurales de los lncRNAs, se comprende una menor longitud en su tamaño (izquierda) y un menor número de isoformas (derecha) cuando se compara con los mRNAs. Modificado de Derrien *et al.*, 2012.

Los experimentos bioquímicos han sido capaces de caracterizar algunos lncRNAs abundantes, tanto regulatorios como estructurales, por localización celular y similitud de secuencias (Rinn & Chang, 2012). Estudios genéticos posteriores han ido elucidando las funciones de estos transcritos. A pesar de que se han reportado miles de lncRNAs, sólo se han estudiado alrededor de 200 transcritos en diversos organismos, con una profundidad de análisis muy variable (Moran, *et al.*, 2012). Aunque las funciones de sólo una minoría se han descrito, su expresión regulada y dinámica sugiere que muchos más son funcionales (Cheetham, *et al.*, 2013).

Patrones de expresión

Se ha observado que muchos lincRNAs presentan un patrón de expresión espacio-temporal específico, que se puede ligar al establecimiento de barreras bien definidas en la expresión génica y a programas regulatorios tipo celular específicos (Fatica & Bozzoni, 2014). Son espacio-específico debido a que en muchos casos son tipo celular específico, a diferencia de lo observado para genes codificantes (Kutter, *et al.*, 2012). Consistente con esta observación, se determinó que los lncRNAs muestran una expresión de tejido mucho más variable a lo largo de todas las líneas y tejidos celulares, con sólo el 1% de los lncRNAs expresados en todos los tejidos (Derrien, *et al.*, 2012; Gibb, *et al.*, 2011). Existen reportes que argumentan que la mayoría de los lncRNAs se transcriben en el cerebro (Mercer, *et al.*, 2008). Interesantemente, en múltiples casos, esta especificidad de tejido se ha atribuido a la presencia de elementos transponibles embebidos en la vecindad del sitio de inicio de la transcripción de los lncRNAs (Fatica & Bozzoni, 2014).

Generalmente, los lncRNAs presentan niveles de expresión bajos en comparación a los RNAs codificantes en todos los tejidos, excepto en testículo, donde los lncRNAs exhiben niveles de expresión relativamente elevados (Derrien, *et al.*, 2012). En el caso de los lncRNAs antisentido, se ha reportado que poseen niveles de expresión más de diez veces menores cuando se comparan con la expresión de su transcrito sentido (Pelechano & Steinmetz, 2013). Sin embargo, hay propuestas de que no se trata de una baja cantidad de transcrito, sino de una población celular particular con expresión precisa de estas moléculas (Mattick, 2013). También se debe contemplar el hecho de que existe variabilidad de una célula a otra en la expresión génica, lo que le permite a una población de un tejido responder diferencialmente a estímulos ambientales idénticos, confiriendo una ventaja adaptativa al organismo (Pelechano & Steinmetz, 2013).

Los lncRNAs están enriquecidos en el núcleo celular. Este hecho podría correlacionar con los crecientes reportes de que los lncRNAs son reclutados a zonas específicas del genoma para modular su estructura cromatínica y así regular la expresión génica a nivel epigenético. El incremento de lncRNAs nucleares sobre citoplásmicos es consistente entre tipos celulares. Además, los lncRNAs aumentan en la cromatina cuando se comparan con mRNAs de todos los compartimentos nucleares. Por último, no parece haber distinción entre el tipo de lncRNA y la propensión a localizarse en el núcleo, lo que indica que el enriquecimiento nuclear es una propiedad general de estos transcritos (Derrien, *et al.*, 2012).

Algo que se ha observado es que la transcripción de los lncRNAs evoluciona rápidamente, guardando especificidad de tejido conservada sólo entre primates, pero no más

allá (Necsulea, *et al.*, 2014). Siguiendo lo anterior, se ha reportado que los lincRNAs podrían contribuir sustancialmente a la evolución de la expresión génica linaje y tejido específico (Kutter, *et al.*, 2012). Interesantemente, una gran proporción de los lincRNA -aproximadamente 11,000- son específicos del clado de los primates, resaltando la velocidad de su evolución (Necsulea, *et al.*, 2014).

Conservación de secuencia y evolución de lincRNAs

Se ha reportado que existen alrededor de 2,500 lincRNA altamente conservados, incluyendo ~400 que parecen haberse originado hace más de 300 millones de años; sin embargo, la mayoría de los lincRNAs evolucionan rápidamente en términos de secuencia y niveles de expresión (Necsulea, *et al.*, 2014). Las secuencias de los lincRNAs se restringen por factores como la posición genómica, la estructura secundaria del RNA y los niveles de expresión. En conjunto, las secuencias que codifican a estos transcritos acumulan menos sustituciones que las secuencias neutrales que los rodean y sumado a esto, la regulación transcripcional de estos RNAs muestra una conservación evolutiva (Pelechano & Steinmetz, 2013).

La aparición de nuevos lincRNAs se encuentra menos limitada evolutivamente que las estrategias biológicas que dependen de proteínas, debido a que no involucran modificaciones de las regiones codificantes. Esto aparentemente permite la rápida generación de lincRNAs y contribuye a la evolución de circuitos regulatorios. Siguiendo con lo anterior, aún en el caso más simple en el que la función regulatoria de un lincRNA se dé por el acto de su transcripción, no existen restricciones evolutivas respecto a la secuencia de RNA por sí misma o a la abundancia final del transcrito. Por lo tanto, en términos evolutivos, la evolución de funciones regulatorias mediada por lincRNAs provee un mecanismo que no requiere la invención de una nueva maquinaria (Pelechano & Steinmetz, 2013).

La selección purificadora de las secuencias genómicas representa una evidencia poderosa de la funcionalidad de un gen (Derrien, *et al.*, 2012). Una de las dificultades que se ha presentado al estudiar a los lincRNAs es que, a pesar de que estos transcritos tienen mayor conservación en su secuencia que la observada en intrones o secuencias intergénicas no transcritas, su identidad a lo largo de la evolución sigue siendo baja (Rinn & Chang, 2012). La excepción podría ser que los promotores de los lincRNAs poseen en promedio una secuencia más conservada que la de sus exones que es muy similar al grado de conservación de los promotores de genes codificantes, sugiriendo restricciones selectivas fuertes a nivel de transcripción (Derrien, *et al.*, 2012; Guttman, *et al.*, 2009; Necsulea, *et al.*, 2014).

Si las funciones de los lncRNAs están conservadas, se espera que sus loci se preserven evolutivamente (Kutter, *et al.*, 2012). Se ha descrito que algunos lncRNAs presentan sintenia – conservación del orden de los genes en un bloque o locus, en dos grupos de cromosomas cuando se comparan entre sí- y estructura secundaria similar en diversas especies; además, pueden mostrar patrones de expresión tejido-específicos, indicando su función en procesos celulares particulares (Qiu, *et al.*, 2013; Ulitsky, *et al.*, 2011).

Un ejemplo de la conservación funcional a pesar de la poca identidad en la secuencia es el descubrimiento de *cyrano* y *megamind*. Estos dos lincRNAs se identificaron en pez cebra, el primero se expresa en sistema nervioso y notocordio, mientras que *megamind* se expresa en cerebro y tejido ocular. El abatimiento de estos transcritos afecta el desarrollo del pez cebra; sorprendentemente, al hacer un rescate con la versión homóloga (ratón y humano) de estos transcritos, se pudo recuperar la función en la mayoría de los casos. Cabe destacar que la estructura exónica de estos lincRNAs es muy similar en todos los organismos, el locus muestra sintenia y existe un fragmento corto cercano al extremo 3' con alta identidad de secuencia (Ulitsky, *et al.*, 2011).

En eucariontes superiores, existen estudios que muestran que la ganancia o pérdida de lncRNAs es más frecuente que la de genes codificantes para proteína. Esta alta tasa de evolución de los lncRNAs sugiere su rápida adaptación como mecanismos regulatorios (Kutter, *et al.*, 2012; Pelechano & Steinmetz, 2013). Existen múltiples ejemplos de lncRNAs cuya función principal es la de regular diversos procesos celulares partiendo de una variedad de estrategias que involucran la asociación con otras moléculas como proteínas, DNA o hasta otros ncRNAs. Estos mecanismos de acción se discutirán a continuación, para dar un panorama más amplio de cómo se implican los lncRNAs en los procesos celulares.

Mecanismos de acción de los lncRNAs

Aunque aún se desconocen las funciones precisas de muchos lncRNAs, en general se piensa que actúan en la regulación transcripcional (Kutter, *et al.*, 2012). Los lncRNAs cumplen funciones regulatorias cruciales en casi todas las etapas de la expresión génica, desde sirviendo de guía para modificadores epigenéticos en el núcleo, hasta modulando la estabilidad de los mRNAs y la traducción en el citoplasma (Mercer & Mattick, 2013). Los lncRNAs poseen las habilidades intrínsecas para funcionar como una molécula de RNA, por lo que es importante entender las características químicas de estos transcritos que permiten su acción. Sumado a los cuatro nucleótidos canónicos, las secuencias de RNA pueden incluir más de cien diferentes modificaciones químicas en sus nucleótidos. A la fecha, estas marcas se habían detectado en

tRNAs, rRNAs y snoRNAs; sin embargo, se ha observado por una serie de técnicas que estas modificaciones podrían localizarse a lo largo del transcriptoma, posiblemente modulando de una manera similar la función de los lncRNAs. Muchas de estas marcas son reversibles y podrían comprender una capa adicional de regulación post-transcripcional, análogo a lo que ocurre con las histonas en el paisaje epigenético (Mercer & Mattick, 2013).

Un atributo elemental de los lncRNAs es su capacidad y propensión a plegarse en estructuras secundarias, y de órdenes mayores, que son termodinámicamente estables. Además de la capacidad del RNA para formar puentes de hidrógeno, siguiendo las reglas de apareamiento de bases Watson-Crick, se pueden formar interacciones tipo Hoogsteen, donde una tercera hebra se une al surco mayor de la doble cadena, o Hoogsteen reverso, donde se forman puentes de hidrógeno con la hebra rica en purinas en el dúplex de DNA (Buske, *et al.*, 2012; Koziol & Rinn, 2010; Mercer & Mattick, 2013).

Estas estructuras confieren un gran potencial y versatilidad para ser reconocidas por proteínas y secuencias blanco (Guttman & Rinn, 2012; Wan, *et al.*, 2012). Una clara predicción es que la estructuración de los lncRNAs dictará su función. Al momento, se han propuesto diversos dominios en este tipo de transcritos, como lo que se observa en las proteínas. Entre los dominios propuestos se encuentran el de unión a RNA, el de unión a proteína y el de unión a DNA, los cuales podrían ser críticos para permitirles llevar a cabo funciones particulares en la célula (Fig. 6) (Guttman & Rinn, 2012). Se propone además, una hipótesis de andamiaje modular que sugiere que los lncRNAs han sufrido una re-distribución por ganancia y pérdida de diversos módulos. Esto significaría que existe una construcción o creación de lncRNAs a partir de un rango diverso de elementos existentes, que provee funciones alternas y más complejas sujetas a selección evolutiva (Fatica & Bozzoni, 2014; Guttman & Rinn, 2012; Mercer & Mattick, 2013). La naturaleza flexible de los lncRNAs permite su unión con factores proteicos que no interaccionarían si sólo dependieran de asociaciones proteína-proteína (Guttman & Rinn, 2012; Mercer & Mattick, 2013).

Modelos de acción de los lncRNAs

La naturaleza intrínseca de los ácidos nucleicos en los lncRNAs les confiere la habilidad dual de funcionar como ligandos de proteínas y mediadores de la interacción por apareamiento de bases, que guían a complejos que contienen lncRNAs a sitios blanco específicos (Fatica & Bozzoni, 2014). Entre los papeles funcionales que se les ha atribuido a estos transcritos largos no-codificantes se enlistan 4 principales mecanismos: anzuelo (de proteínas o de miRNAs), andamio, guía y potenciador (Koziol & Rinn, 2010; Rinn & Chang, 2012).

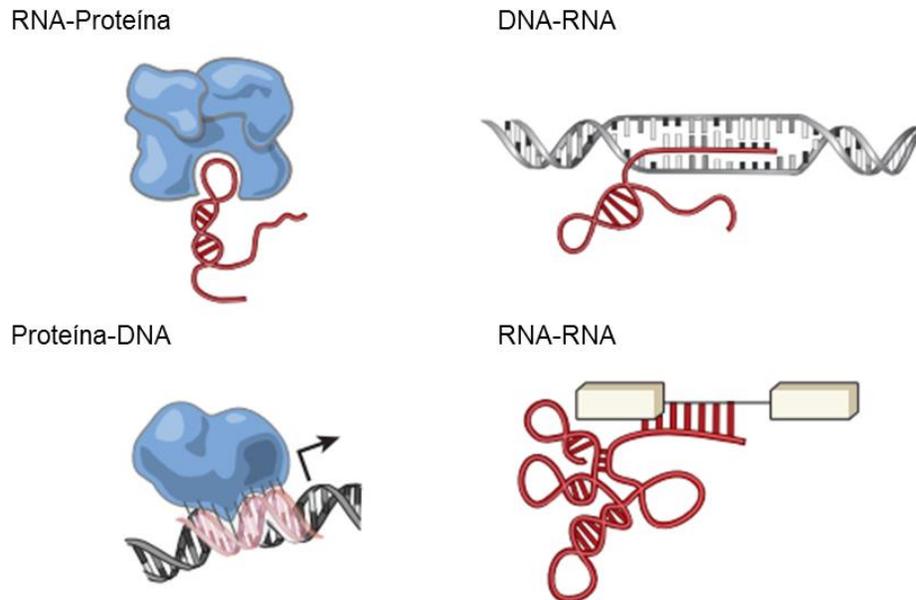


Figura 6: Principios modulares de los RNAs largos no-codificantes. Estos principios se basan en las cuatro posibles interacciones entre ácidos nucleicos y proteínas. Las proteínas representan un compañero importante de los lncRNAs, que al formar ribonucleoproteínas, pueden actuar como chaperonas, transportadores o efectores. Aunque existen pocos reportes de RNAs asociados a DNA de doble cadena, se considera que esta interacción puede dirigir las señales de RNA a loci de manera eficiente y selectiva. La asociación de DNA y proteína se ha descrito a profundidad y se sabe que podría mediar la interacción específica entre un lncRNA y sitios determinados del genoma. Por último, los motivos de unión a RNA poseen la habilidad de aparearse a manera de sensores altamente específicos a mRNAs, miRNAs u otros lncRNAs. Modificado de Guttman & Rinn, 2012.

Mecanismo de anzuelo

Se dice que los lncRNAs actúan como anzuelo o ancla cuando pueden servir para titular la interacción con otra molécula, a manera de inhibidores competitivos endógenos para una proteína de unión al DNA, previniendo su unión a elementos de reconocimiento en su secuencia, o un miRNA, en cuyo caso se denomina esponja de RNA (Kallen, *et al.*, 2013; Rinn & Chang, 2012). A continuación se describen dos ejemplos de lncRNAs a los cuales se les ha asociado este mecanismo: *Gas5* y *H19* (Fig. 7).

Ejemplo: Gas5

Un lncRNA al que se le atribuye este modo es *Gas5*, que se encarga de controlar la supervivencia celular al inducir o sensibilizar a las células para dirigirse hacia apoptosis (Esteller, 2011). *Gas5* contiene un motivo tallo-asa que se asemeja al sitio de unión a DNA del

receptor glucocorticoide; en condiciones de hambruna, la expresión de este lncRNA se induce y ayuda a liberar a este receptor del DNA para prevenir la transcripción de genes metabólicos (Rinn & Chang, 2012). Interesantemente, se han observado bajos niveles de expresión de *Gas5* en cáncer de mama y al sobre-expresar este transcrito se induce apoptosis y se suprime la proliferación celular, lo que le ha dado la clasificación de gen supresor de tumores (Prensner & Chinnaiyan, 2011).

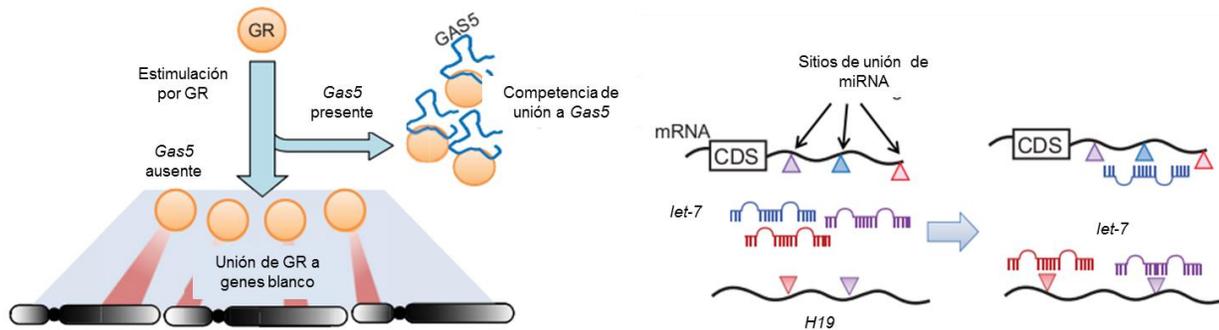


Figura 7: Mecanismo de acción tipo anzuelo. A la izquierda se puede ver cómo en la presencia de *Gas5* se inicia una competencia por GR, comprometiendo su unión en diversos sitios del genoma. A la derecha se muestra la función tipo lncRNA esponja; como *H19* posee secuencias específicas para el reconocimiento de *let-7*, se crea una titulación de este miRNA con respecto a sus sitios en otros mRNAs. Modificado de Prensner & Chinnaiyan, 2011.

Ejemplo: *H19*

Recientemente, se descubrió que el lncRNA *H19*, descubierto a finales de los años 80's, puede asociarse *in vivo* con el miRNA *let-7* durante la diferenciación muscular. *H19* es un lncRNA de 2.3 kb que muestra altos niveles de expresión en tejidos derivados del mesodermo y endodermo durante el desarrollo embrionario y mantiene su expresión en músculo esquelético y tejido cardíaco después del nacimiento (Guttman & Rinn, 2012). Además, este lncRNA se transcribe sólo del alelo materno a partir de la región improntada de *H19-Igf2* y su regulación transcripcional es regulada por una región intergénica de control y una región potenciadora. En 2013, se le atribuyó el papel de RNA-esponja cuando se descubrió que existen 4 sitios semillas para *let-7* dentro de *H19* y se validó que el abatimiento de este lncRNA incrementaba la acción de *let-7* sobre sus blancos, provocando una disminución en los niveles de expresión de los genes en los cuales actúa este miRNA. Algo que sobresale de este estudio es que la asociación entre *H19* y *let-7* parece conservarse evolutivamente, lo que indica que este mecanismo de acción podría ser también general en distintas especies (Kallen, *et al.*, 2013). Esto representa

un nuevo tipo de circuito regulatorio en el que diferentes tipos de RNAs (tanto codificantes como no-codificantes) pueden entablar una comunicación entre sí por medio de competir por miRNAs comunes (Fatica & Bozzoni, 2014).

Mecanismo de andamiaje

El mecanismo de andamiaje indica la formación de complejos lincRNA-RNP, donde el RNA adapta o junta a múltiples proteínas para acercarlas (Rinn & Chang, 2012). Este modo de acción parece ser común, existiendo estudios que reportan a un 30% de lincRNAs actuando de este modo en células troncales (Guttman & Rinn, 2012; Guttman, *et al.*, 2009). Dos ejemplos muy bien descritos de este mecanismo de acción son *Kcnq1ot1* y *HOTAIR*.

Ejemplo: *Kcnq1ot1*

Kcnq1ot1 es un lincRNA expresado del alelo paterno del dominio improntado *Kcnq1*, cuya función es la represión del alelo paterno *Kcnq1* y el silenciamiento de genes improntados. La secuencia de *Kcnq1ot1* le permite adaptarse a las enzimas metiltransferasa G9a y PRC2, para que actúen en conjunto depositando las modificaciones post-traduccionales H3K9me3 y H3K27me3, respectivamente, para reprimir diversos loci (Koziol & Rinn, 2010; Mercer & Mattick, 2013; Prensner & Chinnaiyan, 2011; Rinn & Chang, 2012).

Ejemplo: *HOTAIR*

Otro ejemplo bien documentado de un lincRNA que actúa como andamio es *HOTAIR* (Rinn, *et al.*, 2007; Tsai, *et al.*, 2010). Este lincRNA actúa en los loci HOX, que codifican para factores transcripcionales maestros que tienen la habilidad de establecer y regular la identidad de los segmentos del cuerpo de un organismo (Koziol & Rinn, 2010; Rinn, *et al.*, 2007; Tsai, *et al.*, 2010). La posición relativa de cada gen HOX dentro de su locus refleja su expresión espacio-temporal a lo largo del eje próximo-distal y del eje antero-posterior en embriones en desarrollo, lo que confiere una identidad celular única. La regulación de la expresión de los genes HOX es coordinada por la interacción dinámica de PcG y TrxG (Koziol & Rinn, 2010). *HOTAIR* es un lincRNA que se transcribe en el dominio *HoxC* y ayuda a unir complejos represores. En su dominio 5' une a PRC2, mientras que en su extremo 3' se asocia al complejo LSD1/CoREST/REST, ayudando en el silenciamiento de genes en el dominio *HoxD* (Croce, 2010; Tsai, *et al.*, 2010; Tsai, *et al.*, 2011).

Mecanismo de guía

Los perfiles de expresión de los lncRNAs pueden influenciar estados epigenéticos específicos al interactuar con complejos reguladores de la cromatina, impartiendoles especificidad al llevar a estas proteínas a sus genes blanco (Guttman & Rinn, 2012; Koziol & Rinn, 2010; Mercer & Mattick, 2013). El mecanismo de guía se utiliza frecuentemente y consiste en la formación de un complejo lncRNA-RNP que confiere especificidad para dirigirse a regiones genómicas por medio de asociaciones DNA-proteína o RNA-DNA (Guttman & Rinn, 2012; Rinn & Chang, 2012). Los lncRNAs pueden unirse a diversos complejos reguladores epigenéticos, como PcG y TrxG, para establecer dominios heredables de heterocromatina y eucromatina, respectivamente (Koziol & Rinn, 2010).

Ejemplos: *Kcnq1ot1* y *HOTAIR*

Kcnq1ot1 y *HOTAIR* actúan como andamios y guías, dirigiendo específicamente a las proteínas con las que se asocian hacia blancos génicos particulares para que ejerzan su actividad silenciadora (Rinn & Chang, 2012). En el caso de *HOTAIR*, un estudio reportó que este lincRNA es capaz de llevar a la subunidad PRC2 a 854 genes, asociados con términos ontológicos de señalización célula-célula y desarrollo. Además, se observó que la ocupación de PRC2 inducida por *HOTAIR* tiende a localizarse sobre promotores y en menor medida sobre el cuerpo de los genes (Gupta, *et al.*, 2010). En un segundo estudio, donde se introduce la técnica de ChIRP-Seq, se identificó que existen 832 sitios de ocupación por *HOTAIR* en el genoma, que ocurren en múltiples cromosomas y que están enriquecidos en promotores e intrones (Chu, *et al.*, 2011). Se puede ver un esquema de estos lncRNAs en la figura 8 (Prensner & Chinnaiyan, 2011).

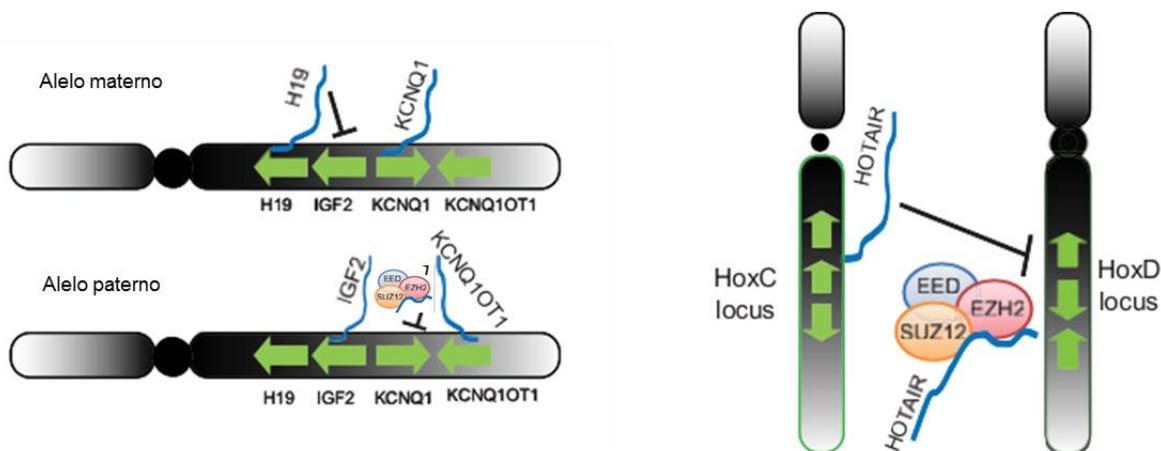


Figura 8: Mecanismos de andamio y guía. Izquierda: *Kcnq1ot1*. Al reclutar a PRC2/G9a y guiar a estos complejos al alelo paterno, logra silenciar al gen *Kcnq1*. Derecha: *HOTAIR* y cómo con sus funciones de andamio y guía reclutan a PRC2/LSD1/coREST hacia el locus *HoxD*. Modificado de Prensner & Chinnaiyan 2011.

Mecanismo de potenciador

Por último, el mecanismo potenciador funciona cuando los lncRNAs tienen la capacidad de incrementar la expresión de genes vecinos. Sin embargo, hay que distinguir que se han identificado dos clases de transcritos potenciadores: los eRNAs –simple producto de la transcripción de secuencias tipo enhancer- y los lncRNAs que juegan un papel en la formación de contactos entre enhancers y promotores para promover la expresión génica. La identificación de este tipo de transcritos se basa en la suma de señales de ChIP-seq para modificaciones como H3K4me1, H3K27ac y la proteína p300. Muchos de los eRNAs encontrados son bidireccionales, carecen de poli-(A) y tienen un bajo número de copias (Rinn & Chang, 2012). Actualmente, se conoce que los tanto los eRNAs como los lncRNAs potenciadores se transcriben por la RNA Pol II a partir de secuencias enhancers activas y existen esfuerzos de consorcios como FANTOM para caracterizarlos (Prensner & Chinnaiyan, 2011).

Ejemplo: HOTTIP

Un claro ejemplo de un lncRNA con esta función es *HOTTIP*, que interactúa directamente con la proteína WDR5/MLL para promover la deposición de la marca H3K4me3 y la transcripción génica. *HOTTIP* es un lincRNA de 3.8 kb codificado en la región 5' del dominio *HoxA* que se expresa *in vivo* en el brote de las extremidades durante el desarrollo en pollo y en regiones distales del cuerpo en humanos (Wang, *et al.*, 2011). Este lncRNA promueve la formación de asas cromatínicas para acercar espacialmente múltiples regiones regulatorias del dominio *HoxA*. Además, existe una inter-dependencia entre *HOTTIP* y WDR5 que crea un circuito de retroalimentación positivo, manteniendo encendido el estado de este locus (Fig. 9) (Rinn & Chang, 2012; Wang, *et al.*, 2011).

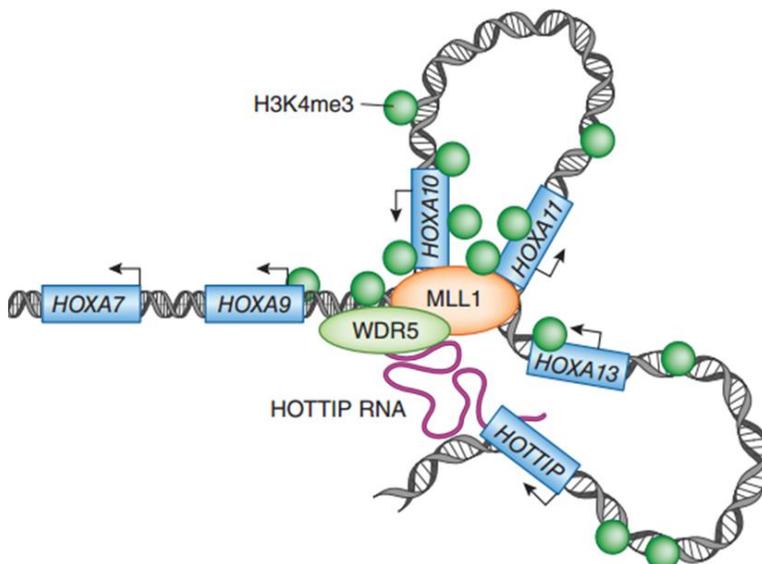


Figura 9: Mecanismo de acción potenciadora. En las células distales humanas, se requiere de *HOTTIP* para mantener la transcripción activa de los genes en el extremo 5' del dominio HOXA. Esto se logra gracias al reclutamiento de MLL1 y WDR5, guiando a estos complejos activadores y promoviendo la formación de asas tridimensionales de la cromatina. Tomado de Guil & Esteller, 2012.

En conclusión, el mecanismo que elige un lncRNA está todavía lejos de ser entendido (Fatica & Bozzoni, 2014). A manera de resumen, se presenta la figura 10, donde se puede visualizar cada uno de los modos de acción descritos anteriormente (Rinn & Chang, 2012). Asimismo, es importante mencionar que los lncRNAs pueden actuar con cualquiera de estos modelos de dos maneras: ya sea ejerciendo su función sobre genes contiguos, es decir en *cis*, o llevando estos mecanismos a locus más distantes, en *trans* (Fig. 11) (Cheetham, *et al.*, 2013; Guttman & Rinn, 2012; Rinn & Chang, 2012).

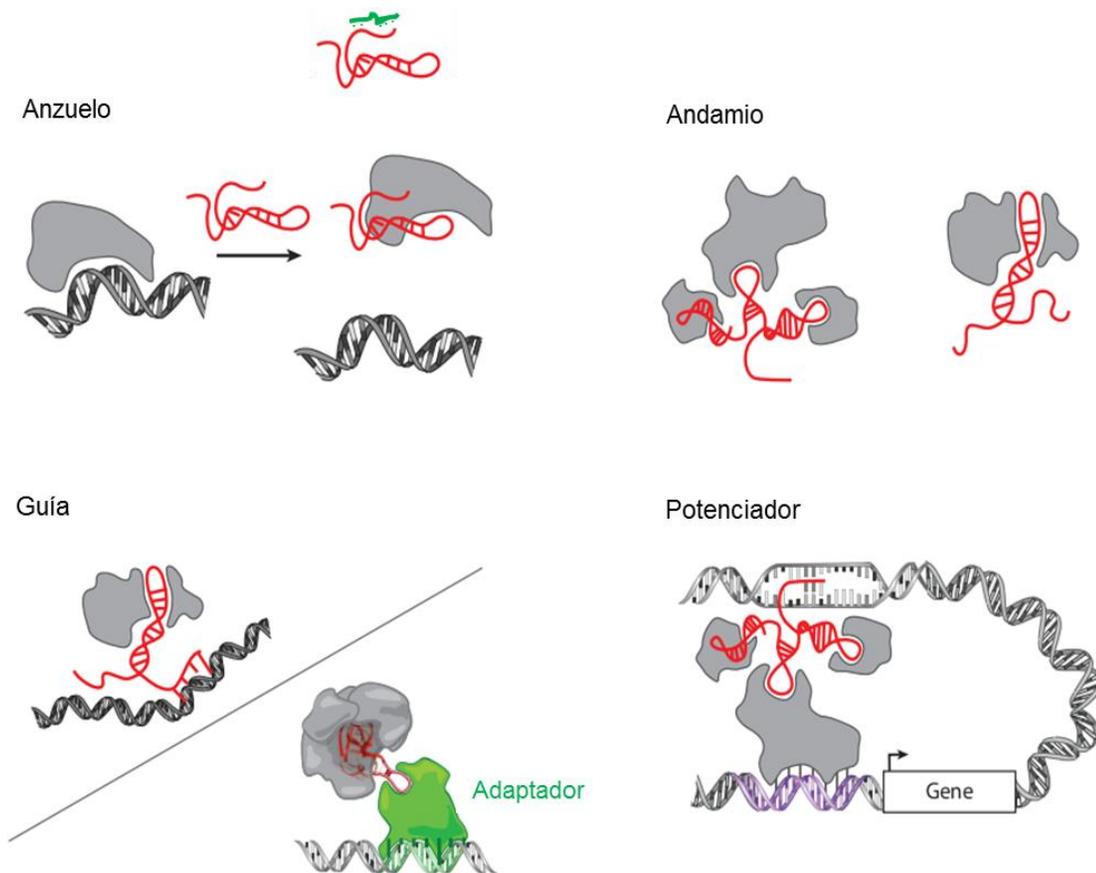


Figura 10: Los cuatro principales modelos de acción reportados para los lncRNAs cuya función ha sido estudiada. Modificado de Rinn & Chang, 2012.

Acción en cis

Se ha definido una regulación en *cis* como una que ejerce su función sobre los genes vecinos en el mismo alelo del que se transcribió el lncRNA (Guttman & Rinn, 2012; Rinn & Chang, 2012). Algunos reportes consideran que sigue nombrándose regulación en *cis* mientras actúe en el mismo cromosoma del cuál fue transcrito (Kornienko, *et al.*, 2013). Para poder considerarse un regulador en *cis*, se deben cumplir tres requisitos: (i) que los niveles de

expresión del gen aledaño correlacionen con la expresión del RNA en todas las condiciones, (ii) que la pérdida de función del lncRNA afecte la expresión del gen vecino y (iii) que el lncRNA afecte la expresión de un gen próximo en el mismo alelo del que se expresa (Guttman & Rinn, 2012).

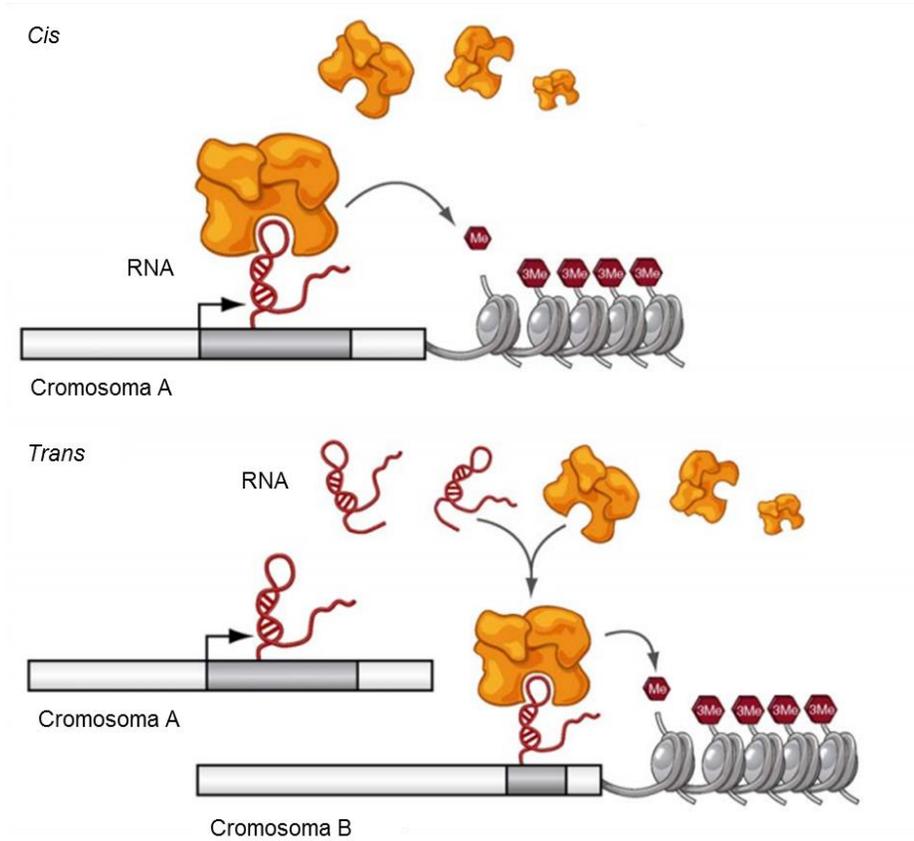


Figura 11: Mecanismos de acción de un lncRNA. En esta imagen se muestran las dos maneras en que un lncRNA puede actuar: arriba se presenta la regulación *cis*, donde el lncRNA se dirige a los genes aledaños para ejercer su función. Abajo se observa la regulación *trans*, donde el producto de lncRNA se dirige a sitios alejados de donde se transcribió, incluso en regiones de otros cromosomas. Modificado de Koziol & Rinn, 2010.

En teoría, un lncRNA tiene la capacidad intrínseca de tener una regulación tipo *cis* porque puede funcionar mientras permanece anclado a su propio locus, mientras que un mRNA tiene que dissociarse, ser exportado y traducido. Debido a esta característica intrínseca, se ha especulado que este tipo de regulación podría ser un mecanismo común para múltiples lncRNAs; sin embargo, la evidencia funcional global sugiere fuertemente que éste no es el caso (Guttman & Rinn, 2012). Los lncRNAs tienen la posibilidad de actuar en *cis* mediante dos formas: la primera depende del lncRNA, ya sea de su producto final o de una función tipo “ancla”, donde el lncRNA nascente permanece unido a la RNA Pol II, pero permite dirigir modificadores epigenéticos hacia el gen vecino. La segunda forma es mediante el acto de transcripción, resultando en un mecanismo *cis-a-priori* (Kornienko, *et al.*, 2013).

El ejemplo más claro de un lncRNA que actúa en *cis* es *Xist*, el segundo lncRNA en ser descrito (hace más de 20 años). Este transcrito es expresado por uno de los dos cromosomas X en las hembras de los mamíferos, se retiene en el núcleo y después de procesarse en el

desarrollo temprano actúa en *cis* para silenciar al cromosoma X completo. *Xist* se encarga de iniciar y esparcir la inactivación de este cromosoma mediante su asociación física con PcG en las células somáticas. Tanto la interacción con PRC2, como la capacidad de *Xist* para promover la hiper-metilación vía la DNA metil-transferasa 3A (DNMT3A), resultan en la condensación de la cromatina y por ende en la represión transcripcional a largo plazo de este cromosoma (Augui, *et al.*, 2011; Guttman & Rinn, 2012; Moran, *et al.*, 2012; Rinn & Chang, 2012). La manera exacta en que *Xist* induce el silenciamiento génico sigue siendo un misterio, sin embargo se sabe que una región conservada de repetidas A (RepA) es crucial para que ocurra dicho proceso. También se conoce que *Xist* es controlado principalmente a nivel transcripcional y algunos eventos de maduración post-transcripcional (Augui, *et al.*, 2011).

Aún se desconoce cómo los lncRNAs que actúan en *cis* se retienen en sus sitios de transcripción (Fatica & Bozzoni, 2014). Consistentemente, se ha observado una co-expresión de lncRNAs con sus mRNAs vecinos y se piensa que existe una co-regulación transcripcional tanto para los genes circundantes como para el lncRNA (Kutter, *et al.*, 2012; Rinn & Chang, 2012). Algo notable es que los loci de lncRNAs en mamíferos y sus genes codificantes adyacentes muestran una tendencia significativa a exhibir perfiles de expresión espacio-temporales similares. Además, un estudio reportó que la transcripción de lncRNAs en hígado ocurre significativamente más frecuente cerca de genes codificantes específicos de este órgano (Kutter, *et al.*, 2012). No obstante, experimentos recientes han demostrado que algunos lncRNAs con función en *cis* también son capaces de actuar en *trans* (Rinn & Chang, 2012).

Acción en trans

Se define una regulación tipo *trans* como aquella que no cumple con el criterio de estar regulada en *cis*, es decir, puede regular a sitios que están a mayor distancia, incluso en otros cromosomas (Guttman & Rinn, 2012). En este tipo de regulación, los lncRNAs difunden a partir de su sitio de síntesis y pueden actuar directamente en múltiples sitios que se localizan a una distancia considerable (Rinn & Chang, 2012; Wahlestedt, 2013). Sin embargo, aún se desconoce cómo un lncRNA que actúa en *trans* encuentra a genes blanco con localización distante (Fatica & Bozzoni, 2014).

Se ha observado que la mayoría de los lncRNAs caracterizados funcionan como reguladores-*trans* (Guttman & Rinn, 2012). Un ejemplo de un lincRNA que ejerce su función en *trans* es *HOTAIR*, cuyo sitio blanco se localiza en el dominio *HoxD*, situado en un cromosoma distinto al locus donde se origina (*HoxC*). *HOTAIR* ayuda a establecer un dominio de heterocromatina en un grupo de genes del dominio *HoxD* que son clave para el correcto desarrollo embrionario (Chu, *et al.*, 2011; Khalil, *et al.*, 2009; Tsai, *et al.*, 2010). El descubrimiento de *HOTAIR* tuvo dos grandes implicaciones biológicas. La primera es que este lncRNA representó el primer ejemplo de una regulación tipo *trans* (Guttman & Rinn, 2012; Koziol & Rinn, 2010; Tsai, *et al.*, 2010). La segunda es que se observó que el aumento en la expresión de este transcrito en cáncer mamario parecía reprogramar el perfil de unión de PcG. Estos estudios sugirieron que la interdependencia entre *HOTAIR* y PRC2 es crucial en la promoción de la invasividad en este tipo de cáncer (Fig. 12) (Croce, 2008; Tsai, *et al.*, 2010; Tsai, *et al.*, 2011).

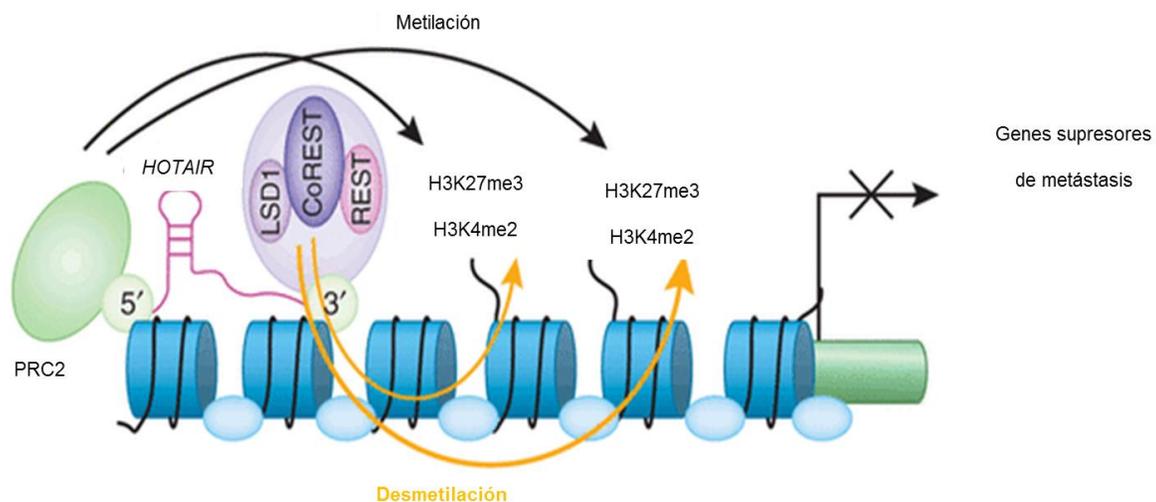


Figura 12: *HOTAIR* y su función en la progresión de cáncer de mama. Se puede observar el mecanismo de andamiaje que asocia en el extremo 5' a PRC2 y en la región 3' a LSD1/CoREST/REST para depositar la marca H3K27me3 y remover la marca H3K4me2, respectivamente. Cuando existen niveles incrementados de este lincRNA, como se observa en cáncer mamario, existirá una guía inespecífica de estos complejos, causando el silenciamiento de genes que deberían estar activos, como genes supresores de tumores. Modificado de Croce, 2010.

En conclusión, los lncRNAs constituyen un grupo altamente heterogéneo de moléculas de RNA, lo que les permite cubrir un amplio espectro de funciones moleculares y celulares al implementar sus diferentes modos de acción (Gutschner & Diederichs, 2012). Se ha observado que las alteraciones en la estructura primaria, secundaria y en los niveles de expresión de los

lncRNAs subyacen a enfermedades que van desde las neurodegenerativas hasta el cáncer (Wapinski & Chang, 2011). Siguiendo con lo anterior, la desregulación de lncRNAs puede impactar de diversas maneras, ya sea durante el desarrollo o progresión del cáncer, hasta asociándose a estadios de metástasis en una gran variedad de tejidos humanos (Cheetham, *et al.*, 2013; Qiu, *et al.*, 2013). Además, los lncRNAs se han implicado en diversos procesos celulares que son comúnmente desregulados en el cáncer, como se discutirá a continuación.

Transcritos no-codificantes largos y su relación con el cáncer

Un fuerte indicio de que los lncRNAs son cruciales en la etiología del cáncer es que los estudios de asociación a lo largo del genoma (GWAS, en inglés) muestran que los polimorfismos en un solo nucleótido, o SNPs, que se relacionan con cáncer se encuentran en su mayoría en regiones no-codificantes. El 40% de estos SNPs se localiza sobre secuencias intrónicas, el 44% sobre regiones intergénicas y sólo un 12% se relacionan con loci de genes codificantes (Cheetham, *et al.*, 2013).

El cáncer comprende un grupo heterogéneo de enfermedades causadas en su mayoría por la expresión génica anormal, resultado de cambios dinámicos en el genoma y de la alteración de redes de genes que mantienen una identidad, crecimiento y diferenciación celular normal (Cheetham, *et al.*, 2013; Gutschner & Diederichs, 2012). Una característica unificadora de esta condición es la creación de células anormales que crecen más allá de sus límites naturales para formar tumores (Gutschner & Diederichs, 2012; Hanahan & Weinberg, 2000). Los tumores son masas de células proliferantes, que pueden ser benignas o cancerosas; en el caso de que sean malignas se formarán tejidos complejos, compuestos de múltiples y distintos tipos celulares que participan en interacciones heterotípicas entre sí (Hanahan & Weinberg, 2011).

La observación del cáncer humano y de modelos animales sugiere que el desarrollo tumoral procede de forma análoga a la evolución Darwiniana. Esto vuelve a la formación de tumores un proceso de varios pasos que se puede entender como una sucesión de cambios genéticos, cada uno confiriendo una ventaja en el crecimiento, que conlleva a la conversión progresiva de células normales a cancerosas (Gutschner & Diederichs, 2012; Hanahan & Weinberg, 2000). Es así como se ha fundamentado el descubrimiento de mutaciones en oncogenes -con una ganancia de función dominante- y genes supresores de tumores, con pérdida de función recesiva (Gutschner & Diederichs, 2012; Hanahan & Weinberg, 2000). Los proto-oncogenes son genes que codifican proteínas involucradas en la proliferación celular e inhibición de la apoptosis y que si se activan por mutaciones se expresan de manera

descontrolada, volviéndose oncogenes. Los genes supresores de tumores representan la contraparte de los oncogenes, debido a que están involucrados en el control negativo del ciclo celular. Normalmente, su función es el inhibir la proliferación celular y el desarrollo de tumores (Croce, 2010).

Dentro de las reglas moleculares, bioquímicas y celulares que gobiernan la transformación de células normales humanas a cánceres malignos y que son compartidas por la mayoría de los tumores, se encuentran seis alteraciones esenciales en la fisiología de la célula (Hanahan & Weinberg, 2000). Estos sellos representan capacidades distintivas y complementarias, que permiten la proliferación celular y la diseminación metastásica, involucrando principalmente defectos en circuitos regulatorios que gobiernan la homeostasis celular (Hanahan & Weinberg, 2000; Hanahan & Weinberg, 2011). A la fecha, la auto-suficiencia en señales de proliferación, la insensibilidad para supresores del crecimiento, la evasión de la apoptosis, el potencial ilimitado de replicación, la angiogénesis sostenida y la invasión de tejidos, continúan representando una guía para entender la biología del cáncer. Sin embargo, hay que tener en cuenta que, más allá de las alteraciones propuestas para el cáncer, la contribución del microambiente tumoral es crucial en la biología de los tumores (Hanahan & Weinberg, 2011).

Interesantemente, en cada una de estas alteraciones o sellos se han visto involucrados lncRNAs (Gutschner & Diederichs, 2012). Dada la gran escala de desregulación de los lncRNAs, es altamente posible que estos transcritos no sean un efecto secundario del cáncer, sino que se involucren en la carcinogénesis o la metástasis (Qiu, *et al.*, 2013). Además, las funciones de los lncRNAs están asociadas muy de cerca con la abundancia de los transcritos (Du, *et al.*, 2013). Esto ha llevado a algunos investigadores a proponer que la transcriptómica de los tumores servirá durante un diagnóstico inicial, permitiendo tratamientos personalizados y un pronóstico más acertado (Cheetham, *et al.*, 2013). A continuación se resume en qué consiste cada alteración del cáncer y se hace referencia a diversos lncRNAs que se han encontrado asociados a este tipo de padecimiento.

Auto-suficiencia en señales de proliferación

Esta es la característica más fundamental de las células cancerosas e involucra la habilidad de mantener una proliferación crónica. En un tejido normal, existe un control riguroso para la producción y liberación de señales promotoras del crecimiento, conocidas como mitógenos, que instruyen la entrada y progresión del ciclo celular, asegurando así la homeostasis y la proliferación celular. Las señales de crecimiento mitogénicas son transmitidas hacia la célula

por receptores que atraviesan la membrana celular y que unen a distintas clases de moléculas señalizadoras. Se ha encontrado que diversos oncogenes modulan la autonomía celular de proliferar. Cuando las células cancerosas desregulan estas señales, se vuelven dueñas de su propio destino (Hanahan & Weinberg, 2000; Hanahan & Weinberg, 2011).

Las células cancerosas pueden adquirir la capacidad de mantener la señal de proliferación de diversas maneras: la primera es mediante la producción de sus propios factores de crecimiento y los receptores correspondientes a estas moléculas, resultando así en una estimulación proliferativa de tipo autócrina. La segunda es que las células malignas manden señales para estimular a células normales dentro del estroma asociado al tumor para que las provean de diversos factores de crecimiento en una señalización parácrina (Gutschner & Diederichs, 2012; Hanahan & Weinberg, 2011). La tercera es que los receptores de factores de crecimiento se encuentren en niveles elevados o las cascadas de señalización estén alteradas, causando una respuesta exacerbada en las células cancerosas (Gutschner & Diederichs, 2012). La última es que exista una activación constitutiva de vías de señalización que existen debajo de los receptores o que se interrumpan mecanismos negativos de retro-alimentación, causando que las células se vuelvan completamente independientes de factores exógenos (Gutschner & Diederichs, 2012; Hanahan & Weinberg, 2011).

Un tejido frecuentemente afectado por el desarrollo del cáncer es el mamario, donde existe una auto-suficiencia en señales de proliferación. Este tipo de cáncer es el más común y representa la segunda principal causa de mortalidad a nivel mundial en muertes relacionadas con cáncer en mujeres (Marella, *et al.*, 2009). Cuando se desarrolla un cáncer en esta glándula se alteran algunos receptores como el del factor de crecimiento epidérmico (EGF-R/*erbB*) y el receptor de HER2/*neu* (Hanahan & Weinberg, 2000). Otros receptores que se ven comprometidos comúnmente en la glándula mamaria son los de hormonas esteroideas como estrógeno (ER), progesterona (PR) y andrógeno (AR). Un coactivador transcripcional único para los receptores esteroides PR, ER, AR y receptor glucocorticoide (GR) es el activador del receptor esteroide (SRA), que actúa como un lncRNA (Gutschner & Diederichs, 2012; Lanz, *et al.*, 1999). SRA es parte de un complejo RNP que contiene al coactivador SRC-1 y que completa su función de activación en *trans* por medio del dominio variable amino-terminal AF-1 de los receptores esteroides nucleares (Gutschner & Diederichs, 2012). El papel de estos receptores activados es dirigir el ensamblaje y la estabilización del complejo de pre-iniciación en un ambiente transcripcional permisivo en el promotor de un gen blanco (Lanz, *et al.*, 1999).

Se ha detectado la expresión de SRA en tejidos humanos mamarios normales y en niveles elevados en tumores de esta glándula. El aumento de SRA correlaciona con los niveles

de ER/PR en subgrupos tumorales particulares (Leygue, *et al.*, 1999). Por esta razón, se especula que los cambios en la expresión de SRA podrían estar involucrados en los mecanismos subyacentes a la progresión tumorigénica y a la resistencia hormonal observada en cáncer de mama (Gutschner & Diederichs, 2012; Leygue, *et al.*, 1999).

Insensibilidad para señales supresoras del crecimiento

Una propiedad altamente complementaria a la capacidad de mantener las señales proliferativas es la habilidad para evadir los programas supresores del crecimiento (Gutschner & Diederichs, 2012). En un tejido normal, existen múltiples señales inhibitoras del crecimiento que mantienen la homeostasis celular al bloquear la proliferación por dos mecanismos distintos: ya sea al forzar a la célula a salir del ciclo celular y mantenerse en un estado quiescente (G_0) o llevando a la célula a entrar a un estado post-mitótico, que se asocia a la adquisición de características específicas de diferenciación (Hanahan & Weinberg, 2000).

Las señales supresoras de la proliferación incluyen inhibidores solubles y fijos, embebidos en la matriz extracelular y en otras superficies de células circundantes. Estas señales anti-crecimiento son recibidas por receptores que atraviesan la membrana celular y que se comunican con vías de señalización asociadas (Hanahan & Weinberg, 2000). Se ha observado que los inhibidores de la proliferación dependen en muchas ocasiones de genes supresores de tumores como *RB*, *Tp53* y *PTEN*, frecuentemente mutados en los tumores (Gutschner & Diederichs, 2012; Hanahan & Weinberg, 2011).

Además, algunos lncRNAs se han visto asociados a la inhibición del crecimiento celular en contextos tumorales. Un claro ejemplo es el RNA no-codificante antisentido en el locus INK4, cuyas siglas le dan el nombre de *ANRIL*. Este lncRNA bloquea la actividad de genes supresores de tumores, como p15 (INK4B), al interactuar con SUZ12 y reclutar de esta forma a PRC2 para reprimir su expresión. *ANRIL* también se asocia a CBX7, resultando en el reclutamiento de PRC1 hacia el locus de *p15/CDKN2B-p16/CDKN2A-p14/ARF* (INK4B/INK4A/ARF), silenciando esta región en *cis* mediante la adquisición de la marca en la histona H3K27me3 (Cheetham, *et al.*, 2013; Gutschner & Diederichs, 2012; Tsai, *et al.*, 2011). El abatimiento de *ANRIL* incrementa la expresión de p15 e inhibe la proliferación celular, además CBX7 y *ANRIL* tienen niveles elevados en cáncer de próstata humano (Gutschner & Diederichs, 2012).

Otro lncRNA relevante que actúa para inhibir la proliferación celular es *lincRNA-p21*, que reside junto al gen *p21* (*Cdkn1a*) en ratón (Huarte, *et al.*, 2010). La expresión de este lincRNA es inducida por *Tp53*, el cual se une al promotor de *lincRNA-p21* cuando existe daño en DNA en diversos modelos tumorales. *LincRNA-p21* se asocia a la RNP heterogénea nuclear K

(hnRNP K), modulando la localización de este complejo y confiriéndole especificidad para reprimir a genes de la ruta de *Tp53*. También se ha reportado que *lincRNA-p21* juega un papel fisiológico en la regulación de la viabilidad celular después de daño a DNA, funcionando como un inductor de apoptosis (Cheetham, *et al.*, 2013; Gutschner & Diederichs, 2012; Huarte, *et al.*, 2010; Tsai, *et al.*, 2011).

Evasión de la apoptosis

La habilidad de las poblaciones tumorales de expandirse y crecer en número está determinada no sólo por la tasa de proliferación celular, sino también por la tasa de muerte celular (Hanahan & Weinberg, 2000). Existen tres vías que pueden llevar a la muerte celular y su activación debe ser controlada cuidadosamente por las células tumorales: la autofagia, la necrosis y la apoptosis (Gutschner & Diederichs, 2012).

La autofagia opera comúnmente a bajos niveles en las células, pero puede activarse por diversos tipos de estrés celular, como deficiencia de nutrientes. Este mecanismo puede verse como un programa de reciclado que permite a las células degradar a sus organelos y usar los productos degradados para propulsar a rutas biosintéticas o para la producción de energía (Gutschner & Diederichs, 2012; Hanahan & Weinberg, 2011).

Durante la necrosis, las células se vuelven grandes y explotan, liberando sus contenidos en el micro-ambiente local del tejido. Existe evidencia de que la necrosis puede tanto eliminar a células cancerosas como promover su expansión mediante la atracción de células pro-inflamatorias, que a su vez pueden activar la angiogénesis, la invasividad y la proliferación de células cancerosas (Gutschner & Diederichs, 2012).

La apoptosis o muerte celular programada es una barrera grande que debe ser sobrepasada por las células cancerosas (Hanahan & Weinberg, 2000). Dentro de las diversas señales a estrés que inducen la apoptosis está la señalización alterada por niveles elevados de oncogenes y daño a DNA, asociado a una proliferación exacerbada. Además, la apoptosis se atenúa en aquellos tumores que son exitosos en proceder a estadios de alto grado (que son más malignos) y que se vuelven resistentes a terapias (Gutschner & Diederichs, 2012; Hanahan & Weinberg, 2011). La maquinaria apoptótica se compone tanto de reguladores como de efectores. Los reguladores, a su vez, se dividen en dos circuitos principales, uno que recibe y otro que procesa señales extracelulares que inducen la muerte (programa extrínseco) y el otro que siente e integra una variedad de señales de origen intracelular (programa intrínseco). Cada uno de ellos culmina con la activación de unas proteasas conocidas como caspasas. La caspasa 8 es específica en el programa extrínseco y la 9 en el intrínseco; estas enzimas

proceden para iniciar una cascada de proteólisis en la cual la célula es desintegrada progresivamente y luego consumida por sus vecinos y por fagocitos. Se involucra principalmente al programa intrínseco como barrera en la patogénesis cancerosa (Hanahan & Weinberg, 2011). El inductor apoptótico que integra señales entre los reguladores y efectores se controla por el balance entre miembros tanto pro como anti-apoptóticos de la familia Bcl-2 de proteínas reguladoras. Bcl-2 es un inhibidor de la apoptosis que actúa en gran parte mediante la unión y supresión de dos proteínas pro-apoptóticas (Bax y Bak, embebidas en la membrana externa mitocondrial). Cuando dejan de estar inhibidas, Bax y Bak rompen la integridad de la membrana externa mitocondrial, causando la liberación de proteínas de señalización pro-apoptóticas, como el citocromo c. Una vez que se libera el citocromo c, éste activa una cascada de caspasas que inducen los múltiples cambios celulares asociados al programa apoptótico (Hanahan & Weinberg, 2011).

Las células cancerosas pueden crear una resistencia hacia la apoptosis por una serie de estrategias, como la pérdida del regulador pro-apoptótico y supresor de tumores *Tp53*, que induce la expresión de las proteínas pro-apoptóticas Noxa y Puma (Gutschner & Diederichs, 2012; Hanahan & Weinberg, 2000; Hanahan & Weinberg, 2011). El 50% de los cánceres humanos tiene mutaciones o pérdida del gen *Tp53*, lo que vuelve a las células más resistentes a estas señales de estrés celulares. Alternativamente, los tumores muestran una expresión mayor de factores de sobrevivencia o de los reguladores anti-apoptóticos Bcl-2 y Bcl-x_L (Gutschner & Diederichs, 2012).

El daño a DNA puede inducir cinco lncRNAs del promotor del gen *p21*, uno de ellos es conocido como *PANDA* (ncRNA asociado a *p21* activado por daño a DNA). *PANDA* juega un papel crítico en la inhibición de la apoptosis mediada por *Tp53*, actúa en *trans* mediante la interacción y titulación con el factor de transcripción NF- κ B para favorecer la sobrevivencia celular durante daño al DNA y limitar la expresión de los genes pro-apoptóticos (Gutschner & Diederichs, 2012; Rinn & Chang, 2012).

Potencial ilimitado de replicación

Tanto la auto-suficiencia en señales de proliferación, como la insensibilidad para señales supresoras del crecimiento y la evasión de la apoptosis llevan al desacoplamiento del programa de crecimiento celular por medio de señales en su ambiente. En principio, esta desregulación en el programa de proliferación debería bastar para permitir la generación de poblaciones celulares extensas. Sin embargo, se ha demostrado que esto no asegura el crecimiento tumoral expansivo (Hanahan & Weinberg, 2000).

Las células cancerosas requieren un potencial de replicación ilimitado para poder generar tumores macroscópicos. Esta capacidad contrasta con la conducta de la mayor parte de los linajes celulares normales, en los cuales existe un número finito de ciclos celulares de crecimiento y división (alrededor de 60-70), conocido como el límite de Hayflick (Hanahan & Weinberg, 2000; Hanahan & Weinberg, 2011). Esta limitante se ha asociado a dos tipos de barreras para la proliferación: senescencia y crisis. La primera es la entrada irreversible a un estado viable pero no-proliferativo celular y la segunda involucra a la muerte celular (Hanahan & Weinberg, 2011).

La parte final de los cromosomas, conocida como telómeros, es crucial para este límite de replicación. En los vertebrados, los telómeros se acortan progresivamente con cada ciclo celular; sin embargo, las células tumorales han encontrado dos maneras de mantener el tamaño de los telómeros: 1) incrementando la expresión de la telomerasa, encargada de sintetizar a los telómeros (estrategia observada en el 90% de los tumores) y 2) empleando una alternativa para alargar los telómeros por medio de recombinación por intercambio inter-cromosómico (Gutschner & Diederichs, 2012; Hanahan & Weinberg, 2000).

Adicionalmente, existe un grupo de lncRNAs llamados *TERRA* (RNA telomérico que contiene repeticiones) que son transcritos de loci sub-teloméricos, los cuales participan en la formación de heterocromatina. Se cree que *TERRA* puede actuar global o localmente en los telómeros como un inhibidor directo de la telomerasa, al interactuar con HnRNP A1 y POT1 para promover el capping telomérico y preservar la integridad genómica. En células cancerosas se ha observado que los altos niveles de metilación sub-telomérica resultan en una menor expresión de *TERRA*, lo cual causa una desregulación en el alargamiento de los telómeros (Gutschner & Diederichs, 2012).

Angiogénesis sostenida

La vascularización es crucial en la sobrevivencia y función celular, al proveer oxígeno y nutrientes y permitir la evacuación de desechos metabólicos y dióxido de carbono. Esta necesidad obliga a todas las células en un tejido a residir a 100 μ m o menos de un capilar sanguíneo. Durante la embriogénesis, la vascularización involucra el nacimiento de nuevas células endoteliales y su ensamblaje en tubos, además del brote de nuevas ramificaciones partiendo de las existentes o angiogénesis. Usualmente, estos procesos se restringen al desarrollo embrionario, pero puede reactivarse bajo condiciones específicas en el organismo adulto, como en la cicatrización (Gutschner & Diederichs, 2012; Hanahan & Weinberg, 2011).

Durante la progresión tumoral, la angiogénesis puede reactivarse y permanecer encendida, causando un continuo brote de nuevos vasos sanguíneos que ayudan a sostener y expandir crecimientos neoplásicos. Esto se logra porque las células inducen factores pro-angiogénicos y bloquean señales anti-angiogénicas. Los reguladores de este proceso son proteínas señalizadoras que se unen a receptores de membrana presentes en la superficie de las células vasculares endoteliales (Gutschner & Diederichs, 2012; Hanahan & Weinberg, 2000; Hanahan & Weinberg, 2011).

Recientemente, se reportó un lncRNA llamado *MVIH* (asociado a invasión microvascular en carcinoma hepatocelular) sobre-expresado en tejidos hepáticos cancerosos. *MVIH* se asocia a la fosfoglicerato cinasa 1 (PGK1), enzima que además de catalizar un paso en la glucólisis se puede secretar por células tumorales y contribuir a la inhibición de la angiogénesis. La inducción de la angiogénesis se inicia en una etapa temprana en los carcinomas hepatocelulares y se piensa que debe existir una regulación negativa de PGK1 para promover la progresión tumoral. El incremento en la expresión de *MVIH* correlaciona con el decremento en la secreción de PGK1, que a su vez activa la angiogénesis; sin embargo, se desconoce el mecanismo exacto por el cual actúa. Interesantemente, se reportó que *MVIH* puede servir para predecir el pronóstico de pacientes y se observó que los niveles de este lncRNA correlacionan positivamente con la densidad microvascular (Yuan, *et al.*, 2012).

Invasión de tejidos y metástasis

Tarde o temprano durante el desarrollo de la gran mayoría de los cánceres humanos, algunas células que surgen de las masas tumorales primarias se disocian del tumor y se dirigen a explorar otros tejidos, viajando a lugares distantes para establecer nuevas colonias (Hanahan & Weinberg, 2000). La cascada de invasión y metástasis consiste en múltiples pasos discretos que siguen una sucesión de cambios en la biología de la célula (Hanahan & Weinberg, 2011).

Se comienza con la invasión local, seguida de la invasión por la membrana basal hacia los vasos sanguíneos y linfáticos, el tránsito de células cancerosas por los sistemas linfático y hemático, siguiendo por el escape de células por el lumen de los vasos hacia el parénquima de tejidos distantes (Hanahan & Weinberg, 2011). Durante el tránsito por estos sistemas, las células malignas deben escapar la vigilancia del sistema inmune y mostrar capacidades de crecimiento/supervivencia independiente (Gutschner & Diederichs, 2012). Después, se comienzan a formar pequeños nódulos de células cancerosas (micro-metástasis) que finalmente crecen para convertirse en tumores macroscópicos; este último paso se denomina colonización (Hanahan & Weinberg, 2011).

Un ejemplo de un lincRNA que regula la progresión metastásica es *HOTAIR*, donde su expresión ectópica en un contexto de células cancerosas parece remodelar el estado de la cromatina, permitiendo así la expresión de programas de expresión génica que conducen a la motilidad celular y a la invasividad (Gupta, *et al.*, 2010). Otro ejemplo es *MALAT1* (Transcrito 1 de adenocarcinoma de pulmón asociado a metástasis), que también se conoce como *NEAT2*. La expresión de *MALAT1* aumenta en muchos tipos de cáncer (Prensner & Chinnaiyan, 2011); sin embargo, la manera en que este gen se involucra con la metástasis no se termina de entender. Este lincRNA ha servido de marcador pronóstico para la metástasis y supervivencia de pacientes en cáncer pulmonar.

MALAT1 es muy abundante en diversos tipos celulares y se encuentra altamente conservado en la evolución, mide 8 kb y puede ser procesado en RNAs pequeños citoplásmicos parecidos a los tRNAs (Gutschner & Diederichs, 2012). *MALAT1* se retiene en el núcleo y forma parte de los cuerpos espiralados o speckles, estructuras que juegan un papel en el procesamiento de los pre-mRNAs. Este lincRNA parece secuestrar factores de procesamiento ricos en residuos de serina y arginina hacia estas estructuras nucleares y la desregulación en la expresión de *MALAT1* también altera un conjunto de pre-mRNAs (Cheetham, *et al.*, 2013). Asimismo, se reportó que este RNA puede interactuar con CBX4 para controlar la relocalización de genes que controlan el crecimiento, al dirigirlos de los cuerpos de Polycomb hacia gránulos de intercromatina, es decir, llevarlos de sitios sin expresión génica a lugares donde se activa (Gutschner & Diederichs, 2012; Wahlestedt, 2013). Algunas pistas que nos pueden guiar en la función de *MALAT1* son que este transcrito promueve la movilidad en las células cancerosas de pulmón por regulación transcripcional y post-transcripcional de genes relacionados con movimiento. También se sabe que este lincRNA mantiene la proliferación e invasión de células cancerosas cervicales. Esto relaciona a *MALAT1* en funciones de proliferación, apoptosis, migración y regulación génica, abriendo diversas posibilidades para su función (Gutschner & Diederichs, 2012).

En conclusión, si se desean entender a los mecanismos subyacentes a la carcinogénesis, metástasis y resistencia a fármacos, se debe de considerar a los lincRNAs, los cuales añaden un nuevo nivel de complejidad (Fig. 13). Una manera en que se ha propuesto empezar a estudiar a un lincRNA que presente cambios en su expresión en un contexto tumoral es mediante análisis de pérdida y ganancia de función. Experimentos subsecuentes de fenotipos tumorales relevantes, como la proliferación, migración, viabilidad celular o apoptosis, pueden proveer un primer acercamiento a su función (Gutschner & Diederichs, 2012). Siguiendo con la propuesta anterior, este trabajo propone la identificación y el estudio de un lincRNA

desregulado en cáncer mamario para proveer indicios de su función mediante este tipo de análisis.

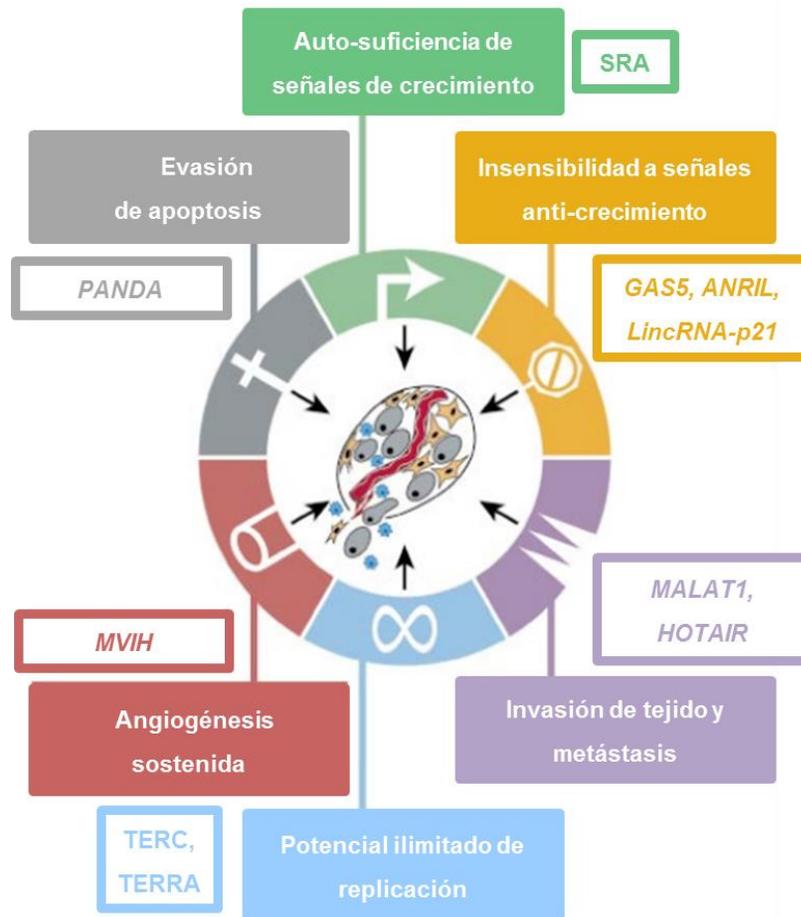


Figura 13: Seis alteraciones encontradas en el cáncer y ejemplos de lncRNAs cuya función se ve involucrada en cada proceso. En los recuadros del mismo color se mencionan ejemplos de lncRNAs que participan en cada una de estas alteraciones. Modificado de Hannahan & Weinberg, 2000.

ANTECEDENTES PARTICULARES

Estudio de perfiles de expresión de lncRNAs desregulados en diversos tipos de cáncer

Actualmente, se conocen diversos ejemplos de lncRNAs desregulados en cáncer, como los mencionados anteriormente. Sin embargo, no existe una investigación clínica sistemática sobre la expresión de lncRNAs en tejidos cancerosos. Sumado a esto, las funciones de la mayoría de estos transcritos y su relación con el cáncer están lejos de ser claros (García Solís, 2006). No obstante, algunos estudios recientes han tomado ventaja del gran número de microarreglos y librerías del transcriptoma reportadas para comenzar a establecer mapas -o atlas- que muestren la expresión de los lncRNAs en múltiples tejidos. Aunque la intención original de estas mediciones no era buscar a los lncRNAs, estos conjuntos de datos pueden ser reanalizados para comenzar a establecer perfiles de expresión génica de este tipo de transcritos, partiendo de cientos de muestras tumorales (Du, *et al.*, 2013; Gibb, *et al.*, 2011).

Un estudio pionero, llevado a cabo en 2011 por el grupo del Dr. Lam en la Universidad de Columbia Británica, Canadá, tuvo como objetivo entender qué tan extensa es la contribución de la expresión diferencial de los lncRNAs para el cáncer (Gibb, *et al.*, 2011). En este estudio, se analizaron 272 librerías humanas tipo SAGE para delinear los patrones de transcripción de los lncRNAs a través de un amplio espectro de tejidos normales y cancerosos (Fig. 14). Cabe destacar que en este trabajo se interesaron principalmente en identificar lncRNAs previamente anotados, sobretodo de la base de datos europea *Ensembl*, más que en la identificación de nuevos genes. Resulta importante precisar que SAGE es un método basado en la secuencia para enumerar la abundancia de los transcritos poliadenilados (Gibb, *et al.*, 2011).

Se integraron perfiles de expresión de 26 tejidos humanos normales, 19 tumorales y 9 líneas celulares cancerosas (Gibb, *et al.*, 2011). Dentro de sus descubrimientos, ellos reportan que existe una expresión extensa, tipo tejido específica de los lncRNAs en tejidos normales, que se ve significativamente alterada durante el cáncer. Sumado a la construcción de los perfiles de expresión en diversos tejidos humanos, en este trabajo pudieron determinar los niveles de expresión diferencial de lncRNAs en tejidos de glándula mamaria, cerebro y pulmón. Esto es debido a que se contaba con cinco o más librerías tipo SAGE en contexto normal y canceroso para cada uno de estos tejidos, permitiéndoles hacer inferencias con soporte estadístico. En sus resultados muestran algunos de los lncRNAs que presentan la mayor alteración en sus niveles de expresión de tejido normal a tumoral en estos 3 tejidos (Fig. 14) (Gibb, *et al.*, 2011).

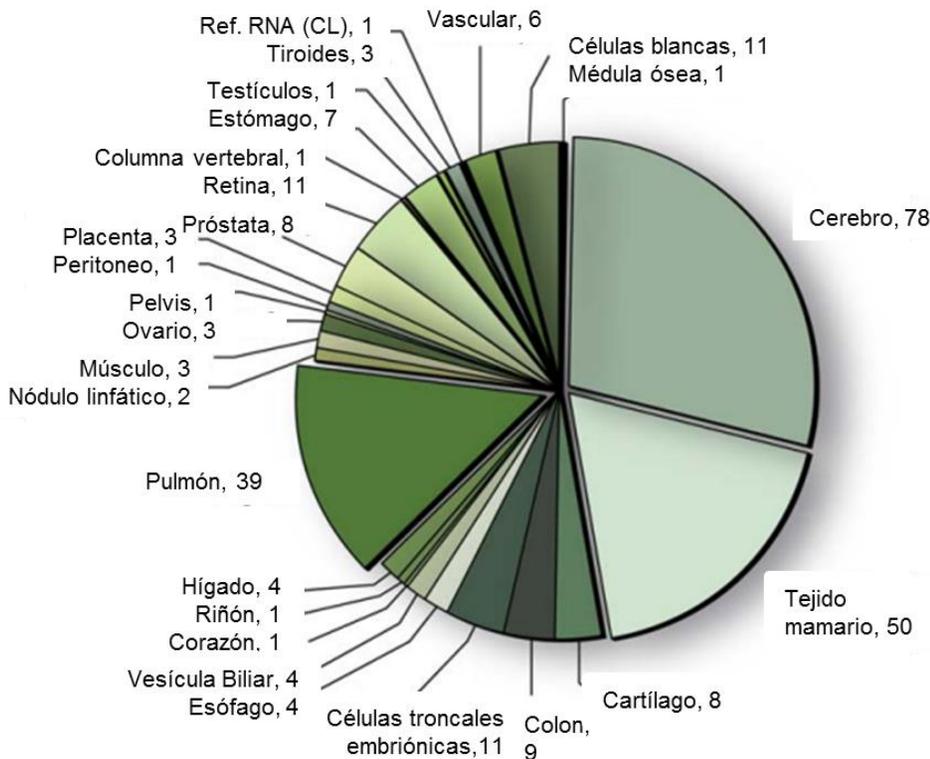


Figura 14: Distribución de los tipos de tejidos utilizados para construir el atlas de lncRNAs. Se reportó el uso de 272 librerías de SAGE, para las cuales se analizaron los perfiles de lncRNAs. En el caso de pulmón, cerebro y tejido mamario, existen suficientes experimentos para poder analizar cuáles transcritos cambian de un tejido normal a uno canceroso. Modificado de Gibb, *et al.*, 2011.

Un segundo estudio a destacar, reanalizó 1,300 tejidos tumorales a partir de microarreglos en cuatro tipos de cáncer. Dado que ellos construyeron perfiles de expresión para 10,207 lncRNAs, esta base de datos provee un recurso para identificar lncRNAs clínicamente relevantes (Du, *et al.*, 2013). Se espera que los lncRNAs puedan proveer una serie de aplicaciones una vez que se profundice en su caracterización, por ejemplo: funcionar como biomarcadores novedosos para el diagnóstico y pronóstico en diversos tipos de cáncer, además de servir como blancos específicos para terapias o para predecir la respuesta de diversos fármacos (Du, *et al.*, 2013; Gutschner & Diederichs, 2012; Qiu, *et al.*, 2013).

Tomando ventaja de la información generada en el trabajo del atlas de los lncRNAs (Gibb, *et al.*, 2011), el estudio descrito en esta tesis utilizó la base de datos generada (los perfiles de expresión del tejido mamario normal y canceroso) en conjunto con criterios adicionales para seleccionar un sub-conjunto de lncRNAs. Más adelante se ahondará en la descripción de la búsqueda de estos candidatos y se mostrarán los resultados.

Estudio de asociación de lncRNAs a complejos de regulación epigenética

Uno de los primeros estudios, que tenía como objetivo examinar el mecanismo bioquímico por el cual funcionan los lncRNAs, fue realizado en 2009 por el grupo de John Rinn (Khalil, *et al.*, 2009). Este estudio estuvo inspirado en *HOTAIR* y su interacción con PRC2, sugiriendo la

posibilidad de que muchos lincRNAs se asociaran físicamente a complejos de regulación epigenética y que actuaran sobre regiones específicas del genoma. Con la finalidad de analizar esta hipótesis de forma global, se realizó una co-inmunoprecipitación de RNA seguida de análisis por microarreglos (RIP-CHIP). Específicamente, se buscó identificar a los lincRNAs que se asociaban a algún complejo de regulación epigenética como PRC2 (en sus proteínas EZH2 y SUZ12) o a CoREST. Entre sus resultados se encontró que el 38% de los transcritos estudiados se asocian a alguno de estos complejos, en particular el 24% se precipitaba con el anticuerpo para PRC2. Ellos también sugieren que parece que los transcritos asociados a este complejo podrían ejercer mecanismos de guía en *trans* (Khalil, *et al.*, 2009).

Cabe mencionar que el estudio se realizó al principio de la era de los lincRNAs, por lo que en ese momento se muestrearon sólo 1,147 transcritos. Dado que no se consideraron otros complejos, se sugiere que este número puede incrementarse considerablemente. Por ejemplo, en este trabajo la atención se centró en los lincRNAs asociados a complejos represivos y no se consideraron los transcritos unidos a MLL o a G9a, que son complejos activadores asociados al grupo de proteínas TrxG (Chu, *et al.*, 2011; Tsai, *et al.*, 2010). Siguiendo con lo anterior, hay que tener en cuenta que también existen ejemplos de transcritos que actúan en conjunto con proteínas de activación de la transcripción, como *HOTTIP* (Wang, *et al.*, 2011).

En años recientes, se han publicado reportes de técnicas novedosas que proponen diversas estrategias para contestar interrogantes relacionadas con la identificación de las proteínas que se asocian a un lincRNA o hacia la búsqueda de los sitios de unión a nivel genómico de transcritos específicos. Las técnicas que más resaltan son el ChIRP y CHART; ambas consisten en el mapeo *in vivo* de células para identificar los sitios de unión de lincRNAs a escala de genomas completos. Primero, logran enriquecer a un lincRNA de interés, junto con sus moléculas asociadas; una vez recuperado el transcrito enriquecido, se puede secuenciar para saber los sitios del genoma donde se unió y complementar esta metodología con técnicas proteómicas para identificar las proteínas asociadas (Chu, *et al.*, 2011; Simon, *et al.*, 2011).

PLANTEAMIENTO DEL PROBLEMA

En la actualidad, existe evidencia que incorpora a los lncRNAs como componentes esenciales del transcriptoma y asocia a estas moléculas en diversos procesos celulares en los organismos eucariontes. Dada la relevancia biológica de estos transcritos, es entendible que existan diversos ejemplos que asocien la desregulación de los lncRNAs con la génesis y patología de diversos tipos de cáncer. Sumado a esto, se ha demostrado que estos transcritos son relevantes en la regulación epigenética por medio de diversos mecanismos, como el de guía para complejos proteicos que regulan a la cromatina, a manera de dirigirlos hacia sus blancos génicos. La alteración de los niveles de expresión en los lncRNAs que se involucren en alguna clase de regulación epigenética podría tener consecuencias globales graves, como su relocalización genómica inespecífica. Por estas razones, resulta importante identificar y estudiar lncRNAs cuyos niveles de expresión se alteren en un contexto canceroso y que además se hayan reportado asociados a complejos de regulación epigenética. En particular, centramos nuestros esfuerzos en cáncer de glándula mamaria. Es así como buscamos entender mejor la función de algún lncRNA y cómo su desregulación impacta a distintos niveles y en diversos procesos celulares.

HIPÓTESIS

Existe una relación entre la desregulación de los transcritos largos no-codificantes, el reclutamiento de complejos de regulación epigenética y la carcinogénesis mamaria, lo cual podría ser debido a que estos transcritos pierden la capacidad de guiar a dichos complejos a sus diversos blancos génicos. Como consecuencia, existirá una expresión anormal de genes a lo largo del genoma, pudiendo impactar en distintos procesos celulares.

OBJETIVOS

Objetivo General

Identificar diversos lncRNAs humanos que hayan sido reportados con asociación a complejos de regulación epigenética y que muestren cambios en sus niveles de expresión en tejidos tumorales mamarios. Después de esta primera etapa, se elegirá algún lncRNA candidato y se buscará entender su función en *cis* por medio de ensayos de ganancia y pérdida de función.

Objetivos Particulares

1. Identificar por una búsqueda computacional a lncRNAs bajo distintos criterios. Partiendo de los resultados, seleccionar un transcrito candidato y realizar una caracterización *in silico* del mismo.
2. Cuantificar la expresión del transcrito largo no-codificante candidato en distintas líneas celulares de tejido mamario.
3. Definir si el lncRNA tiene una acción tipo *cis* e identificar sus blancos génicos.
4. Realizar ensayos de pérdida y ganancia de función para el lncRNA candidato y determinar si existen cambios en la expresión de sus genes blancos.
5. Con base en los mismos ensayos de ganancia y pérdida de función, se evaluará su efecto en la fisiología celular.

MODELO DE ESTUDIO

Dado que uno de los criterios de selección de este transcrita consistió en identificar lncRNAs que mostraran un cambio de expresión de tejido normal a tumoral mamario, consideramos fundamental escoger un contexto apropiado para su estudio. Por tal razón, se estableció un modelo con cultivos celulares humanos como un primer acercamiento, con posibilidad de expandirse a muestras clínicas en un futuro. Dicho modelo consiste en 1 línea celular de epitelio mamario normal (MCF-10A) y 4 líneas celulares tumorales de tejido mamario (MCF7, ZR-75-30, T47-D y MDA-MB-231), las cuales se ordenaron según su grado de invasividad (Fig. 15) (Neve, *et al.*, 2006).

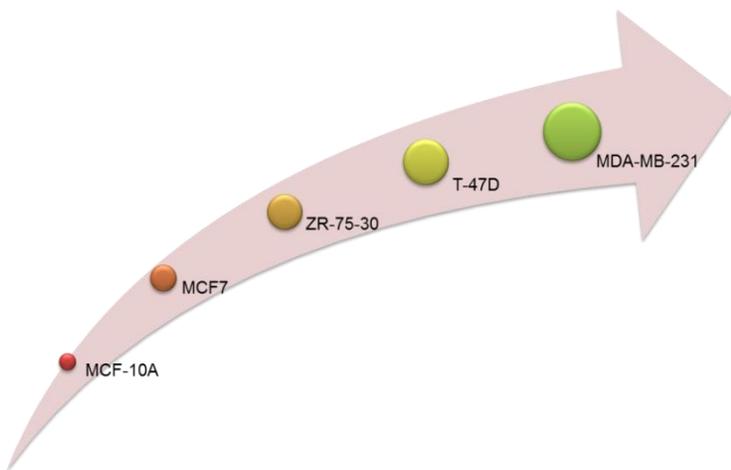


Figura 15: Líneas celulares tumorales de origen mamario utilizadas en el presente estudio.

El modelo consiste en 5 líneas celulares, 1 transformada normal MCF-10A (epitelial) y 4 líneas tumorales con distinta capacidad invasiva: MCF7, ZR-75-30 (luminales, no invasivas), T-47D (luminal, ligeramente invasiva) y MDA-MB-231 (basal, altamente metastásica).

Las células MCF-10 se derivaron de un paciente con mastopatía fibroquística y se inmortalizaron después de cultivarlas en un periodo de tiempo extendido, a baja concentración de calcio, para generar a la línea MCF-10A (Soule, *et al.*, 1990). MCF-10 es diploide, mientras que la línea MCF-10A tiene un cariotipo estable de 47 cromosomas con ganancia en los loci 5q23.1-35.3, 19q13.11-q13.43, 13q32.1-p32.2 y una copia adicional del cromosoma 8. También tiene pérdida de los loci 9p21.3 (que conllevó a la inmortalización de esta línea celular), 3p26.3, 16p11.2, 21p11-q11.2 y 22q11.1 (Marella, *et al.*, 2009). Las células MCF-10A expresan normalmente a p53, no forman colonias en agar suave y no proliferan en ratones inmunosuprimidos (Imbalzano, *et al.*, 2009; Soule, *et al.*, 1990). Además, esta línea desarrolla características importantes del tejido mamario normal cuando se cultiva en una membrana tridimensional y su crecimiento está controlado por hormonas y factores de crecimiento (Imbalzano, *et al.*, 2009; Soule, *et al.*, 1990). Al no presentar alguna característica de invasividad o formación tumoral, se considera a MCF-10A como una línea celular normal epitelial mamaria (Marella, *et al.*, 2009).

La línea celular MCF7 proviene de un adenocarcinoma maligno mamario, extraído de líquido pleural de una mujer caucásica de 69 años de edad (Altogen Biosystems). Es una línea luminal no invasiva, según se observó con ensayos de invasividad por cámara de Boyden (Neve, *et al.*, 2006). Esta línea ha retenido múltiples características particulares del epitelio mamario, incluyendo la habilidad de procesar estrógeno en la forma de estradiol por medio de sus receptores (ER+) y también posee PR+. MCF7 es capaz de formar domos cuando se crece en cultivo y posee un cariotipo que varía entre 66 y 87 cromosomas (Altogen Biosystems; Neve, *et al.*, 2006).

Las células ZR-75-30 se recolectaron de fluido ascítico (líquido seroso que existe entre el peritoneo visceral y parietal) de una mujer negra de 47 años con carcinoma ductal invasivo (Engel, *et al.*, 1978). Esta línea celular tiene aproximadamente 81 cromosomas, con dos grandes co-amplificaciones en los cromosomas 8 y 17; no es invasiva con ensayos de cámara de Boyden y no expresa a p53 (Engel, *et al.*, 1978; Neve, *et al.*, 2006).

La línea celular T47-D se estableció a partir de un derrame pleural de una paciente de 54 años con carcinoma mamario ductal metastásico. Es una línea casi triploide, con aproximadamente 66 cromosomas y con múltiples translocaciones. Expresa tanto a ER como PR y tiene a p53 mutado (American Type Culture Collection, 2014).

Finalmente, la línea celular MDA-MB-231 fue establecida a partir de efusión pleural de una mujer caucásica de 51 años con adenocarcinoma. Tiene un número de cromosomas que varía de 52 a 68, acercándose a ser triploide (Schwartz & Pirrotta, 2013). Es una línea celular altamente invasiva, siendo la que presentaba mayor número de células en análisis por cámara de Boyden de entre 30 líneas derivadas de tejido mamario (incluyendo a todas las de este modelo) (American Type Culture Collection, 2014).

RESULTADOS

1. Identificación *in silico* de un lncRNA con asociación a complejos epigenéticos que muestra niveles de expresión alterados en tejido mamario canceroso

En la actualidad, existe una fuerte evidencia de que los lncRNAs pueden participar de manera importante en la génesis y mantenimiento del cáncer; asimismo, se ha demostrado que estos transcritos tienen gran relevancia en la regulación epigenética. Por estas razones, nuestro interés inicial se basó en identificar lncRNAs con asociación a complejos de regulación epigenética, que también se hayan reportado entre los que más cambian su expresión en tejido mamario tumoral.

Para poder obtener algunos candidatos, se realizó una búsqueda bioinformática basada en dos estudios. Primero se consideraron los resultados de los estudios de Khalil, A., *et al.*, 2009, contenidos en la base de datos de Cabili, M., *et al.*, 2011, que indicarían si existe una interacción demostrada experimentalmente con un complejo represor de la transcripción (PRC2 en este caso). En segundo lugar, se buscó que el lncRNA estuviera contenido en el atlas construido por Gibb, E., *et al.*, 2011. La intersección de estas dos bases de datos nos dio una lista con las coordenadas de 34 lncRNAs (Tabla 1).

Nombre	Chr	Sitio de inicio Gibb	Sitio de Término	Sitio de inicio Cabili	Sitio de Término	+/ -	Clase	Valor Normal	Valor Cáncer
RP11-371I1.2	1	238643684	238649323	238643685	238649317	-1	lincRNA	3.48	0.46
RP11-177F15.1	1	238648745	238701605	238643685	238649317	1	processed_t	1.28	0.55
AC122136.1	2	216476286	216708445	216397855, 216582765	216578618, 16584147	-1	processed_t	1.97	8.92
RP11-446H18.3	3	106959539	107045811	106959538	107045810	1	lincRNA	16.56	2.61
AC069513.3	3	195415518	195451630	195435001	195438746	1	processed_t	0	4.17
AC093323.3	4	6675178	6677774	6672409	6675587	1	lincRNA	6.7	12.18
CTD-2201E18.3	5	43014516	43067521	43066161	43067888	-1	lincRNA	5.37	1.95
RP11-65F13.2	5	92746995	92921354	92906434	92910714	-1	lincRNA	4.47	7.9
AL031123.2	6	6695027	6700720	6659634	6707166	-1	lincRNA	7	6.21
RP5-1120P11.3	6	44041650	44045689	44041649	44045689	1	processed_t	2.53	1.91
AC005154.6	7	30555925	30604233	30586781	30590397	-1	processed_t	7.28	8.38
AC004945.2	7	79082198	79100524	79085480	79096779	1	processed_t	29.34	11.39
PVT1	8	128806779	129113499	128988493	129009287	1	lincRNA	0.94	3.17
NCRNA00051	8	143279717	143290364	143279716	143290362	1	processed_t	10.08	7.52
RP11-492E3.1	9	132245730	132275947	132262771	132270807	1	lincRNA	17.38	16.51
RP11-202P11.1	10	80703085	80828817	80703082	80737963	-1	lincRNA	0	2.24
CTC-497E21.4	11	12984057	13011095	12995666	13011095	-1	lincRNA	4.4	6.23
RP11-113K21.5	11	82783108	82817761	82783118	82802318	1	lincRNA	0	1.92
RP5-1096D14.2	12	2038368	2045740	2038367	2045740	-1	lincRNA	1.87	0.43
AC124303.3	15	25224704	25620623	25247917, 25355583, 25361691	25281705, 25361586, 25367621	1	processed_t	4.08	3.57
AC021755.4	15	40331512	40359491	40331486	40366005	1	lincRNA	1.41	1.08
AC009108.1	16	86508135	86542466	86508050, 86513746	86542466, 86519552	-1	lincRNA	1.25	3.41
AC007639.1	17	70380443	70588943	70380442	70588943	-1	lincRNA	1.25	2.8
AC006504.2	19	28281402	28284848	28192407	28285209	-1	lincRNA	9.06	6.01
RP4-564F22.2	20	37049235	37063996	37062992	37063976	-1	lincRNA	91.43	54.18
SNHG11	20	37075221	37079564	37075529	37076654	1	lincRNA	4.68	4.67
AL109616.1	21	26931716	26936957	26931715	26936957	-1	lincRNA	7.89	17.1
AP001046.5	21	44778027	44782229	44777429	44782327	-1	lincRNA	0	1.36
NCRNA00319	21	44886481	44873773	44859454	44900667	1	processed_t	1.04	0.91
NCRNA00313	21	44881974	44899414	44859454	44900667	-1	processed_t	0.89	0
AP001048.4	21	44885189	44887178	44859454	44900667	1	processed_t	0	0.88
TUG1	22	31365643	31374831	31365261	31375377	1	lincRNA	29.48	50.59
XIST	X	73040486	73072588	73040449	73072588	-1	lincRNA	197.07	266.99
JPX	X	73164177	73220326	73167079	73225243	1	processed_t	6.39	4.17

Tabla 1: Lista de coordenadas de lncRNAs con una interacción al grupo de proteínas represoras PRC2 y alteraciones en sus niveles de expresión. Nótese que se muestran los valores obtenidos de tejidos normales y cancerosos. De manera importante, XIST aparece en esta lista, dando soporte a los otros lncRNAs encontrados.

El último criterio de selección fue que el lincRNA se desregulara en cáncer de mamá, ya que nuestro laboratorio se encuentra dentro de un consorcio cuyo fin es abordar preguntas relacionadas con este tipo de cáncer. Una vez cumplidos los criterios de selección, se identificaron tres transcritos: *RP11-446H18.3*, *SNHG17* y *AC122136.1*. De estos 3 lincRNAs, se eligió a *RP11-446H18.3*, el cual muestra la mayor tasa de cambio si comparamos tejido normal con tejido tumoral mamario. Para facilitar su descripción, decidimos denominarlo *lincRNA_mr* a lo largo de este estudio. Asimismo, este transcrito se enlistaba entre los 20 más alterados en tejido mamario según el atlas de lincRNAs en cáncer (Gibb, *et al.*, 2011), disminuyendo 7 veces su expresión. Por estas razones, decidimos preguntarnos qué se había reportado acerca de este lincRNA.

1.1 Características de *lincRNA_mr*

LincRNA_mr se expresa en diversos tejidos normales, con niveles que varían de un tipo celular a otro. Se ha reportado que existe expresión en tejido mamario, pero se observó que sus niveles se reducen en casi siete órdenes de magnitud en una serie de cánceres mamaros (Gibb, *et al.*, 2011). Las coordenadas de este lincRNA, según el ensamblaje GRCh37/hg19, son Chr3:106,959,539-107,045,811. Este lincRNA recibe nombres como *RP11-446H18.3*, *linc00883*, *ENSG00000243701*, *LOC344595* y *NR_028301* dependiendo de la base de datos en la cual se encuentre. Aunque el transcrito está anotado, se desconoce su función, genes blanco o mecanismo de acción.

El tamaño del *lincRNA_mr* es de 86.47 kb con todo e intrones; posee 7 exones pero se reporta que se procesa en 5 isoformas con tamaños de 1149 nt (*lincRNA_mr1*), 2737 nt (*lincRNA_mr2*), 738 nt (*lincRNA_mr3*), 801 nt (*lincRNA_mr4*) y 411 nt (*lincRNA_mr5*) (Fig. 16). Asimismo, validamos experimentalmente que *lincRNA_mr* es un transcrito que presenta poli-(A) -al poder amplificarlo partiendo de una RT-PCR con oligo-d(T). Los dos genes codificantes adyacentes al *lincRNA_mr* son el gen *cbl-b*, localizado a 1.3 Mb río arriba y el gen *ccdc54*, que se encuentra a 50 kb río abajo.

Interesantemente, el *lincRNA_mr* parece tener en su organización genómica una co-regulación con un RNA largo no-codificante divergente *-linc00882-* localizado a tan solo 53 bp río arriba de su sitio de inicio de la transcripción (Fig. 16). La corta distancia existente entre ambos lincRNAs fue un hecho que captó nuestra atención, ya que sugería la presencia de un promotor común para ambos transcritos. Por tal razón, se propuso realizar un análisis *in silico* de esta región para comenzar a elucidar los elementos regulatorios de este locus y averiguar cómo lucen el o los promotores de estos transcritos.

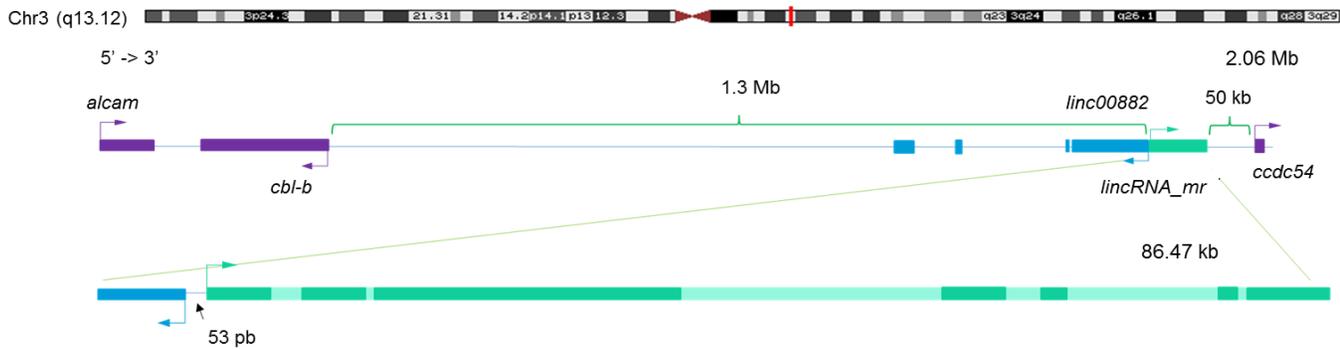


Figura 16: Mapa de la localización genómica de *lincRNA_mr*. Arriba se puede apreciar el locus Chr3 (q13.12). En color morado se muestran los genes codificantes *alcam*, *cbl-b* y *ccdc54*; en azul cielo está *linc00882*, divergente a *lincRNA_mr*, que se muestra en color verde. Al hacer un acercamiento se puede ver el lincRNA que se estudiará en este trabajo y que mide 86 kb, las 53 bp que separan a los dos transcritos y el inicio de *linc00882*.

1.2 Caracterización *in silico* de los elementos regulatorios del *lincRNA_mr*

Con la finalidad de estudiar a los elementos regulatorios responsables de la regulación transcripcional del *lincRNA_mr*, se realizó una investigación *in silico*, utilizando experimentos reportados en bases de datos genómicas como ENCODE, que se visualizaron con ayuda del UCSC genome-browser (UCSC GB). En este ensayo, se analizaron 2,400 bp que contenían la región con las 53 bp que separaban a los transcritos y el comienzo de ambos lincRNAs. Los elementos analizados consisten en las modificaciones post-traduccionales H3K4me3, H3K27ac, presencia de RNA Pol II, hipersensibilidad a la DNasa I, metilación de DNA y estados de la cromatina, según los reportes de Ernst, J., *et al.*, 2011 (Fig. 17) (Ernst, *et al.*, 2011).

Entre los resultados observados resaltan los estados de la cromatina de esta región, los cuales indican que la expresión de estos transcritos es diferencial y depende del tejido o línea celular, como se ha reportado en la literatura. Esto se puede observar en la figura 17, donde se muestran unas barras horizontales con un código de colores que representan los estados de la cromatina en 9 líneas celulares (Ernst, *et al.*, 2011). Si bien, algunas de ellas indican una cromatina abierta y transcripción activa, otras muestran silenciamiento y un promotor inactivo. Si se analiza el resto de la figura 17, en especial la parte superior, se siguen obteniendo indicios de cromatina abierta -como la hipersensibilidad a la DNasa I- pero también metilación del DNA, que denota un posible silenciamiento transcripcional en algunas líneas celulares. Esto es contundente con lo descrito arriba. De esta manera, existen líneas donde se percibe una transcripción activa, con RNA Pol II unida en esta región y enriquecimiento de la marca

H3K4me3, mientras que en otros contextos celulares se presentará mediana expresión o no la habrá.

chr3:106,958,771-106,961,170

2.4 kb

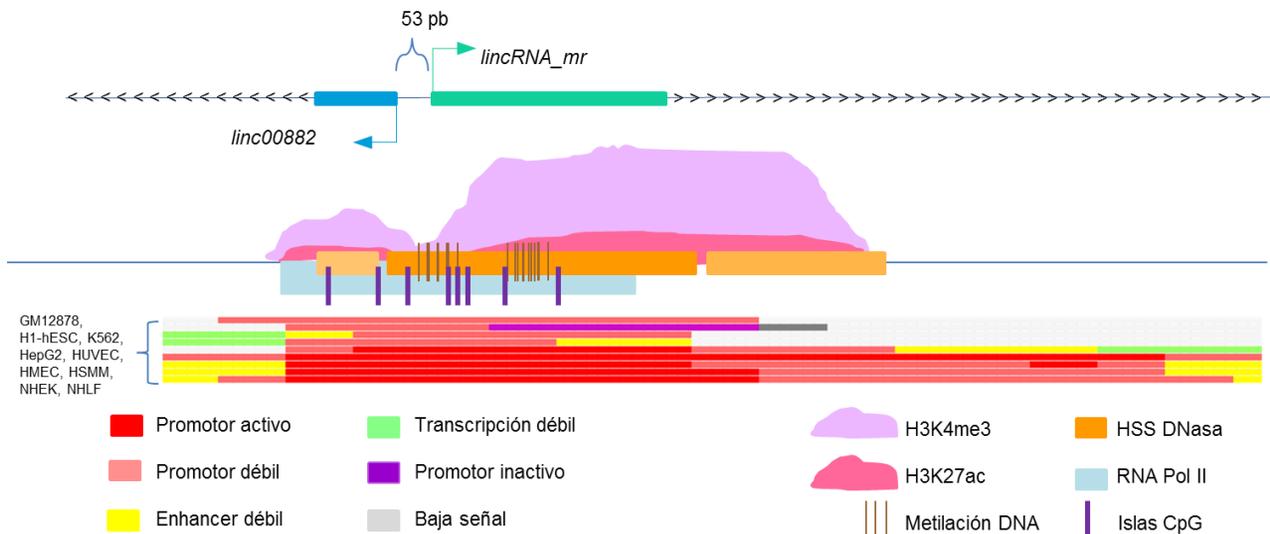


Figura 17: Los elementos regulatorios en la posible región promotora del *lincRNA_mr* y *linc00882*.

En esta figura se ilustran los resultados obtenidos al estudiar la región que contiene el inicio de los dos transcritos *linc00882* y *lincRNA_mr*. En la parte superior se observa un esquema del primer exón de cada lincRNA y un poco del primer intrón, en una región de 2.4 kb. Debajo de los lincRNAs se muestra un resumen de cómo se distribuyen en esta zona las modificaciones de las histonas H3K4me3, H3K27ac, la metilación del DNA, las islas CpG, sitios de hipersensibilidad a la Dnasa I y CHIP de unión a la RNA Pol II. Al tener a elementos asociados tanto con activación como con represión, se deduce que existen líneas donde existe transcripción y otros contextos celulares donde este sitio está silenciado. Debajo de estas marcas se muestran barras horizontales que representan a los estados de la cromatina correspondientes a las 9 líneas celulares enlistadas a la izquierda. De manera relevante, se pueden mencionar dos escenarios: uno donde hay un promotor común para ambos lincRNAs y otro donde existen dos promotores cercanos, cada uno regulando a un lincRNA.

El resultado más relevante es que la suma de los elementos analizados sugiere fuertemente la existencia de uno o varios promotores. En el caso de que existiera uno solo, sería común para ambos transcritos y esto podría indicar una co-regulación para estos lincRNAs. En el segundo caso, podrían existir dos promotores muy cercanos, cada uno regulando a uno de los transcritos no-codificantes. Interesados en la posibilidad de que exista una regulación similar para ambos transcritos y sabiendo que existen pocos estudios sobre la regulación transcripcional de los lincRNAs, nos preguntamos cuáles factores de transcripción (TFs) se unen a esta región. Para contestar esta incógnita, se tomaron dos enfoques: el primero

fue visualizar los datos de ChIP-seq elaborados por el ENCODE en el UCSC GB, la segunda fue realizar un análisis con JASPAR, una base de datos de libre accesos curada y no redundante.

En un primer análisis se utilizó la misma región de 2,400 bp que contiene a los dos lincRNAs y el sitio de 53 bp que los divide. Sin embargo, los resultados dieron una gama grande de posibles TFs unidos a partir de la base de datos JASPAR, por lo que para acotar esta lista se realizó un segundo análisis disminuyendo la región a estudiarse para tener tan solo 420 bp. En esta ocasión, se obtuvieron 7 TFs diferentes con JASPAR y 7 reportados por ENCODE. Al intersectar ambos resultados se obtuvieron 3 TFs en común: NRF1, USF2 y E2F4 (Fig. 18).



Figura 18: Sitios de unión de algunos factores de transcripción en el posible promotor bidireccional. A partir de un análisis con JASPAR y datos del consorcio ENCODE, se interrogaron 420 bp; en negritas y subrayado se muestran los TFs que se reportaron en ambos recursos: USF2, NRF1 y E2F4. Los TFs se dibujaron en la posición exacta donde se reporta el motivo al cual se unen y también indican el sentido de la cadena. En negro se muestra la secuencia del *linc00882*, en rojo las 53 bp de separación y en azul la secuencia de *lincRNA_mr*.

El tener una lista de TFs que posiblemente regulan a estos lincRNAs comienza a sugerir la manera en que este locus se regula. Por otra parte, para averiguar si como parte de la regulación de estos transcritos existían miRNAs involucrados, se corrió un análisis con el algoritmo miRBase (Kozomara & Griffiths-Jones, 2014). Este programa ayuda a predecir

secuencias de miRNA, por lo que se realizó un ensayo dentro del locus Chr3: q13.11-q13.12, el cual contiene a los 2 lincRNAs e incluye a los genes codificantes *cbl-b* y *ccdc54*. El resultado indica que no existen miRNAs codificados en esta región (datos no mostrados). A manera de predecir semillas de miRNAs dentro de las secuencias de estos genes, se utilizó el programa DIANA-microT v3.0 (Maragkakis, *et al.*, 2009). Los resultados reportaron que no existen secuencias que codifiquen miRNAs en este locus. Tampoco existen sitios blancos en el *lincRNA_mr*, el *linc00882* y el gen *ccdc54*, pero sí en el gen *cbl-b* para los miRNAs miR-30a, miR-30b, miR-30c, miR-30d, miR-30e, miR-101 y miR-27b. A este nivel, nuestros resultados sugieren que el *lincRNA_mr* no posee un mecanismo de acción tipo anzuelo de miRNAs.

Pensando en los otros modos por los cuales el *lincRNA_mr* puede ejercer su función, decidimos hacer un ensayo de interacción RNA-proteína *in silico* para seguir elucidando la posibilidad de que la asociación de este lincRNA con complejos de regulación epigenéticos resulte importante para que lleve a cabo su función.

1.3 Predicción computacional de interacción entre el *lincRNA_mr* y complejos epigenéticos

Una forma de analizar una posible asociación entre el *lincRNA_mr* y proteínas como las de PRC2 fue mediante el algoritmo catRAPID graphic. Este programa predice la propensión de interacción entre una pareja proteína-RNA, reportando la fuerza de interacción con respecto a evidencia experimental de interacción, que sirve de referencia. Entre los parámetros que se toman en cuenta para calcular el perfil de interacción están la estructura secundaria del RNA, puentes de hidrógeno y fuerzas de van der Waals (Bellucci, *et al.*, 2011).

Debido a que PRC2 es el complejo reportado con interacción al *lincRNA_mr* por RIP-ChIP (Khalil, *et al.*, 2009), se realizó un ensayo bioinformático con el algoritmo catRAPID graphics (Bellucci, *et al.*, 2011) para 4 de las distintas isoformas del *lincRNA_mr* y las 4 proteínas canónicas de PRC2 (EZH2, SUZ12, EED y RBAP48). Los resultados muestran una potencial asociación fuerte para la mayoría de las isoformas, con distintos valores de interacción dependiendo del transcrito y de la proteína analizadas (Fig. 19a).

Una limitante de este algoritmo es que sólo se permite introducir secuencias de entre 50-1,200 nt, por lo que se utilizó el algoritmo catRAPID omics para entender con cuáles proteínas se predice asociación para el *lincRNA_mr2*, que mide 2.7 kb. Este programa predice la fuerza de interacción entre un transcrito y todo el proteoma humano. Se reportó que la interacción más fuerte para el *lincRNA_mr2* se da con SUZ12 (Fig. 19b). Interesantemente, cuando se corrió catRAPID omics con las demás isoformas del *lincRNA_mr* no se observó que alguna de las proteínas de PRC2 estuviera contenida en las listas de interacción.

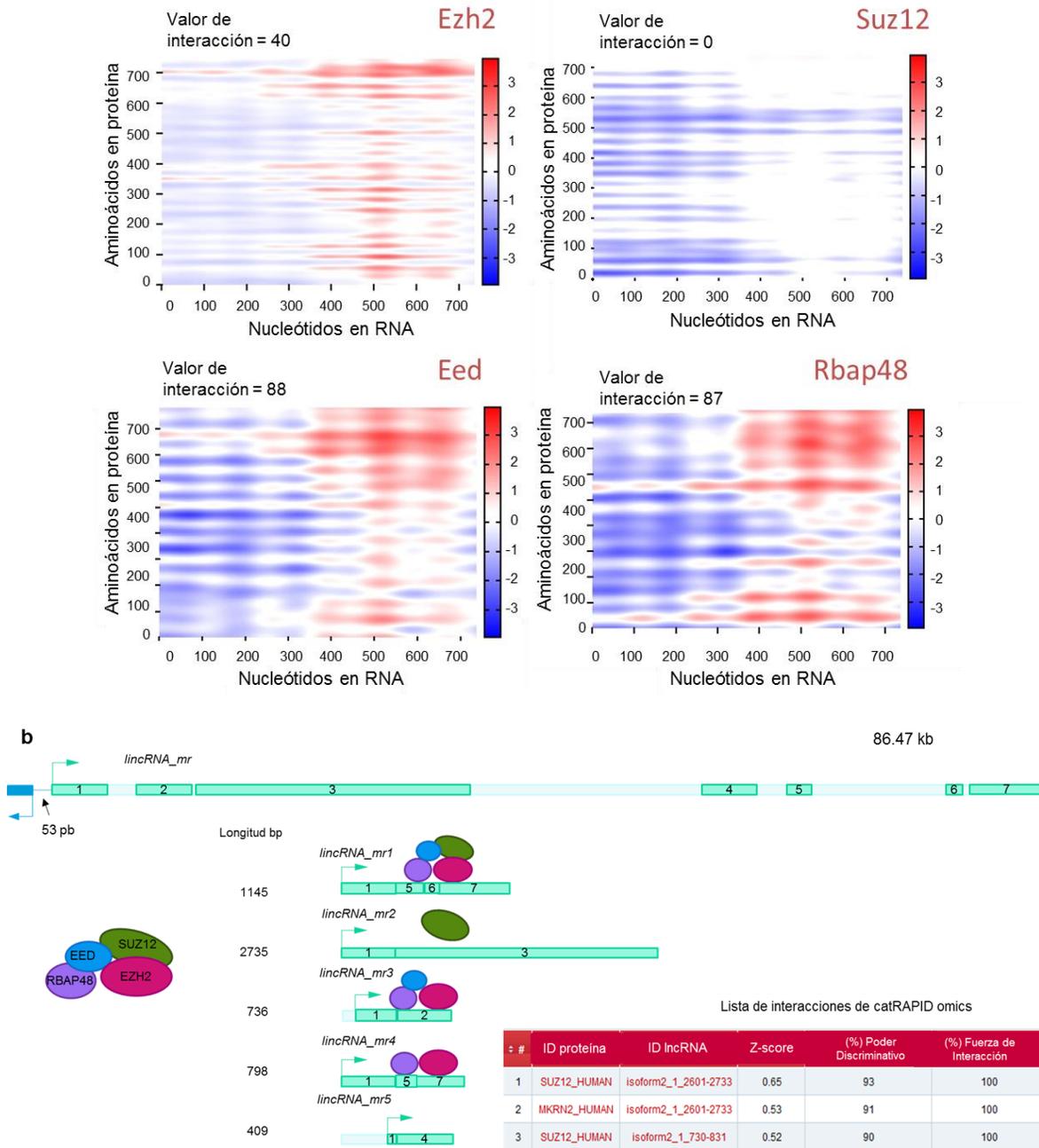


Figura 19: Predicción *in silico* de las interacciones entre las isoformas del *lincRNA_mr* con proteínas del complejo PRC2. a) Esquema de asociación entre proteína (eje y, en aminoácidos) y RNA (eje x, en nucleótidos para el *lincRNA_mr3*) que indica qué regiones en estas moléculas tienen una predicción más fuerte (en rojo) de interactuar. b) Resumen de las subunidades de PRC2 que se reportaron con asociación a las isoformas del *lincRNA_mr*. La zona donde se colocan las proteínas de PRC2 sobre los transcritos está en relación con la región donde se predijo la interacción más fuerte. En la caja de abajo, a la derecha, se muestran los resultados del algoritmo catRAPID omics para el *lincRNA_mr2*, pudiéndose apreciar a SUZ12 en los lugares 1 y 3 de la tabla.

La observación de distintos valores de interacción según la isoforma a un grupo de proteínas nos hizo especular que podría existir una función diferencial según el transcrito, similar a lo demostrado en genes codificantes con el procesamiento alternativo. Partiendo de que los lncRNAs suelen tener estructuras secundarias y terciarias que podrían formar módulos que le permitan una interacción variable con complejos de regulación epigenética, el procesamiento alternativo del *lincRNA_mr* podría conferirle diversas funciones a sus isoformas. Por esta razón, decidimos analizar la estructura secundaria de los transcritos producidos a partir del *lincRNA_mr* con el algoritmo mFold (Zuker, 2003) (Fig. 20).

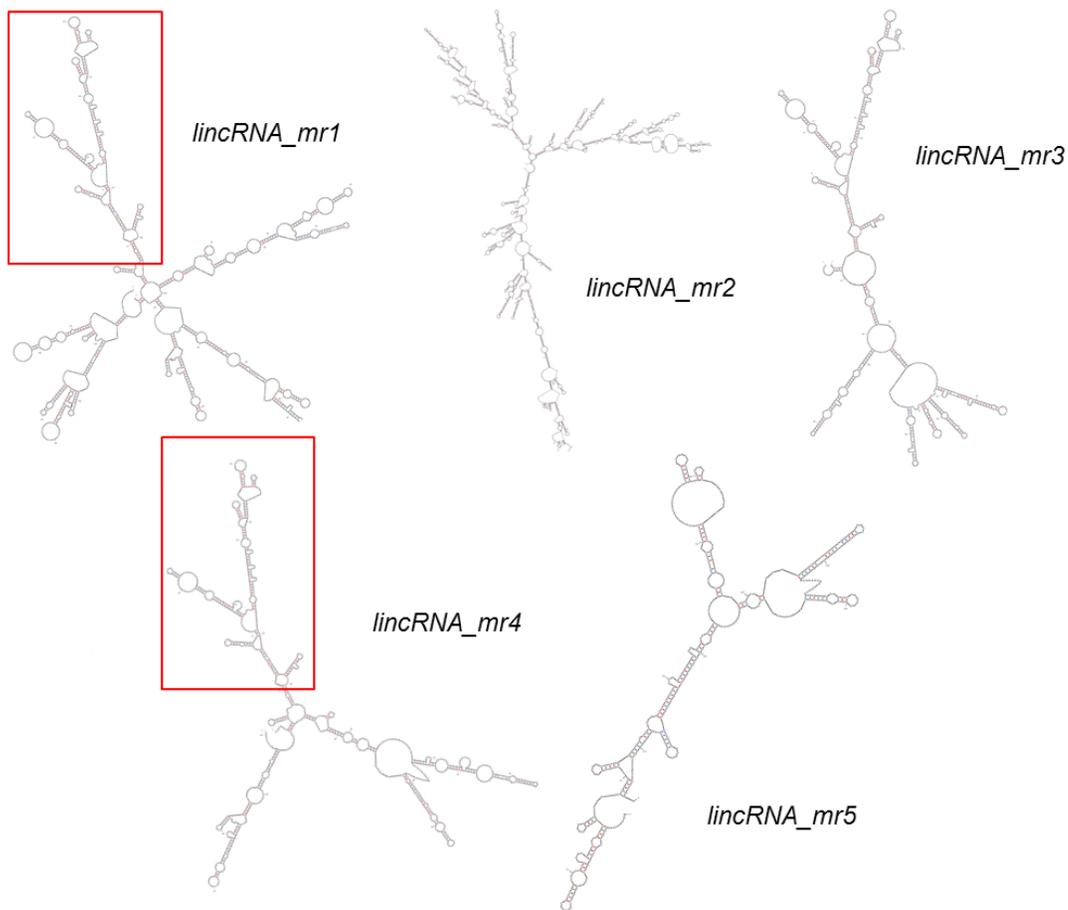


Figura 20: Estructuras secundarias predichas para las isoformas del *lincRNA_mr*. En la figura se muestra la predicción de cómo se pliega cada secuencia que surge de *lincRNA_mr* de acuerdo a características bioquímicas, según el programa mFold. En el recuadro rojo se puede ver cómo para dos isoformas distintas, que difieren en que el *lincRNA_mr1* contiene al exón 5, existen dominios idénticos y otros variables, abriendo la posibilidad de una interacción diferencial a complejos epigenéticos de acuerdo al transcrito expresado. Esto podría conferir una función diferencial.

Se observó que cada una de las isoformas tiene una conformación propia, pero algo que saltó a la vista fue que para el *lincRNA_mr1* y el *lincRNA_mr4*, las cuales sólo difieren en que la primera contiene al exón 5, conservan un dominio idéntico pero otras regiones variables. Si se toman en cuenta los resultados de catRAPID, se observa una interacción diferencial entre estas dos isoformas. Con estos resultados sugerimos que, al igual que lo observado en mRNAs, el procesamiento alternativo de lncRNAs podría generar transcritos con funciones específicas que podrían también expresarse de manera tejida o estadio específico, agregando un nivel adicional de complejidad a estos ncRNAs.

Sin embargo, todos los análisis realizados hasta este punto se trataban de aproximaciones computacionales, por lo que era necesario comenzar a realizar ensayos experimentales. Debido a que se observaba esta asociación diferencial, una pregunta crucial era entender cuáles eran los niveles de expresión del *lincRNA_mr* en nuestro modelo de estudio y qué isoformas son predominantes en los contextos celulares mamarios normal y cancerígeno.

2. Cuantificación de los niveles de expresión del *lincRNA_mr* en líneas celulares de tejido mamario

Con el fin de determinar la expresión del *lincRNA_mr* en nuestro modelo de estudio, se cuantificó experimentalmente a este transcrito por retro-transcripción seguida de reacción en cadena de la polimerasa (RT-PCR). Como un primer enfoque, se diseñaron y sintetizaron oligonucleótidos que permitieran amplificar a nuestro candidato en el exón 1, que es el común a la mayor parte de las isoformas. Una vez que se pudo recuperar al *lincRNA_mr*, se mandó secuenciar y se demostró su identidad.

Una vez que se comenzaron a realizar estos experimentos de RT-PCR en distintos contextos celulares, los primeros indicios sugerían que los niveles de expresión del *lincRNA_mr* eran variables dependiendo de la línea estudiada. El cambio más notorio se observó entre la línea celular MCF7 -que mostraba poca expresión- y MCF-10A, que tenía niveles más elevados. A manera de poder cuantificar la expresión de este transcrito en el modelo establecido, se realizaron experimentos de RT-PCR en tiempo real (qRT-PCR), normalizando los datos con el gen constitutivo *gapdh*. Se observó que la expresión de este transcrito es mayor en la línea celular normal MCF-10A que en MCF7 y MDA-MB-231, pero no en ZR-75-30 y T-47D (Fig. 21).

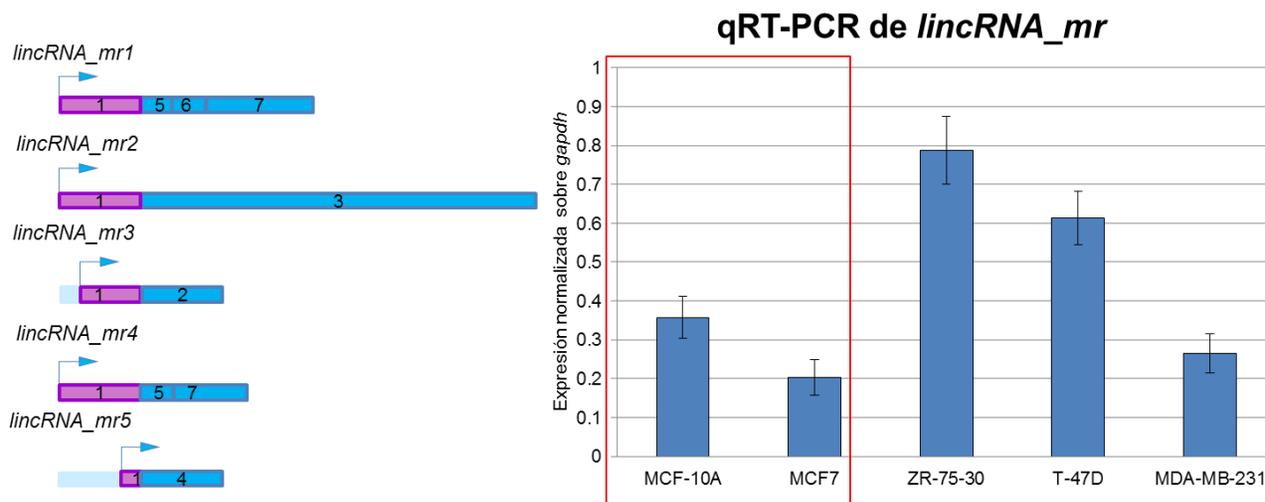


Figura 21: Niveles variables de expresión del *lincRNA_mr* en las líneas del modelo de estudio. En la figura se puede observar la cuantificación por método de Ct comparativo contra *gapdh* para el exón común a todas las isoformas de lincRNA (primer exón, en morado), medida por experimentos de qRT-PCR. Partiendo de estos resultados, se decidió establecer un sub-modelo con MCF-10A y MCF7, que posteriormente se retomará para una serie de ensayos.

Debido a que la disminución en los niveles de expresión del *lincRNA_mr* en MCF7 fue observada de manera reproducible con respecto a MCF-10A (en alrededor de 10 experimentos independientes), se decidió establecer un sub-modelo de estudio con estas dos líneas celulares. Este sub-modelo se podría considerar como uno tipo binario, donde MCF-10A representa la línea donde existe expresión del *lincRNA_mr* o “1”, mientras que en MCF7 la presencia de poco transcrito se representaría con un “0”.

Es importante destacar que antes de este análisis no se había observado la expresión del *lincRNA_mr* experimentalmente. Sin embargo, aún quedaba la incógnita de cuáles isoformas se expresan en el contexto celular normal MCF-10A. Este dato resulta de suma importancia cuando se desean abordar experimentos de ganancia y pérdida de función, así como si se desea comparar la abundancia de estos transcritos en muestras de individuo sano, es decir, en un contexto fisiológico normal.

2.1 Caracterización de la expresión de las 5 isoformas del *lincRNA_mr*

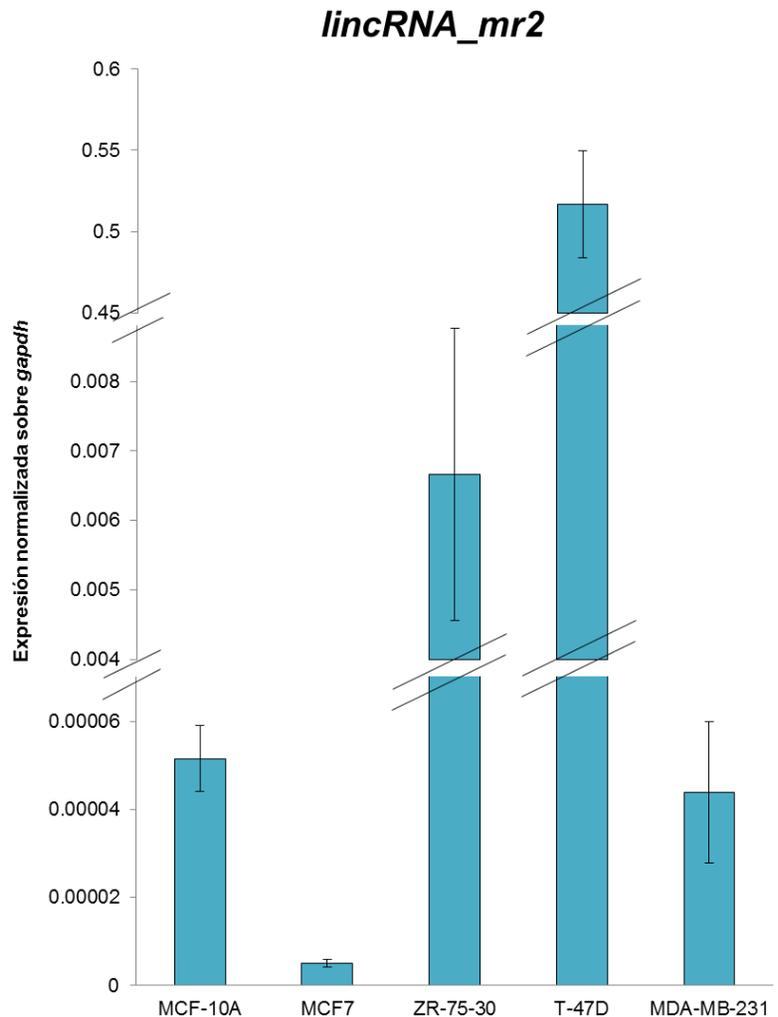
El siguiente objetivo consistió en identificar si existe una expresión diferencial de las isoformas del *lincRNA_mr* en la línea celular MCF-10A. Una vez que se demostró que se podía recuperar al *lincRNA_mr* con oligonucleótidos comunes a partir de cDNA con oligo d(T), se prosiguió a

hacer RT-PCRs, seguidas de ligación en pGEM®-T Easy, clonación y secuenciación, para averiguar qué tipo de transcritos se expresan.

De esta forma, se analizó la expresión de cada una de las isoformas del *lincRNA_mr* en MCF-10A. El resultado fue que se pudo amplificar y validar por secuencia tan solo al *lincRNA_mr2*, indicando que esta isoforma es la única expresada, o la más abundante, en MCF-10A. Datos de RNA-seq en UCSC genome-browser apoyan un enriquecimiento de señal en los exones correspondientes a esta isoforma, así como un estado de transcripción elevado que se observa en los exones del *lincRNA_mr2*, pero no así para las otras isoformas. Interesantemente, los niveles de expresión observados para la isoforma 2 del *lincRNA_mr* por qRT-PCR se asemejan a los valores obtenidos al cuantificar al *lincRNA_mr* utilizando oligonucleótidos que amplifican un exón común a todas las isoformas de este gen (Fig. 22).

Figura 22: Los niveles de expresión del *lincRNA_mr2* varían en las líneas del modelo de estudio. Niveles de expresión de la isoforma 2 del *lincRNA_mr* por método de Ct comparativo contra *gapdh*. Se puede observar que la tendencia en cada línea celular es muy similar a cuando se cuantifica globalmente al *lincRNA_mr* (Fig. 21).

Una vez que se determinaron los patrones de expresión del *lincRNA_mr* en nuestro modelo, resultó importante comenzar a cuantificar a los transcritos aledaños codificantes, para averiguar si existe alguna correlación entre los niveles de expresión de este lincRNA y algún mRNA, pudiendo indicar que se tratan de blancos génicos en *cis*.



3. Cuantificación de los genes codificantes aledaños al *lincRNA_mr*

Una de las preguntas más relevantes al investigar un lincRNA es conocer cuáles son sus blancos génicos y si es capaz de tener una acción en *cis* y *trans*. En este estudio, decidimos enfocar nuestros esfuerzos en analizar el mecanismo en *cis* mediante la cuantificación de los niveles de expresión de los genes codificantes aledaños *cbl-b* y *ccdc54* (Fig. 23). El mecanismo en *trans* quedará como una de las perspectivas de este trabajo de investigación. La estrategia abordada siguió el diseño y la síntesis de oligonucleótidos que estuvieran cercanos al extremo 3' de estos genes, para que pudieran ser retro-transcritos con oligo-d(T), que puede amplificar de manera selectiva a los transcritos poliadenilados.

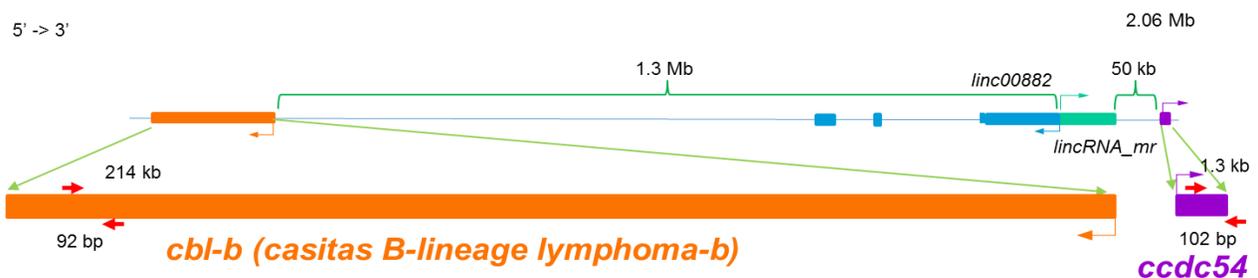
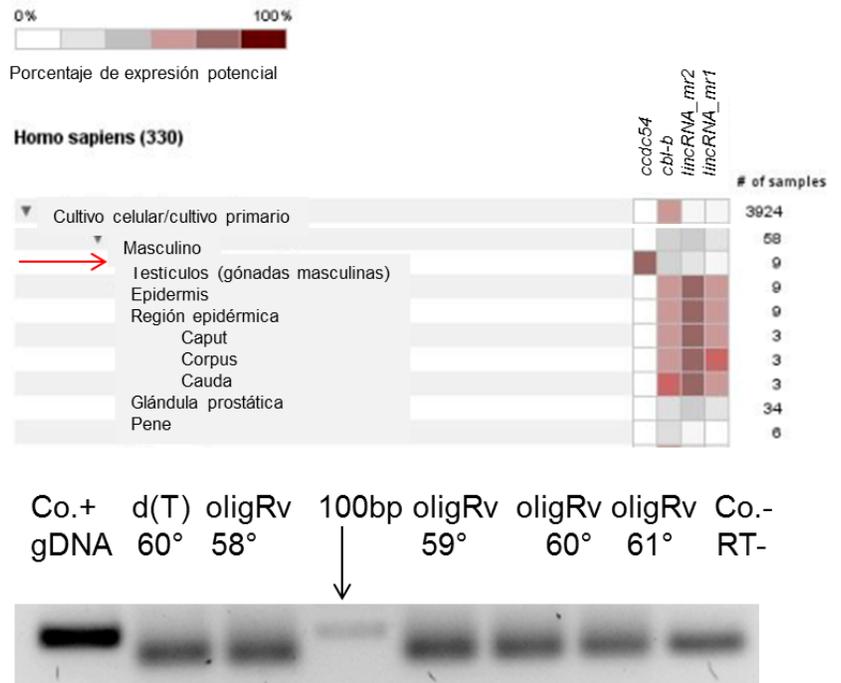


Figura 23: Mapa de los genes codificantes *cbl-b* y *ccdc54* en el locus q13.12. Al igual que en la figura 16, arriba se muestra un esquema correspondiente a esta zona del cromosoma 3 humano. A manera de acercamiento, se puede ver en naranja al gen *cbl-b*, localizado a 1.3 Mb arriba del sitio de inicio de la transcripción del *lincRNA_mr* y en morado al gen *ccdc54*, a 50 kb río abajo. Las flechas rojas representan la región que hibrida con los oligonucleótidos; también se muestra el tamaño del amplicón esperado en las PCRs.

En el caso de *ccdc54*, un gen que codifica para una proteína de desarrollo testicular, se hicieron una serie de RT-PCRs punto final, tanto a partir de cDNA convertido con oligo-d(T), como con oligonucleótidos específicos para este gen, sin tener algún éxito. Los experimentos se realizaron en múltiples contextos celulares que van desde las líneas del modelo de estudio hasta de cerebro y de próstata (Fig. 24). Estos resultados son consistentes con la literatura, en la cual se sugiere que este gen es tejido específico para testículo.

Figura 24: Análisis de expresión del gen *ccdc54*. En la parte superior de la figura se puede ver que, según la base de datos Genevestigator (ETH-Zurich), el único tejido que presenta expresión de *ccdc54* es el testicular. En la parte de abajo se puede observar el resultado de una RT-PCR a 40 ciclos para este gen usando cDNA de la línea celular de próstata LNCaP. El único carril que muestra un producto esperado es el del control positivo, con gDNA de MCF-10A.



Por su parte, el proto-oncogen “Casitas B-lineage lymphoma” (*cbl-b*) codifica para una proteína ligasa E3 de ubiquitina y su proteína también puede utilizarse como andamio en algunas cascadas de señalización celular, al contener un dominio SH3. Para este gen, se logró amplificar un producto del tamaño deseado, validar por secuencia y estandarizar en qRT-PCR. Los niveles de expresión del gen *cbl-b* se cuantificaron en el modelo de estudio y se observó que existe una expresión muy similar entre el gen *cbl-b* y el *lincRNA_mr2* (Fig. 25).

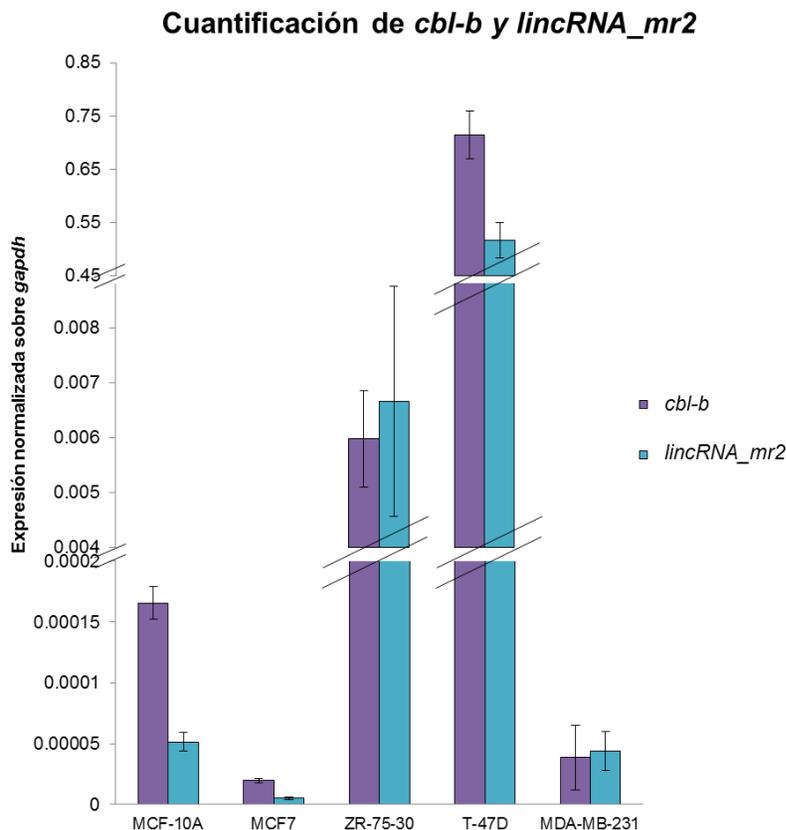


Figura 25: Los niveles de expresión del gen *cbl-b* en las líneas del modelo de estudio se asemejan a los del *lincRNA_mr2*. En esta figura se presentan los niveles de expresión del gen aledaño *cbl-b*, junto con los observados para la isoforma 2 de *lincRNA_mr*. Se utilizó el mismo cDNA de templado para la amplificación de ambos genes y su normalizador. Se puede observar que estos genes poseen niveles de expresión similares entre sí.

Para darle un mayor soporte a la relación que existe entre ambos genes, se realizó un análisis estadístico que demostró que existe una correlación positiva entre los niveles de expresión del *lincRNA_mr2* y el gen *cbl-b*, con un coeficiente de correlación de $\rho = 0.9999899$ (Pearson), $\tau = 1$ (Kendall) y $r_s = 1$ (Spearman) (Fig. 26). Además, al realizar la regresión lineal de los puntos de la gráfica que toma los valores de los niveles de expresión del gen *cbl-b* en el eje X y del *lincRNA_mr2* en el eje Y, se puede graficar una línea que refuerza esta correlación.

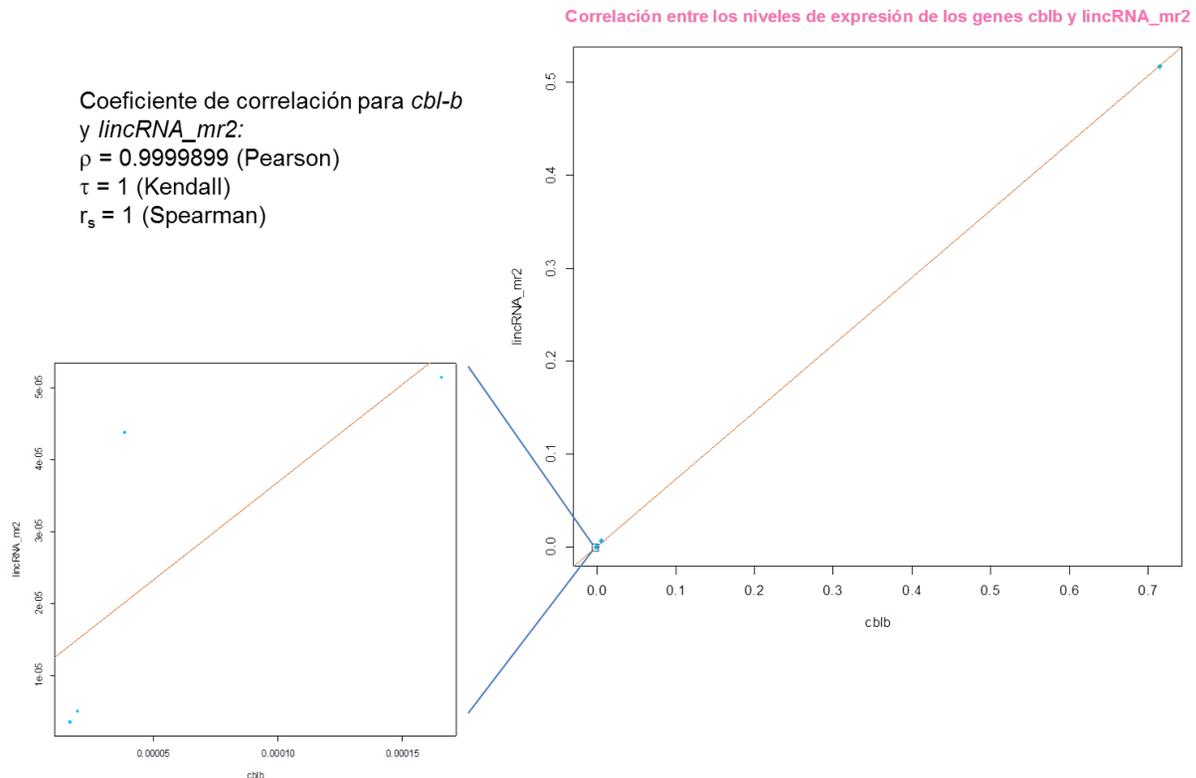


Figura 26: Existe una correlación positiva entre la expresión de los genes *cbl-b* y el *lincRNA_mr2*. En esta figura se puede observar una gráfica que toma los valores de los niveles de expresión del gen *cbl-b* en el eje X y del *lincRNA_mr2* en el eje Y. Al realizar una regresión lineal con estos puntos, se traza una línea recta que indica una correlación positiva. Para poder apreciar mejor a los puntos, se hace un zoom, indicado por las líneas azules. Además, arriba a la izquierda se pueden observar los valores de diversos estadísticos, que soportan este tipo de relación entre ambos genes.

Aun cuando estos resultados sugieren fuertemente que existe un tipo de relación entre el *lincRNA_mr2* y el gen *cbl-b*, es necesario realizar ensayos adicionales que puedan elucidar si existe una regulación tipo *cis*. Por lo tanto, una manera de poder descifrar si el *lincRNA_mr2* está actuando sobre el gen *cbl-b* es mediante la alteración de los niveles de expresión del lincRNA, seguido de la cuantificación de su posible gen blanco, para entender si la desregulación del primero afecta la expresión del segundo gen.

4. Evaluación del gen blanco por medio de ensayos de ganancia y pérdida de función del *lincRNA_mr*

4.1 Ensayos de pérdida de función para abatir al *lincRNA_mr*

Para abordar los análisis de pérdida y ganancia de función, se decidió utilizar al sub-modelo con las líneas MCF-10A y MCF7, respectivamente. Con el fin de disminuir la expresión del *lincRNA_mr* en un contexto normal, simulando lo que se observa en algunos contextos tumorales, se diseñó un shRNA que reconoce a 4 de las 5 isoformas del *lincRNA_mr*. Este shRNA (desde ahora shLincMR) se seleccionó con ayuda de algoritmos bioinformáticos, se sintetizó en forma de sesentámeros para su posterior apareamiento y se clonó en el vector pLentiLox3.7-Puro (Fig. 27). Después de verificar que nuestro shRNA se había clonado adecuadamente tanto por patrones de restricción como por secuencia, se generaron partículas virales que contuvieran a este vector para que así pudiera infectar a la línea MCF-10A. Para seleccionar a las células que integraron al vector shLincMR, se utilizó puromicina.

Una vez que se recuperaron células después de la infección, se extrajo RNA y se demostró por qRT-PCR que shLincMR abate la expresión del *lincRNA_mr* en aproximadamente un 50% con respecto al vector vacío (promedio de 4 experimentos independientes) (Fig. 28).

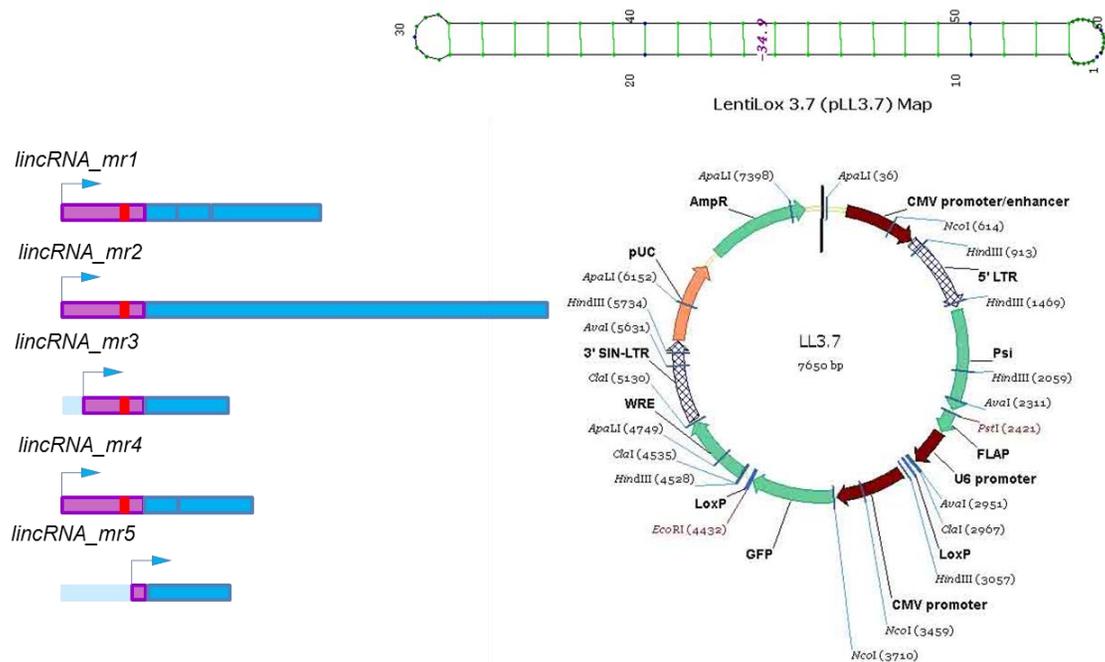


Figura 27: Diseño de shLincMR para abatir a *lincRNA_mr*. A la izquierda de esta figura se puede observar en rojo la región a la cual se une shLincMR en 4 de las 5 isoformas. Arriba a la derecha se muestra la estructura secundaria del shRNA, la cual es óptima para que ejerza su función. También se pueden observar el vector donde se clonó el sesentámero.

Una manera de corroborar que este abatimiento se debe al shRNA fue mediante la infección de otra línea celular. Se eligió utilizar a ZR-75-30, partiendo de que este es el contexto donde existe mayor expresión del *lincRNA_mr* (Figs. 21 y 22). Consistente con los resultados observados en MCF-10A, se pudo observar una disminución de este transcrito en la segunda línea celular, demostrando que shLincMR funciona adecuadamente en distintos contextos celulares. Como ya se había demostrado la reducción del *lincRNA_mr*, nuestra siguiente pregunta fue si existía algún cambio en el nivel de expresión del gen *cbl-b* que pudiera confirmar que este lincRNA actúa en *cis*.

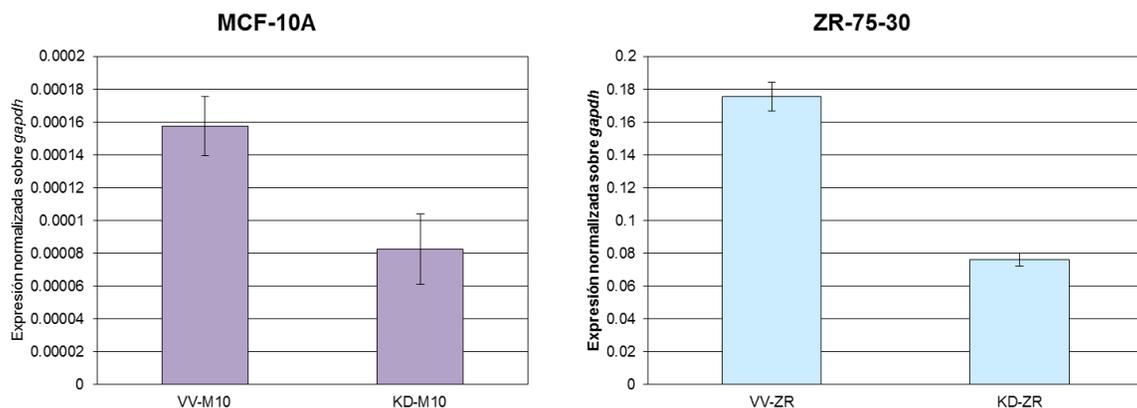


Figura 28: La expresión de *lincRNA_mr2* disminuye en las líneas celulares infectadas con el shLincMR. En ambas gráficas se puede observar una disminución en la línea celular denominada KD (knockdown) con respecto al control (vector sin shRNA, denominado VV). En el caso de MCF-10A se puede observar una reducción del 48%, mientras que en ZR-75-30 existe una disminución en la expresión del 57%.

4.2 Cuantificación de la expresión del gen *cbl-b* para determinar mecanismo en *cis*

Para determinar los niveles de expresión del gen aledaño al *lincRNA_mr2*, se realizó una qRT-PCR en la línea celular donde que presentaba una disminución del transcrito largo no-codificante (KD-M10). Se observó que la expresión del gen *cbl-b* también disminuye en esta línea celular. Este resultado mimetiza lo que se observa en MCF7, donde los niveles de expresión de ambos genes son menores que en MCF-10A. Además, este resultado sugiere que la expresión del *lincRNA_mr* es necesaria para mantener la expresión del gen *cbl-b*, favoreciendo una regulación tipo *cis* entre el *lincRNA_mr* y el gen *cbl-b* (Fig. 29).

El haber realizado un ensayo de pérdida de función nos ayudó a elucidar una serie de posibles consecuencias que surgen del abatimiento del *lincRNA_mr*. Una manera de reforzar estos resultados es mediante la ganancia de función en un contexto que muestre baja expresión

del transcrito, como lo es MCF7. Es así como se esperaría que este experimento funcionara como un rescate de las funciones de este lincRNA.

cbl-b* disminuye en consecuencia a la pérdida de *lincRNA_mr2

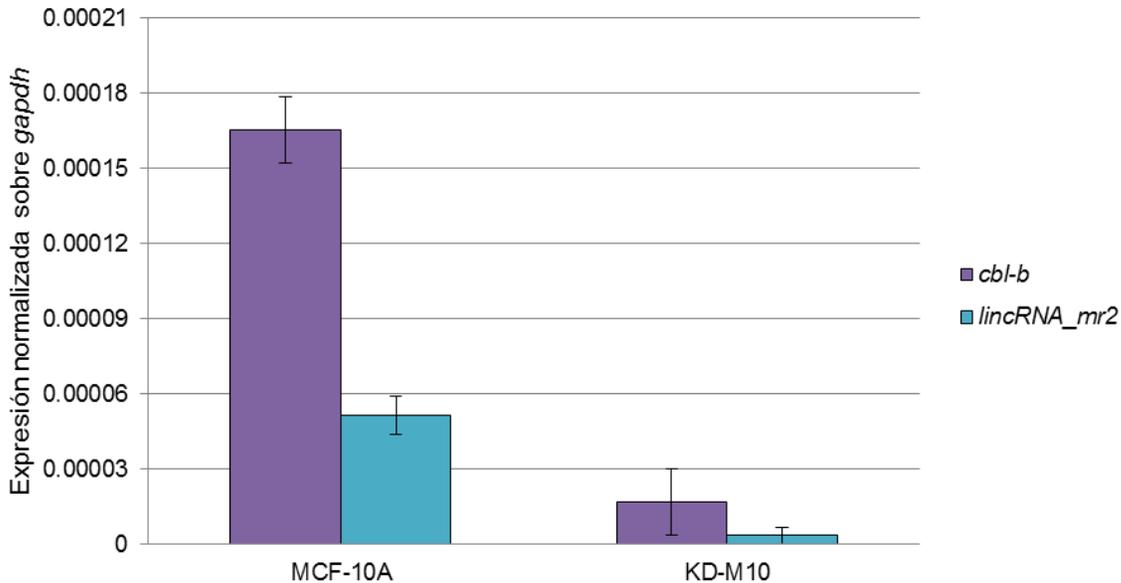


Figura 29: La expresión del gen *cbl-b* disminuye cuando se abate el *lincRNA_mr2*. En esta figura se puede observar que existe una disminución de casi un 90% en el gen *cbl-b* cuando se hace reduce la expresión de este lincRNA.

4.3 Ensayo de ganancia de función al sobre-expresar al *lincRNA_mr2* en MCF7

Con la finalidad de estudiar si el aumento en la expresión de nuestro transcrito de interés en las células MCF7 altera en algún modo el fenotipo tumoral, se procedió a diseñar un experimento de sobre-expresión. Debido a que el *lincRNA_mr2* es la isoforma que se pudo amplificar dentro de las 5 posibles en células MCF-10A, se clonaron ambos exones usando gDNA como templado, para capturar al transcrito en su totalidad (2,737 bp). Se planeó utilizar al vector pcDNA3.1 hygro (+), que posee un promotor fuerte, para luego realizar una transfección estable.

Es importante mencionar que para poder unir a ambos exones se introdujo un sitio XhoI. Aunque se comprobó que no se alteraba drásticamente la secuencia por mFold, no tenemos la certeza de que el tener seis bases extra entre los exones no vaya a afectar la funcionalidad o plegamiento del *lincRNA_mr2*. También decidimos utilizar al vector que sólo contiene al exón 1 a manera de control, ya que podría indicarnos la necesidad de la isoforma completa o si este fragmento basta para ejercer alguna función.

Posterior a la selección con higromicina, se extrajo RNA a las células para validar un posible incremento en la expresión del *lincRNA_mr2* (Fig. 30). Los resultados nos indican que existe un incremento en los niveles de expresión del *lincRNA_mr* en distintos niveles para el plásmido que contiene al exón 1 y la isoforma completa. En ambos casos, el incremento es sustancial, existiendo mayor cantidad de este lincRNA que del gen constitutivo *gapdh*.

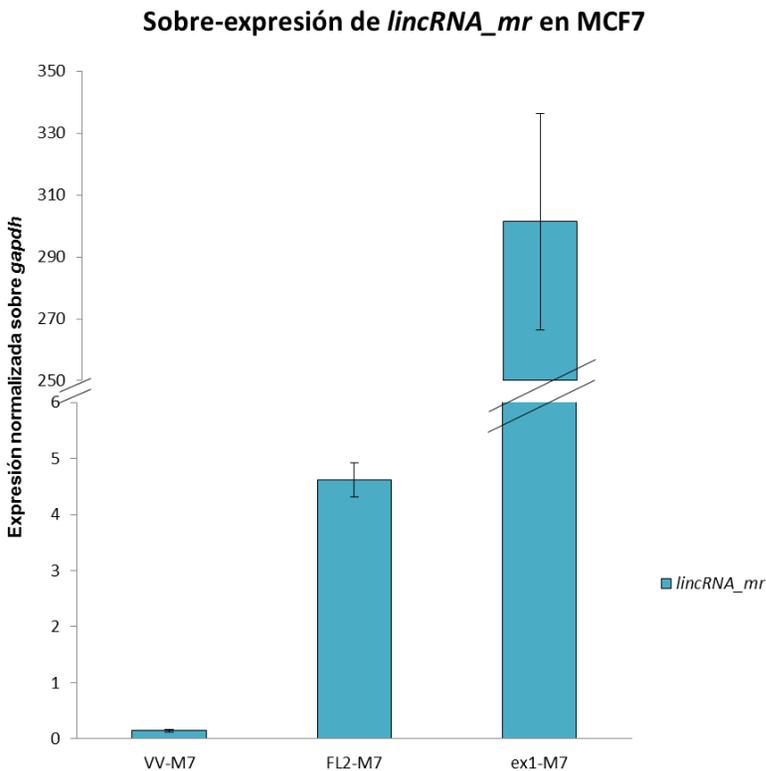


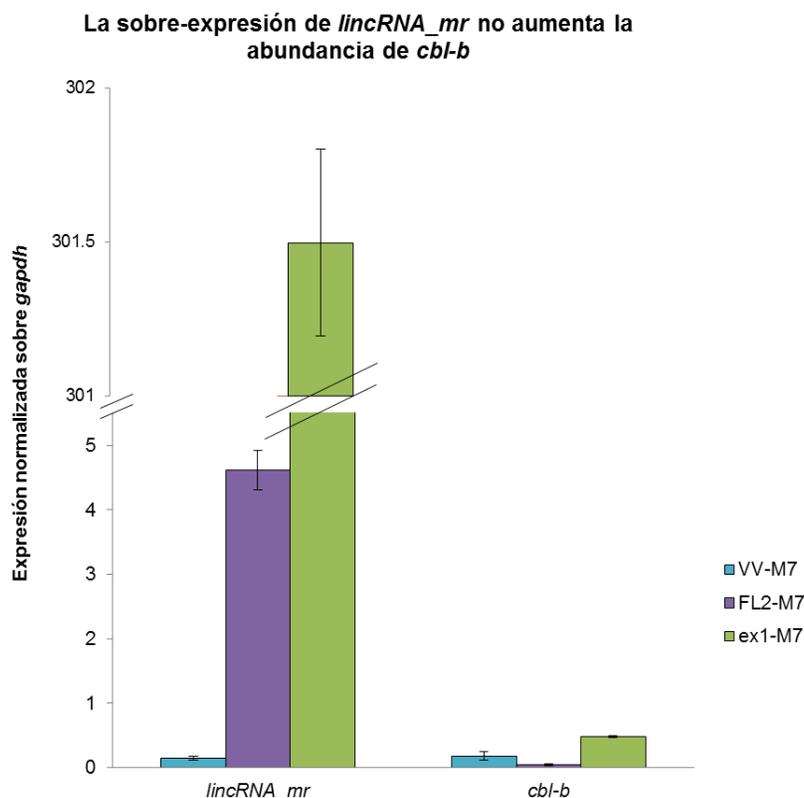
Figura 30: Validación de la sobre-expresión del *lincRNA_mr2* en la línea celular MCF7. En esta gráfica se puede observar un aumento en la expresión de nuestro transcrito candidato en las líneas celulares denominadas FL2-M7 y ex1-M7 con respecto al control (vector vacío).

4.4 Cuantificación de la expresión del gen *cbl-b* en un contexto de sobre-expresión

La siguiente pregunta consistió en entender si existe un cambio en los niveles de transcripción del gen *cbl-b* ante el incremento de la expresión del *lincRNA_mr2*. Interesantemente, en ninguno de los casos se observa una ganancia en la expresión del gen codificante (Fig. 31).

Estos resultados siguen la premisa de que un lincRNA que actúe en *cis* debe localizarse en la vecindad de su blanco, no pudiendo recuperar la función con el producto de su transcrito, quizá quedándose unido a su sitio de transcripción y ejerciendo su función desde este lugar (Guttman & Rinn, 2012; Lee, 2012). Es así como nuestros resultados indican que el *lincRNA_mr2* es un transcrito que regula en *cis* al gen que codifica para la ligasa de ubiquitina E3 *cbl-b*.

Figura 31: Los niveles de expresión del *lincRNA_mr* y el gen *cbl-b* en el ensayo de sobre-expresión. En esta figura se puede observar la cuantificación de ambos transcritos por qRT-PCR. Es evidente que no existe un incremento sustancial en el gen blanco ante la expresión exógena del lincRNA, lo cual podría deberse a que no es el producto del transcrito, sino el transcrito anclado a su sitio de transcripción, lo que resulta necesario para ejercer su función sobre el gen *cbl-b*.



5. Evaluación de la fisiología celular por ensayos de pérdida de función

5.1 Análisis de división celular y fenotipos celulares

Cada vez que se realizaba la infección lentiviral, notamos cambios fenotípicos y alteraciones en el tiempo de división celular en las líneas celulares. Por esta razón decidimos cuantificar la tasa de división en las líneas del sub-modelo, VV-M10 y KD-M10 como un primer ensayo con la idea de definir si la cantidad de este transcrito resulta en el cambio de algún proceso celular.

Con el fin de estudiar si existen cambios en algunos procesos celulares consecuentes al abatimiento de *lincRNA_mr*, se analizó la tasa de división celular para las líneas del sub-modelo de estudio de tejido mamario (Doubling Time, DT). Este análisis resulta en una gráfica donde se observa el número de células existentes con el paso del tiempo, además de un valor de DT que indica la tasa de división celular en horas (Fig. 32).

Los resultados muestran que existe una disminución en la tasa de DT para la línea celular KD-M10, además de una disminución drástica en los números de esta línea celular desde el primer día. Esto no se observa para las células VV-M10 ni en MCF-10A, donde el número de células a lo largo del tiempo permanece muy similar en ambos casos. Dado que los niveles de expresión del *lincRNA_mr* no correlacionan con la tasa de división celular cuando se analizan todas las líneas del modelo de estudio, la división celular no es un proceso que se afecte al cambiar la expresión del *lincRNA_mr*.

Ensayo de división celular

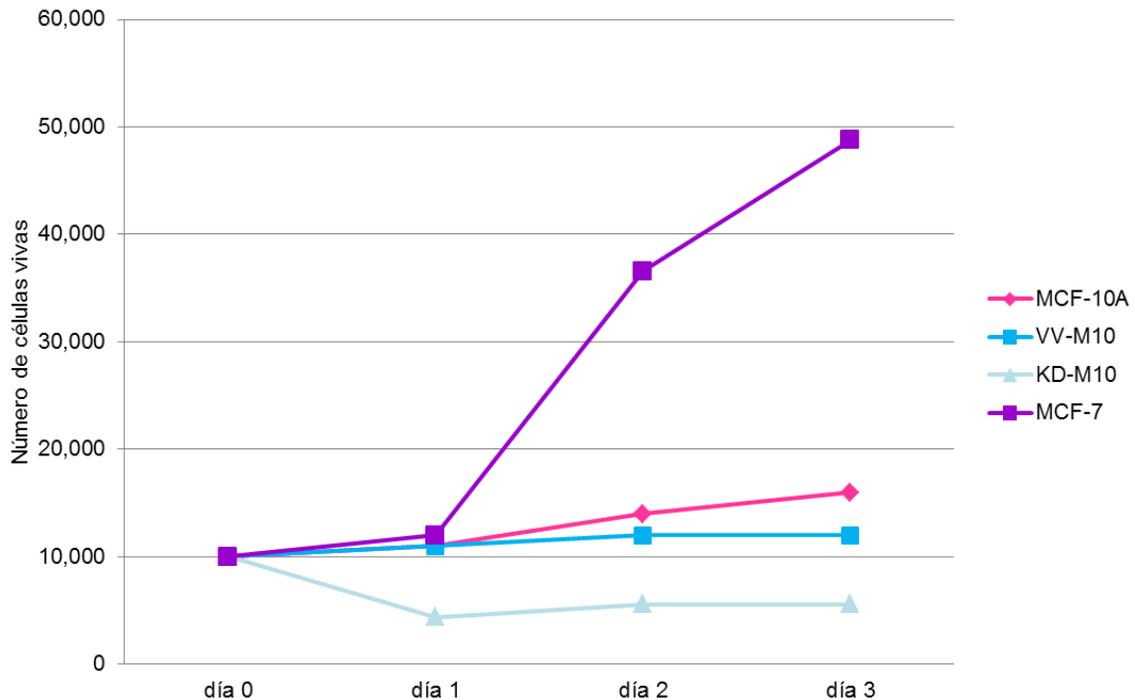


Figura 32: Resultados del ensayo de tiempo de división celular en las líneas MCF-10A, MCF7, VV-M10 y KD-M10. En este diagrama se puede observar el número de células presentes a diferentes días para cada una de las cuatro líneas celulares. Nótese que existe una reducción en el número de células en KD-M10 desde el 1er día, pero no así en VV-M10.

Como un segundo ensayo, se decidió analizar el fenotipo celular de las líneas celulares VV-M10 y KD-M10 por microscopía de interferencia diferencial (Nomarski), para ver si existen cambios macroscópicos. Al tener problemas para alcanzar un número estadístico significativo con VV-M10 y KD-M10, se decidió estandarizar la prueba con VV-ZR y KD-ZR, las líneas celulares que se generaron a partir de la infección de shLincMR en ZR-75-30 y las cuáles proliferaban rápidamente. En breve, se crecen las células sobre un cubre-objetos que se agrega al fondo de una caja con 6 pozos. El día del ensayo se retira el medio de cultivo, se coloca glicerol en el porta-objetos y se coloca cubre-objetos. Se toman fotos de 6-10 campos de cada cubre-objetos y se analizan con diversos pluggins del programa ImageJ.

Desde el comienzo del análisis de VV-ZR y KD-ZR, se observaron cuatro fenotipos en los cuales se podía clasificar a las células: redondas, fusionadas, alargadas y con filopodia (Fig. 33). Después de contar cuántas células de cada tipo existían para ambas líneas con el Cell Counter de ImageJ, se realizó un análisis estadístico de medidas de tendencia central con R 3.0.2

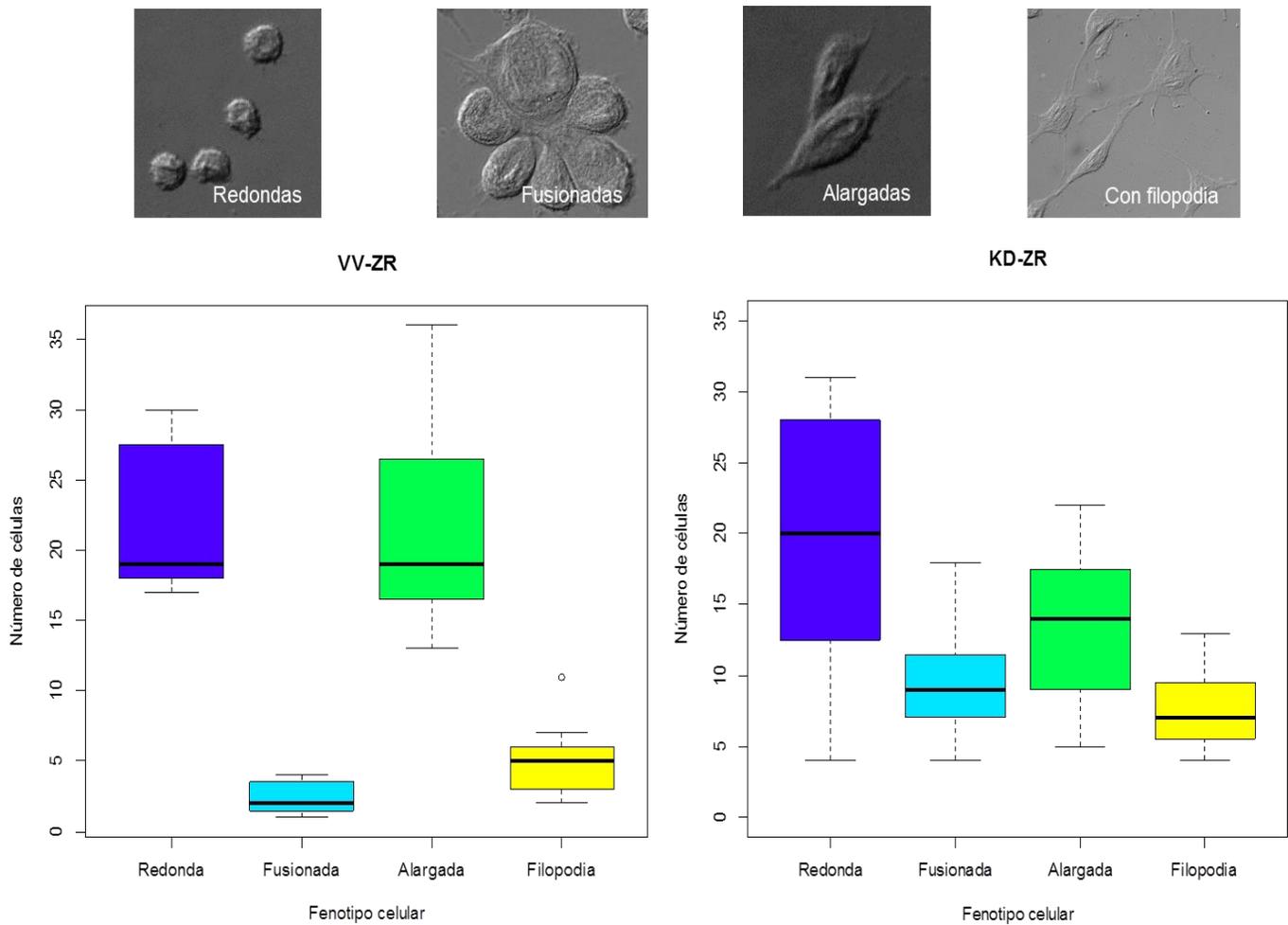


Figura 33: Análisis de microscopía para estudiar cambios en el fenotipo de VV-ZR y KD-ZR. En la parte superior, se pueden apreciar los cuatro fenotipos más comunes que se podrían encontrar en estas líneas celulares. Después de contar ~400 células por línea, se clasificaron en estos 4 tipos con el plugin de ImageJ (Cell Counter) y se hizo un análisis estadístico con R 3.0.2. Se puede ver que existe un cambio en la distribución de los fenotipos de VV-ZR a KD-ZR.

Los resultados muestran que el número de células fusionadas (2.4 en VV-ZR, 9.5 en KD-ZR) y con filopodia (5.1 en VV-ZR, 7.7 en KD-ZR) incrementaba en KD-ZR cuando se compara contra VV-ZR. Interesantemente, el gen *cbl-b* se ha implicado en la redistribución de moléculas señalizadoras en las balsas lipídicas (Xu, *et al.*, 2012) y se ha reportado que su baja expresión puede afectar la red de microtúbulos (Purev, *et al.*, 2009). Esto podría comenzar a sugerir el por qué se observan dichos fenotipos cuando se disminuye la expresión del gen *cbl-b*.

5.2 Análisis de TUNEL en un contexto celular con *lincRNA_mr* abatido

Existen reportes que implican al gen *cbl-b* como una proteína de andamiaje en la ruta de apoptosis. Estos estudios indican que cuando se abate el gen *cbl-b*, existe un incremento en la cantidad de receptores de muerte DR4 y DR5, que están vinculados a la vía extrínseca de la apoptosis inducida por TRAIL (un ligando relacionado con TNF) (Yan, *et al.*, 2012). Dado que se observó una baja en el DT realizado para la línea KD-M10 con respecto a VV-M10 y se ha observado que el silenciamiento del gen *cbl-b* promueve la formación de un complejo señalizador inductor de muerte (DISC) en las balsas lipídicas (Xu, *et al.*, 2012), nos hemos interesado en estudiar si existe un aumento en la muerte celular cuando se abate el *lincRNA_mr* por medio de un ensayo de TUNEL (Fig. 34).

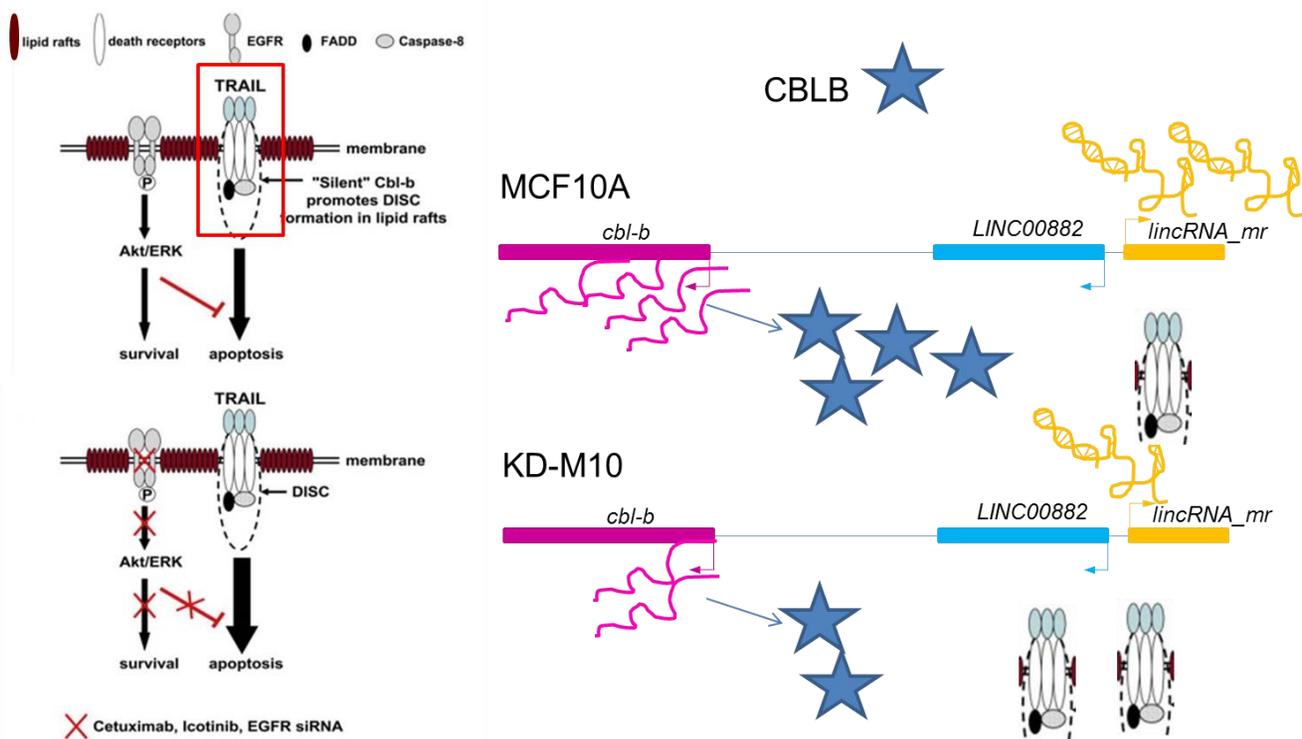


Figura 34: Modelo del gen *cbl-b* en el proceso de apoptosis. A la izquierda se muestra una imagen tomada de Xu *et al.* 2012, donde se observa que el abatimiento del gen *cbl-b* promueve la asociación del complejo DISC en las balsas lipídicas y conlleva al aumento de apoptosis por vía extrínseca. Se podría pensar que la baja expresión del *lincRNA_mr* en la línea KD-M10 presenta el efecto de aumento de DISC que se validó en el trabajo de la izquierda.

Para este ensayo, se realizaron infecciones lentivirales en cajas de 6 pozos que contenían desde un principio cubre-objetos en el fondo. Una vez que se seleccionaron las células con puomicina, se utilizó el kit de detección *in situ* de muerte celular por fluoresceína

(Roche) para detectar si existe un aumento en la apoptosis cuando se abaten los niveles del *lincRNA_mr*. Interesantemente, se observó que existe un incremento significativo en el número de células positivas a TUNEL asociado al abatimiento del *lincRNA_mr* (Fig. 35).

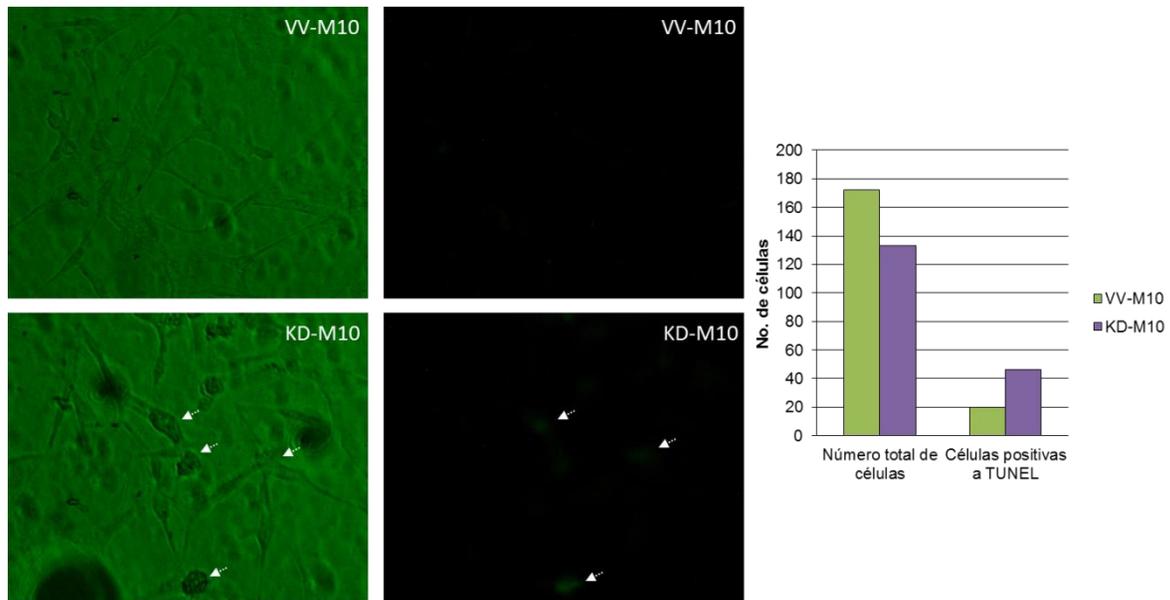


Figura 35: Ensayo de TUNEL muestra un aumento en los núcleos apoptóticos en KD-M10. En esta figura se muestran las imágenes en campo claro (izquierda) y con un filtro para irradiar a la fluoresceína (derecha) en un contexto control (VV-M10) y uno con *lincRNA_mr* abatido (KD-M10). Nótese que existen cuerpos apoptóticos fluorescentes en KD-M10, indicados por flechas blancas. A la derecha, se presenta la cuantificación del número total de células y células positivas a TUNEL para ambas líneas celulares, siendo mayor para la línea KD-M10.

Estos resultados se encuentran en proceso de validación y podrían reforzarse con ensayos que involucren la detección de caspasas específicas para la ruta de apoptosis extrínseca y que confirmen que la desregulación de *lincRNA_mr* se asocia a dicho proceso celular.

DISCUSIÓN

LincRNA_{mr}: sus cambios de expresión en cáncer, su procesamiento alternativo y su posible promotor

Existe una amplia evidencia en la literatura de lncRNAs cuya desregulación en sus niveles de expresión se asocia significativamente a tejidos cancerosos (Mitra, *et al.*, 2012). Esta correlación ha motivado a diversos grupos a re-analizar datos públicos para confirmar los patrones de expresión de los lncRNAs en muestras tumorales, lo que ha permitido la construcción de perfiles de expresión que contienen a miles de estos transcritos en diversos tejidos (Du, *et al.*, 2013; Gibb, *et al.*, 2011). Es ahora que se cuenta con dichos recursos que se pueden abordar preguntas partiendo de una búsqueda inicial en estas bases de datos.

En este trabajo, optamos por elegir a un lncRNA candidato que cumpliera con ciertos criterios de selección, para seguir respondiendo a la pregunta: ¿cómo afecta la desregulación de los lincRNAs que se ha relacionado con complejos de regulación epigenética en un tejido canceroso? Se decidió escoger un transcrito desregulado en cáncer mamario debido a que nuestro laboratorio participa en un consorcio que aborda preguntas en este tipo de cáncer. También porque ésta es la neoplasia maligna más común en la mujer a nivel mundial, existiendo en México una tendencia de mortalidad que se ha incrementado considerablemente en las últimas tres décadas (García Solís, 2006).

Cuando se identificó un transcrito con una alta desregulación en tejido canceroso, al cual denominamos *lincRNA_{mr}* (*linc00883*), se formularon una serie de preguntas acerca de su regulación, su mecanismo de acción y sus genes blanco. Interesantemente, *lincRNA_{mr}* también se reportó en un estudio reciente que se enfocó en identificar lncRNAs clínicamente relevantes para el cáncer humano. En este trabajo, se demostró que la desregulación de *lincRNA_{mr}* se asocia a tejidos ováricos tumorales con subtipo proliferativo, además de tejidos de pulmón y próstata. También se menciona que dicho lincRNA muestra alteraciones en su número de copias somáticas en algunos cánceres, aunque cabe destacar que este estudio no abordó al cáncer mamario (Du, *et al.*, 2013).

Después de una serie de análisis *in silico* que pretendían caracterizar a *lincRNA_{mr}*, un resultado que llamó nuestra atención es que parece existir una función diferencial para las isoformas producidas por un procesamiento alternativo de este transcrito. Esta observación abre una gama de preguntas acerca de su posible estructuración en dominios con asociación preferencial a diversos complejos, haciendo interesante analizar más acerca de su estructura secundaria y terciaria. Cabe mencionar que la posibilidad de que un lncRNA posea isoformas con expresión diferencial ya se ha reportado, para el lncRNA *MEG3* (Zhang, *et al.*, 2010). *MEG3*

es un gen imprintado que expresa un lncRNA de 1.7 kb y que se puede procesar en 12 diferentes isoformas. Estas isoformas contienen a los exones 1-3 y 8-10, pero cada una usa uno o más exones 4-7, en una combinación diferente al centro. Se ha reportado que los patrones de expresión de las isoformas de *MEG3* son tejido y tipo celular-específicas. La función de cada isoforma consiste en estimular la *trans*-activación de *p53* y suprimir el crecimiento tumoral (Zhang, *et al.*, 2010).

Con los resultados obtenidos por los algoritmos de catRAPID, sumado a que sólo se pudo identificar a la isoforma 2 en MCF-10A (Fig. 19b), se sugiere que las isoformas de *lincRNA_mr* presentan una expresión diferencial en tejidos. Dada su interacción variable con PRC2, quizá exista una función para las isoformas de este lincRNA, como lo que se reporta para *MEG3* (Zhang, *et al.*, 2010). Cabe resaltar que el hecho de que los ensayos se hayan realizado con este complejo represor no descarta la posibilidad de que pudiera darse una interacción con un complejo activador o de algún otro tipo. En el caso de la asociación entre *lincRNA_mr2* y SUZ12, esta interacción permite realizar una serie de inferencias funcionales. Si bien, se pueden realizar ensayos bioquímicos adicionales para confirmar dicha interacción, este lincRNA fue co-precipitado cuando se recuperaron los transcritos unidos a EZH2 y SUZ12 en un estudio publicado en el 2009 (Khalil, *et al.*, 2009). Además, el reporte de catRAPID omics realizado en este trabajo sugiere que *lincRNA_mr2* muestra el mayor valor de interacción con SUZ12, aun cuando este algoritmo mide la asociación entre este lncRNA y todas las proteínas humanas (Fig. 19).

Sin embargo, la arquitectura exónica de *lincRNA_mr* no es lo que atrae la mayor atención, sino la cercanía que guarda con el transcrito divergente *linc00882* (Fig. 17). En el genoma de los mamíferos, se estima que el 11% de los genes presentan una estructura cabeza-cabeza, con sus sitios de inicio de la transcripción separados por menos de 1,000 bp (región intergénica) y por eso se le designan promotores bidireccionales. Se cree que los promotores bidireccionales co-regulan la expresión de ambos genes por medio de elementos regulatorios compartidos; sin embargo, la función de los genes no necesariamente está correlacionada (Adachi & Lieber, 2002; Laszkiewicz, *et al.*, 2012).

La posibilidad de que exista un promotor común para *lincRNA_mr* y *linc00882* representaría el primer ejemplo de un promotor bidireccional que regulara a dos lincRNAs. Esta región cumple con todos los criterios necesarios para considerarse un promotor bidireccional que se han reportado en la literatura: una separación por menos de 1,000 bp entre dos genes con acomodo cabeza-cabeza, transcripción divergente (donde los genes en la pareja se transcriben uno de la cadena sentido y el otro de la antisentido) y la presencia de una isla CpG

en la región intergénica y dentro del primer exón de cada gen (Fig. 17) (Adachi & Lieber, 2002). En un segundo escenario, se podrían encontrar dos promotores cercanos en esta región, cada uno regulando la expresión transcripcional de uno de los lincRNAs. Sin embargo, dado que la caracterización se llevó a cabo computacionalmente, deberán realizarse pruebas funcionales para validar que, en efecto, se trata de un promotor único.

Mecanismo de acción tipo *cis* por parte del *lincRNA_mr* sobre el gen codificante *cbl-b*

Uno de los resultados más relevantes de este trabajo es observar que la expresión de *lincRNA_mr2* es necesaria para la expresión del gen no codificante *cbl-b* (Fig. 25). Para entender cómo puede afectar la desregulación del *lincRNA_mr2* a los distintos procesos celulares, decidimos estudiar más acerca de la función de esta ligasa de ubiquitina E3. Entre los reportes más importantes de la literatura, encontramos que el gen *cbl-b* se ha visto implicado en procesos apoptóticos por vía extrínseca, al controlar la activación de algunas MAPKs y otras señales. Además, su gen homólogo *c-Cbl*, regula la degradación de receptores de muerte posterior a la activación por TRAIL (Yan, *et al.*, 2012). Esta asociación del gen *cbl-b* a procesos apoptóticos le ha dado la clasificación de proto-oncogen (Croce, 2008).

Como se menciona en los resultados, algunos estudios reportan que el abatimiento del gen *cbl-b* conlleva consecuencias como la formación del complejo DISC apoptótico en las balsas lipídicas (Xu, *et al.*, 2012) y el aumento en la expresión de los receptores DR4 y DR5 (Yan, *et al.*, 2012). Debido a que la acción del *lincRNA_mr2* se encuentra arriba de la del gen *cbl-b* (ya que si no fuera de esta forma no deberían bajar los niveles del proto-oncogen al abatirse el lincRNA), se puede pensar que la baja expresión del transcrito no-codificante mostrará un resultado similar a un abatimiento de la expresión del gen *cbl-b*.

Para investigar si existía un aumento en la muerte celular de la línea que tiene abatido al *lincRNA_mr2*, se realizó un ensayo de TUNEL, donde se observó una mayor señal de muerte celular en una condición donde este lincRNA disminuye. Para complementar esta observación, se encuentran en proceso réplicas biológicas y de seguirse identificando muerte celular, se podrían utilizar anticuerpos para caspasas 3 y 8, específicos de la vía extrínseca de apoptosis.

Otro resultado interesante es la observación de que el aumento exógeno de la expresión del *lincRNA_mr2* no aumenta la cantidad de transcrito del gen *cbl-b*. Este no es el primer estudio donde se observa que una sobre-expresión no recupera los niveles de expresión de su blanco génico. Por el contrario, resultados como éste han servido de evidencia para argumentar que el transcrito no-codificante actúa verdaderamente en *cis*, pudiendo quizá permanecer anclado a su sitio de transcripción para ejercer su función y no actuando por medio de su

producto (Guttman & Rinn, 2012; Lee, 2012). Aunque faltarían estudios para poder confirmar que ambos genes se transcriben del mismo alelo, basándonos en los resultados obtenidos a lo largo de este estudio podemos argumentar que existe una regulación tipo *cis* del *lincRNA_mr2* sobre el gen codificante *cbl-b*.

Modelo preliminar de acción para el *lincRNA_mr*

Al momento, se podrían proponer 2 posibles mecanismos de acción para *lincRNA_mr*: el primero es que funcione como un lncRNA guía, que ayude a reclutar complejos de regulación epigenética, como PRC2, a sitios blancos en el genoma como se ha descrito para diversos lncRNAs (Rinn & Chang, 2012; Tsai, *et al.*, 2010). Este es un aspecto que queda pendiente por ser demostrado experimentalmente. En este caso, se esperaría que a mayor cantidad del transcrito *lincRNA_mr2* existieran menores niveles de expresión del gen *cbl-b*, ya este gen es uno de sus blancos, pero esto no es lo que se observó. Alternativamente, si existiera menor expresión del *linc00882* a mayores niveles de *lincRNA_mr2*, se podría proponer que *linc00882* fuera un regulador negativo del gen *cbl-b* y esta relación de regulación explicaría la correlación positiva entre el *lincRNA_mr2* y el gen *cbl-b*. También debemos considerar un escenario distinto, donde el *lincRNA_mr2* se asocie a otras proteínas que se involucren en la activación de la transcripción.

En un segundo escenario, podría ser que el *lincRNA_mr* actúe como un anzuelo para complejos remodeladores de la cromatina, asociándose a ellos pero evitando que se unan a regiones del genoma (Esteller, 2011; Rinn & Chang, 2012). De esta manera, cuando existan mayores cantidades del *lincRNA_mr2*, se podría asociar con mayor frecuencia a complejos represores como PRC2 y así permitir la expresión del gen *cbl-b*. Una de las consecuencias que deberían observarse, de ser éste el mecanismo, es que el estado de la cromatina se modificara en el promotor del gen *cbl-b* en contextos que posean niveles variables del *lincRNA_mr2*. Para poder apoyar esta hipótesis, se decidió realizar un ensayo de inmuno-precipitación de la cromatina con el anticuerpo H3K4me3. Aunque el experimento se encuentra en proceso, estos resultados podrían comenzar a elucidar el modelo de acción de *lincRNA_mr2* sobre su blanco en *cis*.

CONCLUSIONES

- ✓ Existe una reducción en los niveles de expresión de *lincRNA_mr* en la línea celular de cáncer mamario MCF7 cuando se compara contra la línea normal MCF-10A.
- ✓ *LincRNA_mr2* es la isoforma que se expresa mayoritariamente en la línea celular MCF-10A.
- ✓ Existe una correlación positiva entre los niveles de expresión del transcrito largo no-codificante *lincRNA_mr2* y el gen codificante *cbl-b*.
- ✓ La expresión del *lincRNA_mr2* es necesaria para mantener la expresión del gen *cbl-b*.
- ✓ La sobre-expresión del *lincRNA_mr2* no incrementa la expresión del gen *cbl-b*.
- ✓ Existe una regulación tipo *cis* entre el *lincRNA_mr2* y el gen *cbl-b*.
- ✓ Preliminarmente, existe un incremento significativo en las células positivas a TUNEL cuando el *lincRNA_mr* se encuentra abatido, posiblemente implicando a este lincRNA en el proceso de apoptosis.

PERSPECTIVAS

Estudiar la expresión de *lincRNA_mr* en tejido primario para comprobar si se comporta de manera similar *in vivo*.

Este estudio se realizó en su totalidad en líneas celulares establecidas y los únicos datos que tenemos de los niveles de expresión del *lincRNA_mr in vivo* provienen de los reportes del atlas de lincRNAs (Gibb, *et al.*, 2011). Sin embargo, resultaría muy interesante validar la expresión de este transcrito en un tejido primario normal y en otro canceroso. Esto nos podría brindar evidencia de si es común que exista una baja en la expresión del *lincRNA_mr* en un tejido canceroso e incluso si se pueden observar cambios en el transcrito estadio-específicos.

Cuantificación del transcrito divergente al *lincRNA_mr* y caracterización del posible promotor divergente

La arquitectura del locus 3q13.12 concuerda con los reportes donde se describe cómo luce un promotor bidireccional (Adachi & Lieber, 2002). Si existiera un promotor común para el *lincRNA_mr* y el *linc00882*, esto implicaría que ambos transcritos podrían estarse regulando juntos y quizá los dos podrían contribuir en un mecanismo de acción. Para abordar esta hipótesis, una de las perspectivas es poder cuantificar la expresión del *linc00882* en nuestro modelo de estudio, para ver si existe alguna relación entre los niveles de expresión de estos dos lincRNAs y del gen *cbl-b*.

Además, se podría realizar un análisis dual de luciferasa reportera (DLR) para identificar si la secuencia predicha como el promotor puede activar a ambos transcritos. En breve, se podrían clonar fragmentos de diversas longitudes del supuesto promotor (para predecir también la región mínima) y probar estas secuencias en ambas orientaciones, dentro de un vector con expresión de un reportero (Laszkiewicz, *et al.*, 2012).

Estudiar qué sucede con los niveles de expresión del *lincRNA_mr*, el *linc00882* y el gen *cbl-b* cuando se abate la proteína “insulator” CTCF.

Recientemente, se ha descrito que la proteína “insulator” CTCF se puede asociar a múltiples lincRNAs (Saldaña-Meyer, *et al.*, 2014). Para este estudio, se utilizaron tecnologías de entrecruzamiento que permitieron identificar los puntos de unión CTCF-lincRNAs. Al analizar si alguno de los transcritos (*lincRNA_mr* o *linc00882*) muestra asociación con CTCF, se observó que existe señal en el cuerpo del gen del *linc00882*. Una pregunta que surgió a partir de este descubrimiento es entender si el mecanismo de acción de estos dos transcritos involucra a CTCF. Para contestar a esta pregunta, se podría hacer uso de plásmidos que al integrarse al

genoma de una línea celular puedan abatir a CTCF ya sea por sistemas constitutivos o inducibles. De observarse alguna desregulación en el *lincRNA_mr* o el *linc00882* en un contexto CTCF-abatido, podría comenzar a describirse un nuevo modelo de acción que podría estar altamente conservado.

Identificación de los blancos génicos en *trans* para el *lincRNA_mr*

Se ha demostrado experimentalmente que muchos lncRNAs intergénicos tienen funciones de regulación transcripcional para genes localizados distalmente, es decir que actúan en *trans* (Kutter, *et al.*, 2012). En este trabajo se pudo determinar que el gen proximal *cbl-b* es blanco génico del *lincRNA_mr* en *cis*, sin embargo, aún resta determinar si este transcrito es capaz de ejercer funciones en *trans*. Para lograr contestar esta pregunta e identificar a genes regulados por el *lincRNA_mr*, se propone realizar ensayos de RNA-seq tanto en contextos donde hay ganancia como cuando existe pérdida de función para determinar los blancos funcionales de este transcrito. Esto se logra al comparar los estados celulares alterados y el contexto celular normal para analizar cuáles genes modifican sus niveles de expresión en respuesta al aumento o disminución del *lincRNA_mr*.

Llevar a cabo un ensayo de CHART o ChIRP para identificar los sitios del genoma donde se une al *lincRNA_mr* y las proteínas con las cuales se asocia.

Para poder verificar la unión del *lincRNA_mr2* en algún sitio del genoma, se podría utilizar un ensayo de ChIRP, utilizando oligonucleótidos específicos para las regiones a estudiar (Chu, *et al.*, 2011; Simon, *et al.*, 2011). Si además, estos análisis se realizaran a una escala global, se podrían abordar preguntas relacionadas con la regulación de estos transcritos, cuáles son las proteínas con las cuáles se asocian y corroborar si el *lincRNA_mr2* también puede actuar en *trans*.

MATERIALES Y MÉTODOS

Búsqueda bioinformática de lincRNA candidatos

Se utilizaron las bases de datos mencionadas en los antecedentes y que fueron descargadas de los materiales suplementarios de los artículos correspondientes (Cabili, M. *et al.* 2011; Gibb, *et al.*, 2011; Khalil, *et al.*, 2009). El primer paso fue seleccionar las coordenadas de los lincRNAs asociados a SUZ12 según se reportaba en el estudio de Cabili, M. *et al.* 2011, en la columna “Khalil_PRC2_bind”. Esta sección denotaba por medio de un sistema binario la interacción entre un lincRNA y PRC2. En un segundo filtro, se utilizó la tabla suplementaria 9 “Matriz de expresión de lincRNAs en tejido canceroso” del estudio de Gibb, *et al.* 2011 y se analizó cuáles de las coordenadas con interacción a PRC2 se encontraban en esta tabla. En el tercer paso, nos enfocamos en identificar a los lincRNAs en esta intersección que se desregulaban específicamente en glándula mamaria. En este caso, se buscó en la tabla suplementaria 11 “principales lincRNAs con expresión diferencial en cerebro, glándula mamaria y pulmón”, que incluye a los 20 lincRNAs con mayor tasa de cambio de tejido normal a canceroso en estos órganos.

Análisis bioinformático del transcrito lincRNA_mr

Análisis de elementos regulatorios en la región promotora

Se utilizó el visualizador del genoma UCSC browser, en su versión GRCh37/hg19 [<http://genome.ucsc.edu/cgi-bin/hgGateway>] (30) para caracterizar el locus donde se encuentra el transcrito *lincRNA_mr*. Para analizar el estado de la cromatina se configuró el track del estudio de las 9 líneas celulares (Ernst, *et al.*, 2011). Se agregaron los tracks con datos del consorcio de ENCODE para las modificaciones postraduccionales H3K4me3, H3K4me1 y H3K27ac, hipersensibilidad a la DNasa I, metilación del DNA e islas CpG. A su vez, se configuraron los tracks de ChIP-seq de ENCODE para una amplia gama de factores de transcripción y los niveles de expresión por RNA-seq. Para el estudio de los factores de transcripción con unión predicha se utilizó la base de datos JASPAR (v. 5.0_ALPHA) [<http://genome.ucsc.edu/cgi-bin/hgGateway>] (Mathelier, *et al.*, 2014). Para predecir secuencias de miRNAs, se utilizó el programa miRBase [<http://www.mirbase.org>] (Kozomara & Griffiths-Jones, 2014) y para identificar semillas DIANA-microT v3.0 [<http://diana.cslab.ece.ntua.gr/microT>] (Maragkakis, *et al.*, 2009).

Estudio de asociación de lincRNAs con diversas proteínas

Se descargaron las secuencias de aminoácidos de SUZ12, EED, EZH2 y RBAP48 humanos de la base de datos UniProt [<http://www.uniprot.org>] (UniProt Consortium, 2014) con identificadores sp|Q15022, sp|O75530, sp|Q15910 y sp|Q09028, respectivamente. A su vez, se obtuvieron las secuencias de cDNA para cada una de las isoformas de *lincRNA_mr* con la herramienta de exportación de datos en la página de *Ensembl* [<http://www.ensembl.org/index.html>] (EMBL-EBI & Wellcome Trust Sanger Institute, 2014). Se utilizó el programa catRAPID [<http://service.tartagliolab.com>] (Bellucci, *et al.*, 2011) para identificar la propensión de interacción entre un par RNA-proteína. En el caso de las isoformas 1, 3, 4 y 5 se utilizó el algoritmo catRAPID graphic, que acepta secuencias de 50-750 aminoácidos para las proteínas y de 50-1200 nt para RNA. Para la isoforma 2, que excede esta longitud, se utilizó la herramienta catRAPID omics [http://service.tartagliolab.com/page/catrapid_omics_group], que verifica la interacción entre un lincRNA y todo el proteoma en un organismo modelo (Bellucci, *et al.*, 2011). El algoritmo reporta una lista ordenada de parejas del RNA analizado con diversas proteínas, de acuerdo a la propensión de interacción (p -value <0.025). En el caso de los ncRNAs primero se encuentra el dominio de unión con HMMscan.

Estructuras secundarias

Para obtener las posibles estructuras secundarias se utilizó el servidor “mfold” del Instituto de RNA, Colegio de Artes y Ciencias, Universidad en Albany, NY [<http://mfold.rna.albany.edu/?q=mfold/rna-folding-form>] (Zuker, 2003).

Diseño in silico del shLincMR

Se identificó una secuencia de 23 bp que hibrida específicamente con el primer exón de *lincRNA_mr* para que pudiera abatir a 4 de 5 isoformas. Se utilizaron diversos algoritmos para elegir las mejores secuencias, filtrarlas y seleccionar un candidato. Los algoritmos son: siRNA Wizard v3.1 [<http://www.siRNAwizard.com/>] (InVivoGen, 2013), siRNA at Whitehead del MIT (N2[CG]N8[AUT]N8[AUT]N2 como plantilla) [<http://sirna.wi.mit.edu/>] (Yuan, *et al.*, 2004), BLOCK-iT RNAi Designer [<http://rnaidesigner.lifetechnologies.com/rnaiexpress/>] (Invitrogen, 2013) y siDESIGN Center de Thermo [<http://dharmacon.gelifesciences.com/design-center/>] (GE Healthcare, 2014). Posteriormente, se diseñaron oligonucleótidos de alrededor de 60 nt que sirvieran para formar el shRNA y que siguieran la estructura: HpaI-sentido-giro/loop-antisentido-terminador-XhoI. Los nombres de los sesentámeros utilizados son shRNA_lincMR251_F y R (tabla de oligos). Para validar formación de estructura secundaria se utilizó el programa GeneBee [http://www.genebee.msu.su/services/rna2_reduced.html] (Brodsky, *et al.*, 1995).

Construcción de vector de pérdida de función

Preparación del vector (pLentiLox3.7-Puro)

Se realizó una doble digestión del vector pLentiLox 3.7-Puro con las enzimas HpaI y XhoI en un volumen final de 30 ul a 37°C-ON y se hizo una purificación por fenol cloroformo con elución en 10 ul. Se agregó fosfatasa alcalina por 3 hrs a 37°C y se purificó el vector por electroforesis seguido de purificación de banda con el kit QIAGEN II (QIAGEN).

Preparación del inserto (shlincMR)

Se resuspendieron los oligonucleótidos shRNA_lincMR251_F y R, ajustándolos a una concentración cercana a las 100 pmol/ul. Se hibridaron ambos cebadores agregando 1ul de cada uno y 48 ul del buffer de hibridación (acetato de potasio 1 M, HEPES-KOH pH 7.4 300 mM y acetato de magnesio 200 mM). La mezcla se incubó a 95°C - 4 min → 70°C - 10 min → decremento de temperatura hasta 4°C, 0.1°C/min.

Construcción del vector pLL_shLincMR

Se ligó vector e inserto usando ligasa T4 (NEB) a 14°C por 9 hrs en proporciones 1:0, 1:1 y 1:3 respectivamente. El producto de ligación se transformó en la cepa TOP10 por shock térmico (incubación de 20 min en baño de hielo → 42°C-45 segundos → 30°C-1 hr). Se recuperaron las bacterias por centrifugación de 6,000 rpm por 2 min y se inocularon cajas de agar con resistencia a ampicilina, las cuales crecieron a 30°C ON. Al día siguiente, se seleccionaron 10 colonias, 7 de las cuales contenían al fragmento. Se validó la construcción por secuenciación y se utilizó pLL_shLincMR obtenido de midi-preps (QIAGEN) para la infección lentiviral.

Cultivo de líneas celulares

Todas las líneas celulares se crecieron a 37°C con 5% de CO₂. En el caso de MCF7, ZR-75-30, T-47D, MDA-MB-231 y LNCaP, se utilizó como medio de cultivo RPMI (GIBCO) con concentraciones finales de suero fetal bovino al 10% y penicilina/estreptomicina al 1%. En el caso de MCF-10A se utilizó un medio de cultivo DMEM/F12 (GIBCO) en proporción 3:1 adicionado con 10% suero fetal bovino, 1X L-glutamina (2 mM), 1X antimicótico (fungizona 250 ng/ml), 1X penicilina/estreptomicina, EGFrh 10 ng/ml (factor de crecimiento epidérmico recombinante humano, GIBCO), hidrocortisona 1 ug/ml (Invitrogen) e insulina humana 100 u/ml (GIBCO). Para tener una mejor adherencia en la línea celular LNCaP, se utilizó un recubrimiento de L-Poli-K en las cajas de Petri.

Infección lentiviral de cultivos celulares

Se obtuvieron y purificaron las partículas lentivirales por co-transfección de los 4 vectores de expresión (incluyendo el vector de expresión) que codifican para las diferentes partículas de ensamble del virus en células HEK-293. El primer día, se realizó una mezcla de los 4 vectores, 400 ul de CaCl 1.25 M (Sigma), HBS 2X (Sigma) y se agregaron 2 ml de la mezcla sobre toda la caja de HEK-293. Las células se incubaron por 4 hrs y se lavaron con PBS 1X (NaCl 137 mM, KCl 2.7 mM, Na₂HPO₄ 10 mM y KH₂PO₄ 1.8 mM) a pH 7.4. Después de incubar de 36 - 48 hrs, se recolectó el sobrenadante de las cajas y se centrifugó 7 min a 2,000 rpm a 4°C. Se recuperó el sobrenadante y se ultra-centrifugó 90 min a 25,000 rpm a 4°C. Se recuperó el sobrenadante y el pellet se dejó secando unos 10 min boca abajo para luego resuspenderlo en 50 ul de PBS pH 7.4. El cuarto día se resuspendieron los virus y se congelaron a -80°C.

Para la infección, se plaquearon 10,000 células; la mañana siguiente se retiró el medio de cultivo y se añadieron 500 ul de medio con polibreno en cada pozo. Se añadió medio y se inocularon 10 ul de virus que se agregaron al medio y se incubó ON a 37°C. Al día siguiente, se agregaron 3 ml de medio y se incubaron las células 2 o 3 días más para permitir la expresión viral. Se seleccionó agregando puromicina 2X [1 mg/ml]. El protocolo se estableció en nuestro laboratorio (122 N IFC, UNAM) por Georgina Guerrero Avendaño y Edgar González Buendía.

Construcción de vector de sobre-expresión

Preparación de los vectores (pcDNA 3.1 hygro (+) y pcLMR_ex1)

Primero se digirieron 2 ug del vector pcDNA 3.1 hygro(+) en una doble digestión con HindIII y XhoI en un volumen final de 30 ul, incubando 4 hrs a 37°C e inactivando 20 min a 65°C. Cuando se tuvo el vector pcLMR_ex1 (como se describe más adelante), este plásmido fue digerido serialmente con ApaI en un volumen de 15 ul por 3 hrs a 25°C y luego con XhoI, ajustando la reacción previa a un volumen de 30 ul, incubando 3 hrs a 37°C e inactivando 20 min a 65°C. Una vez que se realizaron las digestiones, se defosforilaron ambos vectores con fosfatasa alcalina por 1 hr a 37°C, seguido de su inactivación a 70°C por 10 min. Ambos vectores se verificaron con un gel de agarosa al 1% y se purificaron con el kit QIAGEX II.

Preparación de los insertos (lincRNA_mr_exón1 y lincRNA_mr_exón2)

Se realizaron dos PCRs, como se describe anteriormente, logrando amplificar productos que se purificaron de un gel al 1% usando el kit QIAGEX II. Se realizó una doble digestión del exón 1 con HindIII y XhoI en un volumen final de 70 ul, incubando 4 hrs a 37°C e inactivando 20 min a 65°C. El exón 2 tuvo que ser clonado en un primer paso al vector pGEM-T easy (Promega),

donde se validó por secuenciación que era el fragmento esperado y partiendo de este plásmido se realizó una digestión serial. Primero, se digirió con Apal en un volumen de 15 ul por 3 hrs a 25°C y esta reacción se ajustó a un volumen de 30 ul, agregando XhoI, incubando 3 hrs a 37°C. Ambos exones digeridos se corrieron en un gel de agarosa al 1% y se purificaron.

Construcción de los vectores pcLMR_ex1 y pcLMR2_FL

En el caso del vector pcLMR_ex1, se prepararon ligaciones en un volumen final de 20 ul usando proporciones 1:0, 1:1 y 1:3 (vector:inserto). Se utilizó la enzima T4 (NEB) y se incubó ON a 16°C, seguido de su inactivación 10 min a 65°C. Se transformaron 5 ul en 70 ul de células Top10 y se inocularon en un plato con resistencia a ampicilina. Después de seleccionar algunas colonias, se hicieron mini-preps y se verificó que existieran los patrones de digestión esperados con BamHI, Aval y HpaI. Finalmente, se validó por secuencia que se había insertado correctamente el exón 1 completo.

Para el plásmido pcLMR2_FL se siguieron los mismos pasos descritos con ligeros cambios, como el que la ligación fuera en proporción 1:2 y que el crecimiento de las bacterias se llevara a cabo a 30°C.

Transfección estable de cultivos celulares

Se digirieron 20 ug de cada plásmido usando BglII en un volumen de 30 ul ON a 37°C. Al día siguiente, se corrió un gel para validar la linearización de los vectores y se purificó el resto de la muestra por lavados de fenol-cloroformo-alcohol isoamílico y de alcohol 100%. Se partió de cultivos de MCF7 con 50,000 células a las cuales se les agregó una mezcla de lipofectamina (Invitrogen) y DNA de cada vector de interés en medio RPMI sin complementar. Se incubó por 6 hrs a 37°C y se cambió el medio por RPMI con 10% de suero fetal bovino y 1% penicilina/estreptomina. Se incubó a 37°C por 66 hrs y se agregó higromicina a [0.9 mg/ml]. Los cultivos se mantuvieron en selección por 10 días, cuando el mock ya no mostraba células vivas, para luego esperar a que se expandieran los cultivos.

Extracción de ácidos nucleicos

Extracción de RNA

Posterior a la tripsinización, se recolectaron las células por centrifugación a 700 rpm por 5 min y se retiró el sobre-nadante. Se agregó 1 ml de TRIzol (Invitrogen) al pellet y se dejó incubando 5 min. Se agregaron 200 ul de cloroformo, incubando a temperatura ambiente por 2 min y se realizó una centrifugación de 15 min a 4,500 rpm. Se recuperó la fase superior, que contiene al

RNA y se depositó en un tubo Falcon de 15 ml, agregando un volumen de isopropanol (~250 ul). Se mezcló y se dejó incubando 10 min a temperatura ambiente para seguir con una centrifugación de 10 min a 4,500 rpm. Se descartó el sobre-nadante por decantación, dejando un pellet blanco al fondo del tubo y se agregaron 500 ul de alcohol etílico al 70%. Se hizo un último paso de centrifugación a 4,500 rpm por 5 min y se descartó el sobre-nadante. Se dejó secar el pellet por 5-10 min y se resuspendió con agua DEPC al 0.1%.

Electroforesis para RNA

Una vez que se cuantificó la concentración de RNA utilizando un NanoDrop 2,000c, se realizó una electroforesis utilizando un gel de agarosa al 1.2%, con MOPS/EDTA (0.2 M MOPS, 50 mM Acetato de sodio, 10 mM EDTA, pH 7) y 5% de formaldehído al 37%. Se ajustó el volumen de la muestra a 5 ul con agua DEPC, para correr de 1-3 ug de RNA. Se agregó 1 ul de buffer para muestras de electroforesis (formamida desionizada, 10X MOPS, formaldehído, agua DEPC, glicerol y azul de bromofenol al 10%) y se calentaron a 65°C por 15 min. Se agregó 1 ul de bromuro de etidio (200 ug/ml) en la muestra y se cargó en el gel para correrlo a 80 V por 1 hr a temperatura ambiente.

Extracción de DNA genómico

Se recuperaron células de cultivo por centrifugación a 700 rpm por 5 min y el pellet se depositó en un tubo eppendorf. Se agregaron 250 ul del Mix I (Tris 200 mM pH 8.5, NaCL 400 mM) y se resuspendieron las células. Se agregaron 250 ul del Mix II (SDS 0.4%, EDTA 10 mM, RNasa A 50 ug/ml, proteinasa K 50 ug/ml) e incubaron a 50°C ON. Al día siguiente, se realizaron dos precipitaciones agregando 500 ul de fenol-cloroformo-alcohol isomálico, mezclando por 5 min en un vórtex y centrifugando 10 min. Se hicieron dos precipitaciones más con 500 ul de cloroformo, mezclando 3 min en un vórtex y centrifugando 10 min. Se lavó con 500 ul de isopropanol por 2 hrs, y se mezcló con un vórtex para luego centrifugar por 15 min. Se retiró el sobrenadante, cuidando no despegar el pellet de DNA. Se agregaron 500 ul de etanol 70% y centrifugar 15 min para luego retirar el sobrenadante. Se dejó secar el pellet por 10 min a temperatura ambiente y se agregaron 120 ul H₂O. Todos los pasos de centrifugación se realizaron a 13,000 rpm, que es la máxima velocidad.

Extracción de plásmidos por método de lisis alcalina (mini-prep)

Se recuperaron los cultivos celulares, crecidos por más de 6 hrs a 30°C o 37°C, en 3 ml de LB con antibiótico por una centrifugación de 2 min a 13,000 rpm a temperatura ambiente. Se

decantó el sobre-nadante y se agregaron 100 ul de la solución I (glucosa 50 mM, Tris-HCl 25 mM pH 8.0, EDTA 10 mM pH 8.0). Se mezcló con un vórtex hasta resuspender todas las células y se incubó a temperatura ambiente por 3 min. Se agregaron 200 ul de la solución II recién preparada (NaOH 10 N y SDS 10%) y se mezcló por 10 segundos. Se incubó en hielo por 5 min y se agregaron 150 ul de la solución III fría (acetato de potasio 5 M, ácido acético glacial, agua desionizada). Se mezcló invirtiendo cada tubo y se incubó 10 min en hielo. Se centrifugó 10 min y se transfirió el sobrenadante a un tubo limpio. Se añadió un volumen de fenol-cloroformo-isoamílico, se mezcló 30 segundos y se centrifugó por 10 min a temperatura ambiente para separar bien las fases. Se recuperó el sobrenadante en un tubo limpio y se agregó 1 ml de etanol 100%. Se mezcló e incubó de 10-20 min a temperatura ambiente para luego centrifugar 15 min. Se decantó y lavó el pellet con etanol 70% una vez y se centrifugó otros 10 min, seguido de la decantación del etanol. Se dejó secar el pellet 10 min y se resuspendió en 100 ul de agua desionizada. Se agregaron 2 ul de RNasa a cada tubo y se incubó por 1 hr a 37°C. Los pasos de centrifugación se realizaron a 13,000 rpm a 4°C, a menos que se indique otra condición.

Extracción de plásmidos (Midi-prep)

Se utilizó el kit de QIAGEN "Plasmid Midi and Maxi Kits", siguiendo las instrucciones del fabricante. Brevemente, se crecieron cultivos de 100 ml ON en LB con antibiótico y se recuperaron por centrifugación a 6,000 rpm por 15 min a 4°C. Los cultivos bacterianos se resuspendieron, lisaron y se separó la fracción que contenía los plásmidos usando diversos buffers y pasos de centrifugación. Se recuperaron los plásmidos por medio de una columna y pasos de lavado con distintos buffers. Se precipitaron los plásmidos con isopropanol y etanol por centrifugación a 11,000 rpm por 10-30 min a 4°C. Finalmente, se resuspendió el pellet seco en 100 ul de agua desionizada.

Preparación de cDNA por Retro-Transcripción (RT)

Para eliminar cualquier contaminación de DNA genómico, se realizaron reacciones de DNasa (Promega) a 37°C por 30 min (3 ug de RNA, 1X Buffer DNasa y H₂O para un volumen de 20 ul), seguido de 65°C por 10 min para inactivar a la enzima (1X Buffer STOP). En el caso de la RT, las reacciones se optimizaron para quedar a una concentración final de 5 mM de MgCl₂, 1X Buffer RT, 1 mM de cada dNTP, 15 u de AMV (Promega) y 0.5 ug de oligo (dT)₁₅ para un volumen de 30 ul, que se completó con H₂O inyectable. La reacción llevada a cabo se incubó a: 25°C por 10 min → 42°C por 60 min → 95°C por 5 min → 4°C por 5 min.

Una vez finalizada la reacción, se agregaron 50 ul de H₂O libre de nucleasas para llevar a un volumen final de 80 ul.

Reacción en cadena de la polimerasa (PCR)

KAPA Taq

Las reacciones se prepararon con KAPA Buffer 10X, 25 mM MgCl₂ (concentración final 3-4 mM), 10 mM dNTPs, 2.5-5 pmol oligonucleótidos, 1 ul de cDNA y 5 U/ul de enzima, en un volumen final de 20 ul. Las incubaciones se realizaron de acuerdo a las indicaciones del fabricante, con respectivas T_m's (tabla de oligonucleótidos).

Se utilizó esta enzima para poder amplificar los productos *lincRNA_mr_exón1*, *lincRNA_fs*, *lincRNA_mr1_369*, *lincRNA_mr3_242* y *lincRNA_mr5_136*.

Kapa HF Taq

Se prepararon las reacciones con KAPA HF Buffer GC 5X, 10 mM dNTPs, 5 pmol oligonucleótidos, 2.5 ul de gDNA y 1 U/ul de enzima, en un volumen de 50 ul. Las incubaciones se realizaron de acuerdo a las indicaciones del fabricante, con respectivas T_m's (tabla de oligonucleótidos). Se utilizó esta enzima para poder amplificar los productos *LincRNA_mr2_exón1* y *LincRNA_mr2_exón2*.

Dream Taq

Las reacciones se prepararon con Buffer 10X, 25 mM MgCl₂ (concentración final 3 mM), 10 mM dNTPs, 5 pmol oligonucleótidos, 1 ul de cDNA y 1.25 u de enzima, en un volumen de 20 ul. Las incubaciones se realizaron de acuerdo a las indicaciones del fabricante a una T_m de 60°C. Se utilizó esta enzima para poder amplificar el producto de *ccdc54*.

PCR en Tiempo Real (qRT-PCR)

Para los ensayos de qRT-PCR, las reacciones se realizaron en el sistema StepOne (Applied Biosystems) y se siguieron las condiciones óptimas para cada producto, agregando SYBR Green como fluoróforo.

Platinum Taq

Las reacciones se prepararon con Buffer 10X, 50 mM MgCl₂ (concentración final 3 mM), 10 mM dNTPs, 5 pmol oligonucleótidos, 1 ul de cDNA, 4x SYBR GREEN (Sigma) y 1 U/ul de enzima, en un volumen de 25 ul. Las incubaciones se realizaron de acuerdo a las indicaciones del

fabricante, con Tm de 60°C. Se utilizó esta enzima para poder amplificar los productos *lincRNA_mr2_285 bp*, *gapdh* y *cbl-b*.

KAPA Taq

Las reacciones se prepararon como se describe anteriormente, agregando también 4x SYBR GREEN, en un volumen final de 25 ul. Se utilizó esta enzima para poder amplificar los productos *lincRNA_mr_exón1* y *lincRNA_fs*.

Ensayo de Doubling Time

Se usaron 3 puntos muestrales por día por línea en placas de 24 pozos, donde los volúmenes usados fueron: 1ml de medio de cultivo, 250 ul de tripsina y 250 ul de PBS/pozo. En el día 0, se contó el número de células de un cultivo celular confluyente de la línea a analizar y se calculó el volumen que contuviera 10,000 células, las cuales se depositaron en un pozo. La placa se incubó a 37°C y 5% de CO₂ por 24 hrs. Desde el primer día y en adelante, se retiró el medio de los pozos a contar y se realizó un lavado con PBS. Una vez que se retiró el PBS, se agregaron 250 ul de tripsina/pozo. Se incubó a 37°C de 1-5 min para que se despegaran las células y se agregaron 750 ul de medio de cultivo para inactivar la tripsina. Para poder homogeneizar las células, se pipeteó con punta amarilla de arriba hacia abajo, intentando lavar todo el pozo y distribuyendo las células uniformemente. Se tomaron 10 ul por pozo y se depositaron en un tubo eppendorf, al cual se le agregaron otros 10 ul de azul de tripano y se mezclaron para poder cargar en la cámara de Neubauer. El número de células contadas y promediadas se multiplicó por 2 (factor de dilución). Al final, se utilizó la fórmula siguiente para obtener el doubling time: $DT = \frac{T \ln 2}{\ln(\frac{X_f}{X_i})}$, donde T es el tiempo de incubación, Xi = no. de células iniciales y Xf = no. de células finales.

Ensayo de TUNEL

Se utilizó el kit de detección *in situ* de muerte celular por fluoresceína (Roche), el cual detecta y cuantifica la apoptosis al etiquetar la ruptura de DNA (tecnología TUNEL) por un análisis de microscopía de fluorescencia. Brevemente, se crecieron 20,000 células MCF-10A en cajas de 12 pozos con cubre-objetos estériles en el fondo y se realizó la infección lentiviral. Después de la selección por puromicina (5 días), se retiró el medio, se hicieron 2 lavados con PBS y las células se fijaron con paraformaldehído frío al 4% en PBS, pH 7.4 por 1 hr a temperatura ambiente. Se lavó 3 veces con PBS y se incubó en una solución de bloqueo (H₂O₂ al 3% en metanol) por 10 min, seguido de un lavado con PBS. Se añadió la solución permeabilizadora

(Triton X-100, 0.1% en citrato de sodio 0.1%) por 2 min en hielo. Se lavaron las células 2 veces con PBS, se quitaron residuos líquidos con una toallita y se agregaron 35 ul de solución de etiquetado (31.5 ul) y solución de enzima (3.5 ul). Se incubó 1 hr a 37°C y 5% de CO₂ en la oscuridad en una cámara húmeda creada en la misma caja con parafilm y una toalla mojada. Se lavó 2 veces con PBS y se colocaron los cubreobjetos boca abajo en portaobjetos, sellando con barniz de uñas. Las células se observaron en un microscopio de fluorescencia usando un filtro U-MNB2 para fluoresceína.

Oligonucleótidos utilizados

Nombre	Secuencia (5' --> 3')	Tm (PCR)
lincRNA_mr1_SF	GAGAAGAAAAGCTAGACACTG	60
lincRNA_mr1_R	TCAGTTGGACACACATAAAC	60
lincRNA_mr1_LF	CTGGGCACACTGGCTTCAC	60
lincRNA_mr2_R	CTTATTCAGATGGCAGCATAG	60
lincRNA_mr3_R	CCAAGTAAATCATTACAGTACA	60
lincRNA_mr5_R	GCATTTGAATCTGGGCTTTGC	60
lincRNA_mr_ex1_F	AGCATCCTCATCCATTTGGG	60
lincRNA_mr_ex1_R	AGACGGAGCAAATCGGAATC	60
lincRNA_mr1_ex3_R	GAGTGCAGAAGCAGGTAAAG	60
CBLB_F	GAAGATGCTTTTGTATGCCCTC	60
CBLB_R	GAGCCAGGAGGTTTTGAATG	60
CCDC54_F	CAAACATACAACCAGCAAC	60
CCDC54_R	GTGGACTTTGGATAACTGGG	60
shRNA_lincMR251_F	GTTAACGGATCCTTAAGAGGTGAATTTCAAGAGAATTCACCTCTTAAGGATCCTTTTTTGGAACTCGAG	-
shRNA_lincMR251_R	CTCGAGTTCAAAAAAGGATCCTTAAGAGGTGAATTTCTTGAATTCACCTCTTAAGGATCCGTTAAC	-
GAPDH_F	CCACTCCTCCACCTTTGAC	60
GAPDH_R	ACCCTGTTGCTGTAGCCA	60
LncMR2_ex1_F	GATAAAGCTTACCCACGCGGCGCAGCTCC	57
LncMR2_ex1_R	TATACTCGAGCTGGGAATGCGCTTGCTGGTGAG	57
LncMR2_ex2_F	GCTACTCGAGGTGGTCAGGCTCAGGATGGAC	65
LncMR2_ex2_R	TATTGGGCCCTCTGTCTCTCTGGCTTCCTC	65
Chip-cblb-Fw	AGTACACAATGGCCCCAGAG	62
Chip-cblb-Rv	TGCCACCCCTTATTGGTTAG	60
Chip-MyoD-Fw	AGGAAGTGAGGAAGTTAGGG	60
Chip-MyoD-Rv	CAAAACAGGCTACCCCAAGT	62

REFERENCIAS

- Adachi, N., & Lieber, M. R. (2002). Bidirectional gene organization: a common architectural feature of the human genome. *Cell*, 807-809.
- Allis, C., Jenuwein, T., & Reinberg, D. (2007). *Epigenetics*. New York: Cold Spring Harbor Laboratory Press.
- Altogen Biosystems. (s.f.). *MCF-7 Cells*. www.mcf7.com/mcf7.html
- American Type Culture Collection. (2014). *ATCC*. www.atcc.org
- Askarian-Amiri, M. E., Crawford, J., French, J. D., Smart, C. E., Smith, M. A., Clark, M. B., . . . Mattick, J. S. (2011). SNORD-host RNA Zfas1 is a regulator of mammary development and a potential marker for breast cancer. *RNA*, 878-891.
- Augui, S., Nora, E. P., & Heard, E. (2011). Regulation of X-chromosome inactivation by the X-inactivation centre. *Nat. Rev. Genet.*, 429-442.
- Baker, M. (2011). Long noncoding RNAs: the search for function. *Nat. Methods*, 379-383.
- Bellucci, M., Agostini, F., Masin, M., & Tartaglia, G. G. (2011). Predicting protein associations with long noncoding RNAs. *Nat. Methods*, 444-445.
- Bernstein, E., & Allis, D. (2005). RNA meets chromatin. *Gene. Dev.*, 1635-1655.
- Bhutani, N., Burns, D. M., & Blau, H. M. (2011). DNA demethylation dynamics. *Cell*, 866-872.
- Boyerinas, B., Park, S.-M., Hau, A., Murmann, A. E., & Peter, M. E. (2010). The role of let-7 in cell differentiation and cancer. *Endocr-Relat Cancer*, F19-F36.
- Brodsky, L. I., Ivanov, V. V., Kalaydzidis, Y. L., Leontovich, A. M., Nokolaev, V. K., Feranchuk, S. I., & Drachev, V. A. (1995). GeneBee-NET: Internet-based server for analyzing biopolymers structure. *Biochemistry*, 923-928.
- Buske, F. A., Bauer, D. C., Mattick, J. S., & Bailey, T. L. (2012). Triplexator: detecting nucleic acid triple helices in genomic and transcriptomic data. *Genome Res.*, 1-10.
- Cabili, M. N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., & Rinn, J. L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Gene. Dev.*, 1915-1927.
- Cheetham, S. W., Gruhl, F., Mattick, J. S., & Dinger, M. E. (2013). Long noncoding RNAs and the genetics of cancer. *Brit. J. Cancer*, 1-7.
- Chu, C., Qu, K., Zhong, F. L., Artandi, S. E., & Chang, H. Y. (2011). Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol. Cell*, 667-678.
- Croce, C. M. (2008). Oncogenes and Cancer. *New Engl. J. Med.*, 502-511.

- Croce, C. M. (2010). LINCing chromatin remodeling to metastasis. *Nat. Biotechnol.*, 931-932.
- Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., . . . Guigó, R. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.*, 1775-1789.
- Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., . . . Gingeras, T. R. (2012). Landscape of transcription in human cells. *Nature*, 101-108.
- Du, Z., Fei, T., Verhaak, R. G., Su, Z., Zhang, Y., Brown, M., . . . Liu, S. (2013). Integrative genomic analyses reveal clinically relevant long noncoding RNAs in human cancer. *Nat. Struct. Mol. Biol.*, 908-913.
- EMBL-EBI & Wellcome Trust Sanger Institute. (Febrero de 2014). *Ensembl*. www.ensembl.org/index.html
- Engel, L. W., Young, N. A., Tralka, T. S., Lippman, M. E., O'Brien, S. J., & Joyce, M. J. (1978). Establishment and characterization of three new continuous cell lines derived from human breast carcinomas. *Cancer Res.*, 3352-3364.
- Ernst, J., Kheradpour, P., Mikkelsen, T. S., Shores, N., Ward, L. D., Epstein, C. B., . . . Bernstein, B. E. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*, 43-49.
- Esteller, M. (2011). Non-coding RNAs in human disease. *Nat. Rev. Genet.*, 861-874.
- Fatica, A., & Bozzoni, I. (2014). Long non-coding RNAs: new players in cell differentiation and development. *Nat. Rev. Genet.*, 7-21.
- Felsenfeld, G., & Groudine, M. (2003). Controlling the double helix. *Nature*, 448-453.
- García Solís, P. (2006). *Análisis del efecto de dietas ricas en yodo y ácido retinóico sobre la inducción de cáncer mamario en ratas tratadas con el cancerígeno N-metil-N-nitrosourea (MNU)*. UNAM-UAQ, Juriquilla, Querétaro: Tesis doctoral.
- GE Healthcare. (2014). *Dharmacon RNAi & Gene Expression*. <http://dharmacon.gelifesciences.com/design-center/>
- Genome Bioinformatics Group of UC Santa Cruz. (2014). *UCSC Genome Browser*. <http://genome.ucsc.edu/cgi-bin/hgGateway>
- Gibb, E. A., Vucic, E. A., Enfield, K. S., Stewart, G. L., Lonergan, K. M., Kennett, J., . . . Lam, W. L. (2011). Human cancer long non-coding RNA transcriptomes. *PLoS one*, 1-10.
- Guil, S. & Esteller, M. (2012). Cis-acting non-coding RNAs: friends and foes. *Nat. Struct. Mol. Biol.*, 1068-1075.

- Gupta, R. A., Shah, N., Wang, K. C., Kim, J., Horlings, H. M., Wong, D. J., . . . Chang, H. Y. (2010). Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature*, 1071-1076.
- Gutschner, T., & Diederichs, S. (2012). The hallmarks of cancer, a long non-coding RNA point of view. *RNA Biol.*, 703-719.
- Guttman, M., & Rinn, J. L. (2012). Modular regulatory principles of large non-coding RNAs. *Nature*, 339-346.
- Guttman, M., Amit, I., Garber, M., French, C., Lin, M. F., Feldser, D., . . . Lander, E. S. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*, 223-227.
- Hake, S. B., & Allis, C. (2006). Histone H3 variants and their potential role in indexing mammalian genomes: the "H3 barcode hypothesis". *Proc. Natl. Acad. Sci.*, 6428-6435.
- Hanahan, D., & Weinberg, R. A. (2000). The Hallmarks of Cancer. *Cell*, 57-70.
- Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell*, 646-674.
- Huarte, M., Guttman, M., Feldser, D., Garber, M., Koziol, M., Kenzelmann-Broz, D., . . . Rinn, J. L. (2010). A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell*, 409-419.
- Imbalzano, K. M., Tatarkova, I., Imbalzano, A., & Nickerson, J. A. (2009). Increasingly transformed MCF-10A cells have a progressively tumor-like phenotype in three-dimensional basement membrane culture. *Cancer Cell Int.*, 1-11.
- Invitrogen. (2013). *BLOCK-iT RNAi Express*.
<http://rnaidesigner.invitrogen.com/rnaexpress/rnaiDesign.jsp>
- InVivoGen. (2013). *siRNA Wizard V3.1*. www.sirnazard.com/design.php
- Kallen, A. N., Zhou, X.-B., Xu, J., Qiao, C., Ma, J., Yan, L., . . . Huang, Y. (2013). The imprinted H19 lncRNA antagonized let-7 microRNAs. *Mol. Cell*, 1-12.
- Khalil, A. M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., . . . Rinn, J. L. (2009). Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl. Acad. Sci.*, 11675-11680.
- Kornienko, A. E., Guenzl, P. M., Barlow, D. P., & Pauler, F. M. (2013). Gene regulation by the act of long non-coding RNA transcription. *BMC Biology*, 1-14.
- Koziol, M., & Rinn, J. L. (2010). RNA traffic control of chromatin complexes. *Curr. Opin. Genet. Dev.*, 142-148.

- Kozomara, A., & Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucl. Acids Res.*, D68-D73.
- Kutter, C., Watt, S., Stefflova, K., Wilson, M. D., Goncalves, A., Ponting, C. P., . . . Marques, A. C. (2012). Rapid turnover of long noncoding RNAs and the evolution of gene expression. *PLoS Genet.*, 1-15.
- Lanz, R. B., McKenna, N. J., Onate, S. A., Albrecht, U., Wong, J., Tsai, S. Y., . . . O'Malley, B. W. (1999). A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex. *Cell*, 17-27.
- Laszkiewicz, A., Sniezewski, L., Kasztura, M., Bzdion, L., Cebrat, M., & Kisielow, P. (2012). Bidirectional activity of the NWC promoter is responsible for RAG-2 transcription in non-lymphoid cells. *PLoS one*, 1-9.
- Lee, J. T. (2012). Epigenetic regulation by long noncoding RNAs. *Science*, 1435-1439.
- Leygue, E., Dotzlaw, H., Watson, P. H., & Murphy, L. C. (1999). Expression of the steroid receptor RNA activator in human breast tumors. *Cancer Res.*, 4190-4193.
- Maragkakis, M., Alexiou, P., Papdopoulos, G. L., Reczko, M., Dalamagas, T., Giannopoulos, G., . . . Hatzigeorgiou, A. G. (2009). Accurate microRNA target prediction correlates with protein repression levels. *BMC Bioinformatics*, 1-10.
- Marella, N. V., Malyavantham, K. S., Wang, J., Matsui, S.-i., Liang, P., & Berezney, R. (2009). Cytogenetic and cDNA microarray expression analysis of MCF10A human breast cancer progression cell lines. *Cancer Res.*, 5946-5953.
- Margueron, R., & Reinberg, D. (2011). The polycomb complex PRC2 and its mark in life. *Nature*, 343-349.
- Mathelier, A., Zhao, X., Zhang, A. W., Parcy, F., Worsley-Hunt, R., Arenillas, D. J., . . . Wasserman, W. W. (2014). JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucl. Acids Res.*, D142-D147.
- Mattick, J. S. (2005). The functional genomics of noncoding RNA. *Science*, 1527-1528.
- Mattick, J. S. (2013). *The human genome as the zip file extraordinaire*. Heidelberg: EMBO.
- Mattick, J. S., & Makunin, I. V. (2006). Non-coding RNA. *Hum. Mol. Genet.*, R17-R29.
- Mercer, T. R., & Mattick, J. S. (2013). Structure and function of long noncoding RNAs in epigenetic regulation. *Nat. Struct. Mol. Biol.*, 300-307.
- Mercer, T. R., Dinger, M. E., Sunkin, S. M., Mehler, M. F., & Mattick, J. S. (2008). Specific expression of long noncoding RNAs in the mouse brain. *Proc. Natl. Acad. Sci.*, 716-721.
- Mills, A. A. (2010). Throwing the cancer switch: reciprocal roles of polycomb and trithorax proteins. *Nat. Rev. Cancer*, 669-682.

- Mitra, S. A., Mitra, A. P., & Triche, T. J. (2012). A central role for long non-coding RNA in cancer. *Front. Genet.*, 1-9.
- Moran, V. A., Perera, R. J., & Khalil, A. M. (2012). Emerging functional and mechanistic paradigms of mammalian long non-coding RNAs. *Nucleic Acids Res.*, 1-10.
- Nam, J.-W., & Bartel, D. (2012). Long non-coding RNAs in *C. elegans*. *Genome Res.*, 2529-2540.
- Narlikar, G. J., Sundaramoorthy, R., & Owen-Hughes, T. (2013). Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. *Cell*, 490-503.
- Necsulea, A., Soumillon, M., Warnefors, M., Liechti, A., Daish, T., Zeller, U., . . . Kaessman, H. (2014). The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature*, 635-640.
- Neve, R. M., Chin, K., Fridlyand, J., Yeh, J., Baehner, F., Fevr, T., . . . Gray, J. W. (2006). A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell*, 515-527.
- Pelechano, V., & Steinmetz, L. M. (2013). Gene regulation by antisense transcription. *Nat. Rev. Genet.*, 880-892.
- Peric-Hupkes, D., Meuleman, W., Pagie, L., Bruggeman, S., Solovei, I., Brugman, W., . . . van Steensel, B. (2010). Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol. Cell*, 603-613.
- Ponjavic, J., Oliver, P. L., Lunter, G., & Ponting, C. P. (2009). Genomic and transcriptional co-localization of protein-coding and long non-coding RNA pairs in the developing brain. *PLoS Genet.*, 1-14.
- Prensner, J. R., & Chinnaiyan, A. M. (2011). The emergence of lncRNAs in cancer biology. *Cancer Discov.*, 391-407.
- Purev, E., Neff, L., Horne, W. C., & Baron, R. (2009). c-Cbl and Cbl-b act redundantly to protect osteoclasts from apoptosis and to displace HDAC6 from β -tubulin, stabilizing microtubules and podosomes. *Mol. Biol. Cell*, 4021-4030.
- Qiu, M.-T., Hu, J.-W., Yin, R., & Xu, L. (2013). Long noncoding RNA: an emerging paradigm of cancer research. *Tumor Biol.*, 613-620.
- Rajapakse, I., & Groudine, M. (2011). On emerging nuclear order. *J. Cell. Biol.*, 711-721.
- Recillas-Targa, F., & Escamilla-del-Arenal, M. (2004). Participación de la estructura de la cromatina en la regulación de la expresión génica. *Mensaje Bioquímico*, 173-201.
- Rinn, J. L., & Chang, H. Y. (2012). Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.*, 145-166.

- Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Bruggmann, S. A., . . . Chang, H. Y. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell*, 1311-1323.
- Saldaña-Meyer, R., González-Buendía, E., Guerrero, G., Narendra, V., Bonasio, R., Recillas-Targa, F., & Reinberg, D. (2014). CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, Wrap53. *Genes & Dev.*, 723-734.
- Sauvageau, M., Goff, L. A., Lodato, S., Bonev, B., Groff, A. F., Gerhardinger, C., . . . Rinn, J. L. (2014). Multiple knockout mouse models reveal lincRNAs are required for life and brain development. *eLIFE*, 1-24.
- Schuettengruber, B., Martinez, A.-M., Iovino, N., & Cavalli, G. (2011). Trithorax group proteins: switching genes on and keeping them active. *Nat. Rev. Mol. Cell Biol.*, 799-814.
- Schwartz, Y. B., & Pirrotta, V. (2013). A new world of Polycombs: unexpected partnerships and emerging functions. *Nat. Rev. Genet.*, 853-864.
- Simon, M. D., Wang, C. I., Kharchenko, P. V., West, J. A., Chapman, B. A., Alekseyenko, A. A., . . . Kingston, R. E. (2011). The genomic binding sites of a noncoding RNA. *Proc. Natl. Acad. Sci.*, 20497-20502.
- Soule, H. D., Maloney, T. M., Wolman, S. R., Peterson, W. D., Brenz, R., McGrath, C. M., . . . Brooks, S. C. (1990). Isolation and characterization of a spontaneously immortalized human breast epithelial cell line, MCF-10. *Cancer Res.*, 6075-6086.
- Taft, R. J., & Mattick, J. S. (2003). Increasing biological complexity is positively correlated with the relative genome-wide expansion of non-protein-coding DNA sequences. *Genome Biol.*, 5:P1.
- Tsai, M.-C., Manor, O., Wan, Y., Mosammamaparast, N., Wang, J. K., Lan, F., . . . Chang, H. Y. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. *Science*, 689-693.
- Tsai, M.-C., Spitale, R. C., & Chang, H. Y. (2011). Long intergenic noncoding RNAs: new links in cancer progression. *Cancer Res.*, 3-7.
- Ulitsky, I., Shkumatava, A., Jan, C. H., Sive, H., & Bartel, D. P. (2011). Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell*, 1537-1550.
- UniProt Consortium. (2014). *UniProt*. <http://www.uniprot.org/>
- Wahlestedt, C. (2013). Targeting long non-coding RNA to therapeutically upregulate gene expression. *Nat. Rev. Drug Disc.*, 433-446.

- Wan, Y., Qu, K., Ouyang, Z., Kertesz, M., Li, J., Tibshirani, R., . . . Chang, H. Y. (2012). Genome-wide measurement of RNA folding energies. *Mol. Cell*, 1-13.
- Wang, K. C., Yang, Y. W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., . . . Chang, H. Y. (2011). Long noncoding RNA programs active chromatin domain to coordinate homeotic gene activation. *Nature*, 120-124.
- Wapinski, O., & Chang, H. Y. (2011). Long noncoding RNAs and human disease. *Trends Cell. Biol.*, 354-361.
- Wilusz, J. E., Sunwoo, H., & Spector, D. L. (2009). Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev.*, 1494-1504.
- Xu, L., Zhang, Y., Liu, J., Qu, J., Hu, X., Zhang, F., . . . Liu, Y. (2012). TRAIL-activated EGFR by Cbl-b regulated EGFR redistribution in lipid rafts antagonises TRAIL-induced apoptosis in gastric cancer cells. *Eur. J. Cancer*, 3288-3299.
- Yan, S., Qu, X., Xu, C., Zhu, Z., Zhang, L., Xu, L., . . . Liu, Y. (2012). Down-regulation of Cbl-b by bufalin results in up-regulation of DR4/DR5 and sensitization of TRAIL-induced apoptosis in breast cancer cells. *J. Cancer Res. Clin. Oncol.*, 1279-1289.
- Yuan, B., Latek, R., Hossbach, M., Tuschl, T., & Lewitter, F. (2004). siRNA selection server: an automated siRNA oligonucleotide prediction server. *Nucl. Acids Res.*, W130-W134.
- Yuan, S.-X., Yang, F., Yang, Y., Tao, Q.-F., Zhang, J., Huang, G., . . . Zhou, W.-P. (2012). Long noncoding RNA associated with microvascular invasion in hepatocellular carcinoma promotes angiogenesis and serves as a predictor for hepatocellular carcinoma patients' poor recurrence-free survival after hepatectomy. *Hepatology*, 2231-2241.
- Zhang, X., Rice, K., Wang, Y., Chen, W., Zhong, Y., Nakayama, Y., . . . Klibanski, A. (2010). Maternally expressed gene 3 (MEG3) noncoding ribonucleic acid: isoform structure, expression, and functions. *Endocrinology*, 939-947.
- Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucl. Acids Res.*, 3406-3415.